

# Co-Creativity at the Intersection of Place, Poetry, and Generative AI

Ìbùkún Olatúnjí

Submitted to Swansea University in partial fulfilment  
of the requirements for the Degree of Doctor of Philosophy



Swansea University  
Prifysgol Abertawe

Department of Computer Science  
Swansea University

21st September 2025

Copyright: the author, Ìbùkún O látúnjí, 2026



# Declaration

This work has not been previously accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed



Name Ibùkún Olátúnjí

Date 25th February 2026

# Statement 1

This work is the result of my own independent study/investigations, except where otherwise stated. Other sources are clearly acknowledged by giving explicit references. I understand that failure to do this amounts to plagiarism and will be considered grounds for failure of this work and the degree examination as a whole.

Signed



Name Ibùkún Olátúnjí

Date 25th February 2026

# Statement 2

I hereby give my consent for my work, if accepted, to be archived and available for reference use, and for the title and summary to be made available to outside organisations.

Signed



Name Ibùkún Olátúnjí

Date 25th February 2026







# Abstract

This thesis investigates how generative AI can support and reshape human creativity across place, poetry, and performance. It makes three original contributions: (i) introducing cultural geography as a framework for designing and evaluating AI-enabled creative infrastructure; (ii) developing computational proxies such as rhyme density and cadence similarity for assessing poetic and performance creativity; and (iii) pioneering adversarial human–AI interaction as a method for evaluating co-creativity. Methodologically, it integrates spatial inquiry, poetic analysis, and design-based evaluation, positioning creativity as a relational practice that emerges across human, machine, and cultural contexts. Grounded in hip hop principles, Study I examines how physical environments shape creativity through the design and activation of cultural venues in Swansea and London. The study demonstrates how purpose-built infrastructure can function as a test bed for digital cultural systems, a perspective rarely addressed in AI research. Study II investigates AI-assisted poetic composition using computational writing tools. Structured tasks, participant evaluations, and rhyme density as a novel proxy are employed to assess how system types influence creativity. Study III extends into spoken-word performance, presenting a systematic evaluation of AI in freestyle rap. Six experiments assess voice clone similarity, cadence accuracy, and text generation, combining computational analysis with human judgements. The findings demonstrate cadence as a critical dimension for human–AI co-performance. Together, the three studies contribute to a layered model of human–AI co-creativity, spanning spatial, textual, and performance domains. The thesis positions adversarial evaluation as its central contribution, demonstrating how creative sparring provokes originality. Stylistic and improvisational dimensions serve as complementary modes of analysis, offering interpretable lenses on the co-creative process.



## **Author's Statement**

The research reported in this thesis was conducted by the author, who served as principal researcher and interaction designer. The author carried out all user research and requirements gathering; prototype concepts and designs were developed by the author in consultation with collaborators at Arts ARKADE. The design studies were conceived by the author and executed with the support of external experts and collaborators. Contributions by others are explicitly identified and credited in the relevant sections throughout the thesis.

## **Ethics Statement**

As Principal Investigator I took responsibility to explain, in appropriate detail, what the research was about to all participants. Participants were given detailed explanations that outlined the purpose of the studies, and how they would be disseminated and used. Participation in the research was voluntary and informed consent was obtained from all participants.



# Acknowledgements

There are many people to thank and hopefully I have not forgotten anyone.

**Swansea University:** My deepest thanks to **Matt, Alma,** and **Amanda** for your insights, advice, encouragement, and for making the journey so enjoyable. I'm looking forward to what comes next! Also, others in the Computational Foundry who helped and supported in numerous ways over the years: **Simon, Fergus, Matt** (in the first year especially), **Lyds, Ben, Tashi, Sherryl** and **Oliver**.

**Swansea Council:** I am grateful to **Tracey McNulty** for generously opening the door to this research opportunity. Arts ARKADE was born in Swansea, and it would not have been possible without the support of **Jamie Corteen, David Price Deer, Kate Wood, Steve Hopkins, Frances Jenkins,** and many others (**Kingsley, Dougie**).

**The Swansea artistic community:** It was a pleasure to work with and learn from **Jo Furber, Amina Abu-Shahba, Kieron Jones, Johan Butenschøn Skre, Catherine Bennett, Jon Doyle, David Toft, Jason & Becky, Vicky O'Neon,** and **Simon Parton**. I also appreciate **Urban Foundry** and **Ben**, for being so accommodating with time and much more; and **Gareth**, for assisting even when you had no need to. Your kindness was appreciated.

**The Volcano Theatre:** Thank you to Paul, Claude, and especially Catherine for the remarkable performance with the youth group.

**The Crown Estate:** I will forever see Piccadilly differently. Thank you for the chance to reimagine such an iconic part of London. My gratitude goes to **Anna Swaites, Deborah Clark, Beth Eastgate, Tim Allibone, Sara Lee,** and **Joshua Peacock**. It was a pleasure to

work with you, and your support during the most challenging times is deeply appreciated. Thanks also to **Dan Labbad** for your encouraging words.

**Arts ARKADE:** Friends and colleagues here have been a source of inspiration. Shout-outs to **Alex, Anuj, Kelwyn, Becky, Dougi, Gayna,** and **Shaun**. In particular, Dougi for countless technical discussions that brought the building designs to life; Alex for insight into community-building; and Gayna for challenging me on interdisciplinary practice — you were right, though I may not have admitted it often enough at the time.

**Poets and MCs, and Wordsmiths:** With gratitude to all the poets, MCs, and participants who shared their time and creativity in the studies. This thesis would not have been possible without your contributions.

A PhD is perhaps a marathon, and training partners are essential. In **Mark Sheppard**, I could not have been more fortunate: your friendship, guidance, and insights on music and digital signal processing shaped this work and arrived just when they were most needed. **Professor Philip Treleaven** helped me maintain perspective — with good humour, which is perhaps the most efficient critique method of all.

**Finally:** Shaun — more than a business partner and best friend, truly a brother. Ginny and David — I could not ask for more supportive parents-in-law. The thesis may not make up for my many absences, but I hope it gives you a sense of what I have been doing all this time. My daughter, **Amélie**, you have been an incredible source of support, fun, and quite brilliant Netflix recommendations. I've also been inspired by your own educational journey over the past few years. And my wife **Liza**, you have been patient and loving in equal measure. Without your encouragement, I would never have attempted any of this. On occasion, perhaps we both wished this had been the case — but by the time we realised it was too late.

# Contents

Author's Statement . . . . .	v
Ethics Statement . . . . .	v
<b>List of Tables</b>	<b>xii</b>
<b>List of Figures</b>	<b>xv</b>
<b>1 Introduction</b>	<b>3</b>
1.1 Context and Motivation . . . . .	4
1.2 Aim and Research Questions . . . . .	5
1.3 Practitioner Expertise and Researcher Positioning . . . . .	6
1.4 Methods . . . . .	7
1.5 Contributions and Claims . . . . .	9
1.6 Thesis Structure . . . . .	10
<b>Research Outputs</b>	<b>13</b>
<b>2 Literature Review</b>	<b>19</b>
2.1 The Psychology of Human Creativity . . . . .	19
2.2 Creativity as Socio-Cultural System . . . . .	26
2.3 Creative Placemaking . . . . .	28
2.4 Creative, Playable, and Smart Cities . . . . .	32
2.5 Culture: Arts and Language . . . . .	42
2.6 Poetry, Hip Hop, and Lyrical Expression . . . . .	47
2.7 Computational Creativity . . . . .	56
2.8 Co-Creative Spatiality . . . . .	70

<b>3</b>	<b>Study I: Place as a <i>Hip Hop State of Mind</i></b>	<b>81</b>
3.1	Introduction . . . . .	82
3.2	Context and Related Work . . . . .	82
3.3	Research Aim and Study Focus . . . . .	85
3.4	Method . . . . .	86
3.5	Findings . . . . .	113
3.6	Discussion and Future Work . . . . .	117
3.7	Publication and Contribution Note . . . . .	119
<b>4</b>	<b>Study II: Poetry and Writing Assistants</b>	<b>121</b>
4.1	Introduction . . . . .	122
4.2	Related Work . . . . .	127
4.3	Research Aims . . . . .	132
4.4	Method . . . . .	134
4.5	Technology Stack . . . . .	136
4.6	Participants . . . . .	138
4.7	Results and Analysis . . . . .	140
4.8	Discussion and Future Work . . . . .	152
4.9	Publication and Contribution Note . . . . .	156
<b>5</b>	<b>Study III: Generative AI and Co-Creativity</b>	<b>159</b>
5.1	Introduction . . . . .	160
5.2	Related Work . . . . .	164
5.3	Research Aims . . . . .	169
5.4	Method . . . . .	174
5.5	Technology Stack . . . . .	177
5.6	Participants . . . . .	183
5.7	Results and Analysis . . . . .	185
5.8	Discussion and Future Work . . . . .	206
5.9	Publication and Contribution Note . . . . .	209
<b>6</b>	<b>General Discussion and Future Work</b>	<b>213</b>
6.1	Discussion . . . . .	214
6.2	DCTs, Digital Replicas, and Performer Rights . . . . .	222

6.3	Future Work . . . . .	223
6.4	Conclusion . . . . .	229
6.5	Publication and Contribution Note . . . . .	231
	<b>Bibliography</b>	<b>233</b>
	<b>Appendices</b>	<b>266</b>
<b>A</b>	<b>Study I: Supplementary Materials</b>	<b>267</b>
A.1	Artcamp: Output and Artefacts . . . . .	267
<b>B</b>	<b>Study II: Supplementary Materials and Analyses</b>	<b>271</b>
B.1	Rhyme Types and Positions . . . . .	274
B.2	Rhyme Density Method and Worked Example . . . . .	278
<b>C</b>	<b>Study III; Supplementary Materials and Analyses</b>	<b>297</b>
C.1	Extended Rhyme Analysis . . . . .	297
C.2	Speech Systems Taxonomy . . . . .	300
<b>D</b>	<b>Study III: Further Experimental Data</b>	<b>301</b>
D.1	Overview of Study III Experiments . . . . .	306
D.2	E1a (Exploratory): Voice identity checks across conditions . . . . .	307
D.3	Experiment E1: Voice Clone Identity and Similarity . . . . .	308
D.4	Experiment E2: Cadence Similarity across TTS Models . . . . .	309
D.5	Experiment E3: Cadence and Pitch Consistency . . . . .	312
D.6	Experiment E4: Text Style and Rhyme Quality . . . . .	316
D.7	Experiment E5: Poetic Co-Creation (TTS vs Remix) . . . . .	319
D.8	Experiment E6: MC Case Study . . . . .	329
D.9	Experiment E7: LYRICAL KOMBAT Flow Analysis . . . . .	331
D.10	Text Generation: Full Prompts and Judging Protocol . . . . .	333
<b>E</b>	<b>Ethics Approval Forms</b>	<b>348</b>
	<b>Participant Information Sheets and Consent Forms</b>	<b>350</b>

# List of Tables

1.1	Thematic Emphasis Across Thesis Studies . . . . .	12
2.1	Placemaking Expressions Summary . . . . .	29
2.2	A Typology of Urban Play . . . . .	34
2.3	Overview of City Frameworks . . . . .	38
2.4	Key Elements of Rap . . . . .	48
2.5	Rap Flow Model Comparison . . . . .	52
2.6	Markov Chains and N-gram Models . . . . .	62
2.7	Hidden Markov Models (HMMs) . . . . .	63
2.8	Neural Networks (RNNs) and LSTMs . . . . .	64
2.9	Transformer Models . . . . .	67
2.10	Probabilistic Text Generation Overview . . . . .	68
2.11	Computational Story and Poem System Types . . . . .	69
2.12	Dimensions of Co-Creativity . . . . .	71
2.13	Parallels between Hip Hop and Generative AI . . . . .	78
3.1	Study I: Research Questions . . . . .	86
3.2	Study I: Framework for Arts ARKADE Studies . . . . .	89
3.3	Study I: Participants . . . . .	98
3.4	Study I: Arts ARKADE Swansea Outputs . . . . .	102
3.5	Study I Hypotheses and Findings . . . . .	114
3.6	Study I Research Questions to Findings Mapping . . . . .	116
4.1	Study II: Writing Technologies Evolution . . . . .	127
4.2	Study II: Categories of Writing Assistants . . . . .	128
4.3	Study II: Co-creative Design Tensions . . . . .	129
4.4	Study II: Research Questions . . . . .	133

4.5	Study II: Design Method . . . . .	136
4.6	Study II: Writing Assistants used within the Study . . . . .	137
4.7	Study II: Participant Summary . . . . .	138
4.8	Study II Statistical Analyses Overview . . . . .	141
4.9	Study II: Participant Characteristics . . . . .	142
4.10	Study II DAT and Rhyme Score Correlations . . . . .	142
4.11	Study II: Predicting Rhyme Density . . . . .	143
4.12	Study II: Experience Level Effects on Rhyme Scores . . . . .	143
4.13	Study II: Experience Level Effects on Rhyme Scores . . . . .	144
4.14	Study II: Predictors of Perceived Tool Usefulness . . . . .	144
4.15	Study II: Summary of Research Questions and Findings . . . . .	148
5.1	Study III: Design Challenges for Digital Creative Twins (DCTs) . . . . .	168
5.2	Study III: Literature Review Themes . . . . .	169
5.3	Study III: Research Questions and Hypotheses . . . . .	171
5.4	Study III: Rap Battle Interaction Structure . . . . .	179
5.5	Study III: Human MC vs. DCT Across Battle Stages . . . . .	181
5.6	Study III: Experiments Summary . . . . .	186
5.7	Study III: Voice Engine Ratings . . . . .	187
5.8	Study III: Participant Characteristics . . . . .	188
5.9	Study III: Clone Rating Metrics . . . . .	189
5.10	Study III: Clone Metrics Across Battle Rounds . . . . .	190
5.11	Study III: Predicting Clone Rating by Features . . . . .	190
5.12	Study III: Subjective Co-Creativity Ratings . . . . .	191
5.13	Study III: Correlation between Subjective Scores, DAT, and Clone Rating. . . . .	191
5.14	Study III: Results for RQ3-1–RQ4-2 . . . . .	192
5.15	Study III: MC Case Study Scores . . . . .	193
5.16	Study III: MC vs Poet Evaluation Comparison . . . . .	193
5.17	Study III: Summary of Clone Quality and Flow . . . . .	196
5.18	Study III: AI vs. Human Phoneme Comparison . . . . .	197
5.19	Study III: LYRICAL KOMBAT Results . . . . .	198
5.20	Study III: Audio Datasets . . . . .	200
5.21	Study III: Acoustic Comparison of Best vs Worst Clones . . . . .	202
5.22	Study III: Best vs. Worst Clones Performance Profile by Feature. . . . .	204

6.1	Research Questions Addressed Across Studies . . . . .	217
B.1	Spearman Correlation Coefficients . . . . .	271
B.2	Study II: Mann-Whitney U Tests for User Experience . . . . .	273
B.3	Rhyme Type Descriptions . . . . .	275
B.4	Rhyme Positions and Descriptions . . . . .	276
B.5	Rhyme Types and their Positions . . . . .	277
B.6	Study II base poems chosen for qualitative inspection. . . . .	292
C.1	Study II vs. Study III rhyme analysis . . . . .	298
C.2	Speech Systems Taxonomy . . . . .	300
D.1	Rap Flow Adams vs. Condit-Schulz . . . . .	303
D.2	Rap Flow Global Features . . . . .	304
D.3	Study III experiments overview . . . . .	306
D.4	Voice Identity Matching Tests . . . . .	307
D.5	Voice Model Evaluation . . . . .	310
D.6	Supplementary Table <i>p</i> -values for TTS Clone Comparisons . . . . .	314
D.7	Effect Sizes of Best vs. Worst Clone Comparisons . . . . .	314
D.8	LLM used in E4 . . . . .	316
D.9	E4 data generation and selection schema. . . . .	317
D.10	E7 Paired Tests . . . . .	319
D.11	Exploratory Correlations . . . . .	322
D.12	MC Case Study . . . . .	329
D.13	LK Descriptive Deltas . . . . .	331
D.14	LK Per-Pair MOS and Rhymes . . . . .	331
D.15	WER Components and Rates. . . . .	342

# List of Figures

2.1	Key Literature Concept Map . . . . .	20
2.2	The Dual Pathway to Creativity Model. . . . .	24
2.3	The Metacontrol State Model . . . . .	25
2.4	Creativity as Socio-Cultural Process . . . . .	27
2.5	Urban Play Example: Stop Smile Stroll Intervention . . . . .	37
2.6	The Cultural City Model . . . . .	41
2.7	Schematic of Ohringer’s Flow Model . . . . .	50
2.8	Speech–Song Continuum in Rap Delivery . . . . .	53
2.9	Eminem transcription on the speech–song continuum. . . . .	54
2.10	Self-Attention Mechanism . . . . .	66
3.1	Arts ARKADE: Launch, February 2024 . . . . .	81
3.2	Study I Venn Diagram . . . . .	85
3.3	Former Cranes Music Shop . . . . .	91
3.4	Arts ARKADE Swansea Mood Board . . . . .	92
3.5	Arts ARKADE Swansea Window Design . . . . .	93
3.6	Arts ARKADE Swansea Immersive Screens . . . . .	94
3.7	Arts ARKADE Swansea: Building Layout and Activity Zones . . . . .	96
3.8	Arts ARKADE Swansea Artcamp . . . . .	99
3.9	Arts ARKADE Swansea Hip Hop . . . . .	101
3.10	Arts ARKADE London Development . . . . .	105
3.11	Arts ARKADE London Main and Immersive Spaces . . . . .	106
3.12	Arts ARKADE London Interior Design . . . . .	107
3.13	Arts ARKADE London Metaverse Prototype . . . . .	108
3.14	Arts ARKADE London Exterior during Long Dark Tunnel Exhibition . . . . .	110
3.15	Arts ARKADE London Interior during Long Dark Tunnel Exhibition . . . . .	111

3.16	The Beatsora Cypher at Arts ARKADE London. . . . .	112
4.1	Jeremy Deller 'More Poetry' Installation . . . . .	121
4.2	Study II Venn Diagram . . . . .	123
4.3	Study II: Key Topics Map . . . . .	125
4.4	Study II: Writing Technologies Concept Map . . . . .	157
5.1	Human–AI Battle Rap Illustration . . . . .	159
5.2	Continuum of expressive forms in Study III . . . . .	160
5.3	Study III Venn Diagram . . . . .	161
5.4	Study III: Key Topics Map . . . . .	162
5.5	Creative Beam Search Generation Process . . . . .	164
5.6	DCT Response Overview . . . . .	178
5.7	Text-to-Text Remix Pipeline . . . . .	180
5.8	Summary of LYRICAL KOMBAT Effects . . . . .	198
5.9	Study III Concept Map . . . . .	210
6.1	Future Work Venn Diagram . . . . .	213
6.2	DCT–MC Battle Visualisation . . . . .	226
A.1	Study I: Artcamp gesture experiment example . . . . .	268
A.2	Study I: Artcamp pose estimation test example . . . . .	269
B.1	Study II: Summary of Statistical Findings . . . . .	274
D.1	Beat-aligned Stress . . . . .	302
D.2	Bar-level Rest . . . . .	302
D.3	Tupac transcription on the speech–song continuum. . . . .	305
D.4	Voice Engine × Rating Type . . . . .	310
D.5	Voice Quality by Clone . . . . .	310
D.6	Cadence Similarity by Clone . . . . .	310
D.7	Listener Ratings: quality and cadence . . . . .	310
D.8	Best clones: cadence DTW distance heatmap . . . . .	313
D.9	Worst clones: cadence DTW distance heatmap . . . . .	313
D.10	Phoneme complexity by role and test . . . . .	320
D.11	Per-item questionnaire distributions . . . . .	321

D.12 Mean questionnaire scores . . . . . 321



# List of Abbreviations

- AI** Artificial Intelligence
- ASR** Automated Speech Recognition
- CBS** Creative Beam Search
- CADT** Creative Adversarial Test
- DAT** Divergent Association Task
- DCT** Digital Creative Twin
- DTW** Dynamic Time Warping
- HCAI** Human-Centred Artificial Intelligence
- HCI** Human-Computer Interaction
- LLM** Large Language Model
- MC** Master of Ceremonies
- MFCC** Mel-Frequency Cepstral Coefficient
- MVA** Minimum Viable Artwork
- STT** Speech-to-Text
- TTS** Text-to-Speech
- WER** Word Error Rate



# Chapter 1

## Introduction

Evidence suggests that cultural activity drives development most effectively where two local conditions coincide: (i) a *creative environment* (talent, practices, industries) and (ii) a *cosmopolitan local identity*, trust, tolerance, and openness that enable exchange [1]. Cities that combine both (“creative cosmopolitan” places) are measurably better at turning cultural participation into wider urban and even regional gains, producing “pull” effects beyond the city boundary [1]. In this view, culture is not simply programming; it is part of a city’s territorial capital: the social, relational, and cognitive assets through which places learn, innovate, and prosper [1]. The central question, then, is not whether culture matters, but how specific cultural infrastructures and tools can amplify practice and translate participation into shared value for people and place. Alongside conventional notions of ‘placemaking’, this thesis draws directly on recent cultural geography that engages with technology.

Swansea’s City of Culture bid set out an ambition that culture should play a *transformational* role that unifies people and place [2]. This thesis takes that proposition and builds on it theoretically and practically: it focuses on urban culture, often called *street culture*, and examines how digital cultural infrastructure and generative AI can be designed to stimulate creativity rather than merely imitate it [3, 4]. Swansea is both the starting point and the testing ground for this inquiry: Arts ARKADE—a *pop-up, low-cost, AI-enabled creative space and programme of activities*, funded in partnership with the local authority, provided the core empirical setting for this thesis, and shaped the research questions about how AI tools sit within creative practice and place. Arts ARKADE served as a laboratory in which place, poetry, and technology-enhanced performance were developed with

communities in situ. The policy intuition that creative, cosmopolitan cities can pull wider regional growth is well-established; importantly, it depends on participation and open, trusting civic climates that let culture travel from venues and scenes into everyday life.

This work situates Swansea’s cultural programme in a broader set of ideas from creative geographies and urban media. Hybrid public spaces, where physical settings are interlaced with digital layers, have been shown to reshape how people meet, express, and govern themselves, reframing parks, streets and venues as programmable civic interfaces [5]. Playful, participatory approaches to citymaking demonstrate how co-design and public authorship of digital experiences can cultivate agency and belonging [6] A recent scoping review of *creative and digital placemaking* calls out fragmentation across disciplines and a shortage of robust evaluation linking cultural aims to socio-technical interventions, gaps this thesis addresses empirically [7]

To connect place to creative tooling, the lens of *co-creative spatialities* is employed; this is the idea that creativity is distributed across people, places and systems. AI is part of a *more-than-human* ensemble rather than a drop-in replacement author. This approach argues for collaboration, situated use, and the embodied environments under which creativity is made and judged [8] Beyond documenting asymmetries in access to AI, this thesis advances the notion of *culturally grounded* AI systems: tools designed and evaluated in situ, with local practices, resources, and constraints treated as first-class design parameters.

Hip hop is employed as framework that links place and AI. As a culture, hip hop also binds text to rhythm to body to crowd; as Bradley argues, “every rap song is a poem waiting to be performed” [9]. This is echoed by Schmidt, who suggests that how poems reach their audience has an effect on how poems are conceived in technological terms [10]. Hip hop thus provides a context for claims about AI creativity, and a route to study *evaluation as stimulation*: does a system help a human make something better, braver, or *more their own*?

### 1.1 Context and Motivation

Physical interventions (Arts ARKADE Swansea, then London) are approached as *digital cultural infrastructure*: publicly visible, community-embedded settings where creative practice is scaffolded across physical and digital space. This responds to three linked needs. First, policy and scholarship emphasise that cultural vibrancy and cosmopolitan identity co-produce urban and regional development; participation is the mediating channel. Second, creative and digital placemaking lacks integrative, evaluable pathways

from intention (e.g., inclusion, authorship) to outcomes; methods are required that travel between field deployments and laboratory-style studies [11]. Third, AI has entered public culture not only as a tool but as an actor in hybrid spaces; design must therefore reckon with co-agency, and more-than-human intelligences in situated practice [8, 12].

## 1.2 Aim and Research Questions

The goal of the research was to develop evidence and frameworks to construct *situated* environments, across physical and digital settings, which nurture creative practice and agency. This goal was articulated by the questions below.

1. **Place and Practice.** How does digital cultural infrastructure (e.g., Arts ARKADE) influence creative practice in place-based contexts?
2. **Tools and Process.** How do different classes of writing tools (support, co-creative, autonomous) affect fluency, accuracy, and perceived creativity in poetic composition?
3. **Proxies and Evaluation.** In what ways can textual and performance measures (e.g., rhyme density, cadence similarity) serve as *interpretable* proxies for co-creativity?
4. **Adversarial Co-Creation.** How can adversarial human–AI interaction (e.g., rap battles with a Digital Creative Twin) be used to evaluate and *stimulate* creativity via a Creative Adversarial Test (CADT)?

### Relation of Studies to Research Questions

The three empirical studies address the four primary research questions in a staged and progressive manner. Study I examines how creative practice unfolds in situated, hybrid environments and therefore speaks directly to **RQ1 (Place and Practice)**. Study II introduces structured writing tools and constrained poetic tasks, addressing **RQ2 (Tools and Process)** while operationalising rhyme density as an interpretable proxy for **RQ3 (Proxies and Evaluation)**. Study III builds directly on the methods and findings of Study II by extending rhyme- and cadence-based evaluation into live, adversarial human–AI performance, thereby addressing **RQ3** and **RQ4 (Adversarial Co-Creation)**. Taken together, the studies form a cumulative research programme in which each contribution advances the corresponding research question(s) and informs the design of those that follow.

### 1.3 Practitioner Expertise and Research Positioning

The interdisciplinary orientation of this thesis is shaped directly by the author's longstanding practice as a poet, lyricist and creative producer. Over more than fifteen years, this work has included professional publications, invited festival performances, creative commissions, and collaboration with leading cultural institutions across the United Kingdom and internationally. These experiences form a substantive practitioner background that informs the design, interpretation and evaluation of the creative processes studied in this thesis.

As a writer and performer, the author has been selected for programmes at the Donmar Warehouse (Future Forms), Criterion Theatre (Criterion New Writing 15), the Jaipur Literature Festival, and has written libretti and narrative texts for the London Philharmonic Orchestra and the Southbank Centre. The author's original narrative poem, *The Battle of the Wordsmiths*, was nominated for a PRS New Music Award [13], and later supported by the PRS for performance at the Midland Arts Centre. The author's other creative works include theatre (*Motherland*), music (*The Butterfly's Wings*, *The Word Garden*), and spoken-word performance (with The Versability Shop at the Cheltenham Literature Festival). As a producer and creative director, the author has led projects combining narrative, performance, and technology, including adaptations of *The Battle of the Wordsmiths* [14], *The Wild Kingdoms*, and digital games [15].

In parallel, the author has a sustained research trajectory in human–AI interaction and creative technologies. This includes peer-reviewed publications and submissions to CHI, IUI, and ICCV, covering topics such as co-creative writing systems, evaluation of lyrical AI, and the use of battle rap as a framework for human–machine interaction. Prior work at UCL Knowledge Lab involved participatory design with youth communities, literacy technology development, and educational game research, complementing the creative practice with methodological experience in mixed-methods studies.

This background, creative and computational, positions the author not only as an investigator, but as an experienced practitioner within the domains under study. It serves three roles in the thesis: (i) informing the design of creative tasks and study protocols, (ii) shaping the analytic framing of rhyme, cadence and performance-based evaluation, and (iii) grounding claims in practitioner insight consistent with practice-led research models such as Smith and Dean's iterative cyclic process [16]. The studies presented in this thesis therefore draw on practitioner expertise as an analytic and methodological

resource, shaping how creativity, co-design, and performance dynamics are understood across place-based interventions, poetic tasks and adversarial human–AI interaction.

## 1.4 Methods

Given the interdisciplinarity (creative geography, computational creativity), a mixed-methods design is employed. The thesis combines practice-led deployments with controlled studies. First, case studies in Swansea and London treat venues, exhibitions and workshops as field sites for observing and shaping participation, authorship, and identity in hybrid public–digital spaces [5, 6, 7]. Second, structured evaluations interrogate contemporary LLM based assistants on constrained poetic tasks, blending quantitative text analysis with qualitative user-experience measures [17]. Third, live adversarial experiments (spoken word, freestyle rap) pair human MCs with DCTs, using cadence-aware models and human ratings to study co-creative *stimulation* rather than imitation [18].

### Ethical Framework

This research adopts a cross-study ethical framework grounded in principles of informed consent, participant agency, cultural sensitivity, and responsible innovation. Given the interdisciplinary nature of the work, spanning creative geography, computational creativity, voice technologies, and human–AI collaboration, ethical considerations extend beyond standard research governance to encompass questions of authorship, identity, and the social implications of generative systems. Participants across all studies were treated not merely as data sources but as active co-creators.

Across all three studies—Study I (placemaking and digital cultural infrastructure), Study II (structured poetic tasks), and Study III (voice-based human–AI performance)—the same core ethical principles were applied, with study-specific safeguards introduced where required by the nature of the interaction and the sensitivity of the data. In Study I, informed consent was obtained prior to all workshops, public activities, and interviews, with attention to power dynamics in community–researcher relationships. In Studies II and III, which involved structured poetic tasks and performance-based interactions with AI systems, additional safeguards were introduced. These included briefing participants on the nature of AI involvement, offering the right to withdraw at any time, and ensuring

## *1. Introduction*

---

they retained control over their creative outputs. Particular care was taken with voice-based data. Spoken-word performances, freestyle raps, and AI-generated vocal outputs were anonymised and securely stored. For Study III, which included the use of voice cloning, participants were provided with clear documentation explaining how their voice samples would be used, and were given the opportunity to review and approve generated content prior to inclusion in analysis. A minimal-exposure protocol was implemented to limit internal access to sensitive audio data, and cloned voices were used solely for non-commercial, research-oriented evaluation.

More broadly, the thesis anticipates emerging frameworks for ethical AI deployment in creative contexts. Questions of voice ownership, authorship, and attribution were foregrounded throughout the research design. Where applicable, participants were consulted on how they wished to be credited, and study artefacts were treated as co-authored where appropriate. This aligns with current work on ethical generative systems, including the UKRI principles on Responsible Innovation, the UNESCO Recommendation on the Ethics of Artificial Intelligence, and recent Creative IP Lab guidance on digital performance rights and synthetic media governance.

Formal ethics approval was granted by Swansea University (Research Ethics Approval Number: 2 2025 9520 12891). Specific protocols and consent forms are provided in Appendix E. Each empirical chapter also includes references to the ethical procedures and decisions relevant to that study.

---

## Software Provenance and Authorship

Component	Authorship	Notes
Rhyme analysis notebooks	Author (this thesis)	
TTS/remix game design	Author (this thesis)	Voice data not retained
Prompt sets and judging scripts	Author (this thesis)	Internal use only
ASR and TTS configuration	Author (this thesis)	Configs retained; no audio released
Cadence feature analysis	M. Sheppard (proprietary)	Summary features only
Audio feature analysis	M. Sheppard (proprietary)	Derived metrics only
Poetry/MC survey tools	Author (this thesis)	Anonymised data deleted
LYRICAL KOMBAT	Author (this thesis)	Experimental prototype only

*Note.* The tools and scripts used in this study were authored by the researcher or collaborators. These components were developed specifically for the experimental framework described in Chapters 4 and 5. In line with the approved ethics protocol and participant requests, no data nor code has been made publicly available. All sensitive audio files, transcripts, and related metadata were permanently deleted following the conclusion of the study.

## 1.5 Contributions and Claims

**Claim 1 (Place as active variable).** When configured as digital cultural infrastructure, place measurably reshapes participation and authorship. The city’s creative and cosmopolitan climate—and the extent of public cultural participation—condition how such benefits travel beyond the immediate site [1, 7]. Demonstrated in Chapter 3.

**Claim 2 (From imitation to stimulation).** Evaluation should prioritise whether systems *help* people create (stimulation) rather than simply match style (imitation). This principle is operationalised through constrained writing tasks and the Creative Adversarial Test (CADT) for performance [18, 17]. Demonstrated in Chapters 4 and 5.

**Claim 3 (Interpretable proxies).** Rhyme density and cadence similarity function as interpretable, task-aware proxies that bridge automated analysis and human judgement while keeping proxy limits explicit [17, 18]. Demonstrated in Chapters 4 and 5.

**Claim 4 (More-than-human co-agency).** Treating AI as part of a more-than-human ensemble, rather than as a human replacement, supports designs (e.g., voice-first DCTs) and methods (e.g., adversarial sparring) that better align with creative geographies and public culture [6, 8]. Demonstrated in Chapters 3 and 5.

Conceptually, the thesis reframes creativity as relational across place, poetry, and performance, shifting evaluation from imitation to *stimulation*. Methodologically, it proposes a layered evaluation stack that combines interpretable proxies (rhyme density; cadence similarity) with task-aware human judgements (CADT). Technically, it operationalises rhyme density for structured poetic tasks, extends toward cadence-sensitive measures, and introduces Digital Creative Twins (DCTs) as adversarial sparring partners for improvisation within performance settings. Practice-led artefacts developed during the research include large-scale creative prototypes, public workshops, and experimental digital tools.

## 1.6 Thesis Structure

The thesis is organised to move from context and theory, through practice-based studies, to synthesis and future work. Each chapter advances a single argument: that creativity in hip hop, anchored in place, negotiated through text, and realised in performance can guide the design and evaluation of culturally-grounded AI systems.

**Chapter 2. Literature Review.** The chapter positions the work at the intersection of placemaking, computational creativity, and HCI. It develops three lenses *place* as digital cultural infrastructure, *text* under constraint, and *performance* as embodied representation of cadence and audience negotiation; it uses the lenses to frame evaluation from imitation toward provocation. The review synthesises debates on 'AI creativity' and argues for interpretable proxies (e.g., rhyme density and cadence similarity) layered with human judgement, and motivates voice interaction as a design stance for culturally specific practice. The chapter closes by specifying why public-realm interventions are an appropriate

methodological starting point, thereby setting up Chapter 3.

**Chapter 3. Study I: Arts ARKADE as Place.** Here the thesis moves into the field, treating Swansea and London Arts ARKADE sites as living laboratories. Using design ethnography, exploratory case study methods, participant observation, and interviews/focus groups, the chapter examines how interventions, workshops, and exhibitions reorganise participation, authorship, and presence. From these findings, the chapter derives situated prompts, genre cues, and constraints that travel into subsequent studies, and articulates design principles for small, steerable, voice tools grounded in community norms. This empirical grounding bridges directly to Chapter 4, where these prompts and principles are operationalised in structured writing tasks.

**Chapter 4. Study II: Writing Assistants for Poetry.** The second study narrows to text as a tractable surface for measuring tool effects. It introduces a controlled experimental frame that compares support, co-creative, and autonomous assistants on constrained poetic tasks seeded in Chapter 3. Methods combine quantitative text analysis with qualitative user experience and process accounts; rhyme density is operationalised as an interpretable proxy for patterned constraint, alongside fluency/accuracy checks and reflective accounts of agency and voice. The chapter also explores perception on AI vs human authored text; this motivates a shift from page to stage, and from plausibility to timing and responsiveness, which in turn frames Chapter 5.

**Chapter 5. Study III: Human–AI Co-Creativity.** The third study relocates evaluation to live, adversarial collaboration, using freestyle/battle rap as a rigorous test bed for real-time co-creation. It introduces DCTs aligned to a performer’s style and evaluates them through a process that scores provocation, responsiveness, and originality.

**Chapter 6. Discussion and Future Work.** The final chapter synthesises work across place, poetry and generative AI into a layered model of co-creativity, revisiting the research questions and clarifying contributions. It reflects on limitations, as well as issues around ethics and consent. The discussion ends by going from public-realm experimentation, to real-time co-creation, establishing a potential roadmap for situated, accountable human–AI

## 1. Introduction

---

practice.

Study / Chapter	Place	Poetry	Generative AI	Performance
Study I: Arts ARKADE	● ● ●	● ●	●	● ●
Study II: Writing Tools	● ●	● ● ●	● ●	●
Study III: Creative Twins	●	● ● ●	● ● ●	● ● ●

Table 1.1: **Thematic Weighting Matrix.** This table summarises the four core dimensions of the conceptual framework—Place, Poetry, Generative AI, and Performance—across Studies I–III. It provides a structural overview showing how each study contributes to the arguments developed throughout the thesis.

**Key:** Each dot indicates relative emphasis of a theme in the given study. More dots = greater prominence. Dot colours correspond to:

● Place, ● Poetry, ● Generative AI, ● Performance.

# Research Outputs

The thesis was developed and disseminated through the following activities, events, and publications.

## **Arts ARKADE Swansea (2022–2023)**

The Arts ARKADE Swansea intervention contributed directly to the practice-based components of this thesis. Activities included collaborative creative workshops, installations, and public performances that engaged participants in experimental poetic and narrative production. Further detail is contained in Chapter 3.

## **Workshops, Exhibitions, and Performances**

**Artcamp I.** Collaborative residency and ideation event for poets, artists, and technologists. 16–17 February 2022.

**Press Launch.** Official opening by Swansea Council attended by the Welsh Government's Deputy Minister for Arts and Sport. 9 March 2022.

**Artcamp II.** Extended community co-creation focused on site-specific storytelling. 4–6 May 2022.

**Hip Hop for Health.** Public engagement event merging health messaging with freestyle and written lyricism. 24–25 June 2022.

**Rebooked.** Interactive poetry and memory installation. 15 July–16 September 2022.

**Volcano Theatre.** Performed poetic responses using early co-creative tool prototypes. 31 October–1 November 2022.

**Dylan Thomas Centre Writing Workshop.** Writing workshop with poets and community participants. 28 January 2023.

## Arts ARKADE London (2023–2025)

Arts ARKADE London emerged as a continuation of the Swansea intervention, expanding the design, deployment, and user evaluation of writing systems. With a larger number of events and exhibitions, this phase broadened both the cultural reach and the empirical grounding of the research. Further documentation and evaluation are referenced in Chapter 3.

### Reports, Events, and Exhibitions

**Launch Event.** Opening of Arts ARKADE London with performances and public demonstrations. 25 January 2024.

**Walking the Line Book Launch.** Jimmy C. 23 May 2024.

**The Vault Exhibition Space.** Ongoing exhibition of collaborative work. 5 June 2024–10 July 2025.

**Grime Culture Photography Exhibition.** ALI AD3N. 20 June 2024.

**The Crown Estate Annual Report 2023/2024.** Arts ARKADE case study feature. 24 July 2024.

**Arts ARKADE Annual Report 2023/2024.** Documenting activities and outcomes. 19 August 2024.

**The Evening Standard and Samsung: The Imagination Playground.** Public installation and performances. 12 September 2024.

**Original London Style Book Launch.** Giuseppe *U Net* Pipitone. 25 October 2024.

**21 Minutes Art Battle.** Live performance event featuring Nathan Bowen. 7 December 2024.

**Long Dark Tunnel Exhibition.** Contemporary art show featuring never-before-seen works by 10Foot, Tox, and Fume DDS. 21 March–10 April 2025.

**Beatsora x Arts ARKADE Live Rap Cypher.** 17 May 2025.

## **Invited Conferences, Workshops, and Panel Discussions**

**HAI-GEN Workshop at IUI 2023.** 4th Workshop on Human–AI Co-Creation with Generative Models. 27 March 2023.

**The Second Workshop on Intelligent and Interactive Writing Assistants.** CHI 2023. 23 April 2023.

**GenAICHI 2023.** Generative AI and HCI at CHI, Hamburg. 28 April 2023.

**AVARA 2023.** The Centre for Arts Innovation (CAI), Oulu University of Applied Sciences, Finland. 16–17 November 2023.

**The Crown Estate: Together '24.** Panel discussion. 9 May 2024.

**The Evening Standard and Samsung: The Imagination Playground.** Panel discussion. 12 September 2024.

**Creative IP Lab – UCL x CoSTAR Foresight.** Workshop. 12 June 2025.

**Digital Creativity and Placemaking.** Panel. RGS–IBG Annual International Conference, Birmingham, UK. 26–29 August 2025.

## Publications

Specific explorations are contained in Chapters 3 (*placemaking*), 4 (*writing tools*), and 5 (*human–AI co-creativity*).

Ibùkún Ọlátúnjí. **Why try to build a co-creative poetry system that makes people feel that they have ‘creative superpowers’?**. IUI Workshops, pp. 67–80. 2023.

Ibùkún Ọlátúnjí. **Interactive writing systems and why small(er) could be more beautiful.** *The Second Workshop on Intelligent and Interactive Writing Assistants*, co-located with CHI 2023, Hamburg. 2023.

Ibùkún Ọlátúnjí. **Humans feeling creative and the dangers of large language models.** GenAICHI 2023 at CHI (non-archival workshop paper). 2023.

I. Olátúnjí, M. Sheppard, A. Rahat, M. Jones, and A. Rogers. **Battle Rap as a Framework for Human–Machine Co-Creation.** In *Proceedings of the 16th International Conference on Computational Creativity (ICCC '25)*, to appear.

I. Olátúnjí and M. Sheppard. **Measuring Creativity in Co-Writing with AI: Rhyme Density and the Limits of Computational Proxies.** In *Proceedings of the 16th International Conference on Computational Creativity (ICCC '25)*, to appear.

I. Olátúnjí and M. Sheppard. **Next Token Prediction Is a Dead End for Creativity: Why It’s Impossible to Lose Yourself in the Moment.** arXiv preprint. 2025.

Ọlátúnjí, Ibùkún. **Arts ARKADE: Digital Placemaking, Urban Arts, and the Future of Public Space.** Accepted for presentation at the RGS–IBG Annual International Conference, Birmingham, UK. 2025, to appear.

Ibùkún Ọlátúnjí. **Why try to build a co-creative poetry system that makes people feel that they have ‘creative superpowers’?**. UKAIRS. 8–9 September 2025 (presented version of IUI Workshops 2023 paper).

M. Sheppard and Ibùkún Olátúnjí. **Eruption: The Concept of Feel as a Framework for Human–Machine Co-Creativity.** arXiv preprint. 2025, pending.

The outputs document the development of the research across practice, performance, and publication, establishing a body of work that supports both scholarly and creative contributions.



## Chapter 2

# Literature Review

The Literature Review spans several intersecting domains across place, poetry and generative AI. To orient the reader within this interdisciplinary space, Figure 2.1 maps the key research fields and their relationships, forming the structure for the sections that follow.

### 2.1 The Psychology of Human Creativity

Human creativity is not fully understood and has been described as being ‘a mystery’ and ‘a mysterious and complex phenomenon’ [19, 20]. Nevertheless, there is a body of research describing the concepts surrounding creativity. These can be used to provide a foundation for working definitions and models [21, 22, 23]. Boden, for instance, argues that creativity is “the ability to come up with ideas or artifacts that are new, surprising, and valuable. It enters virtually every aspect of life as an aspect of human intelligence in general” [19]. According to this definition, creativity is grounded in everyday abilities such as conceptual thinking, perception, memory, and reflective self-criticism. This conceptualisation of creativity argues that everyone is creative, to a varying degree. It is a near-universal feature of *humanness* as opposed to a quality that a subset of humans have. Adopting Boden’s definition, creativity has the elements of:

- Novelty (N): the idea or artifact is new within the domain.
- Surprise (S): The creative idea should defy expectations or offer an element of unpredictability.
- Value (V): the idea or artefact is meaningful or serves a purpose in its context.

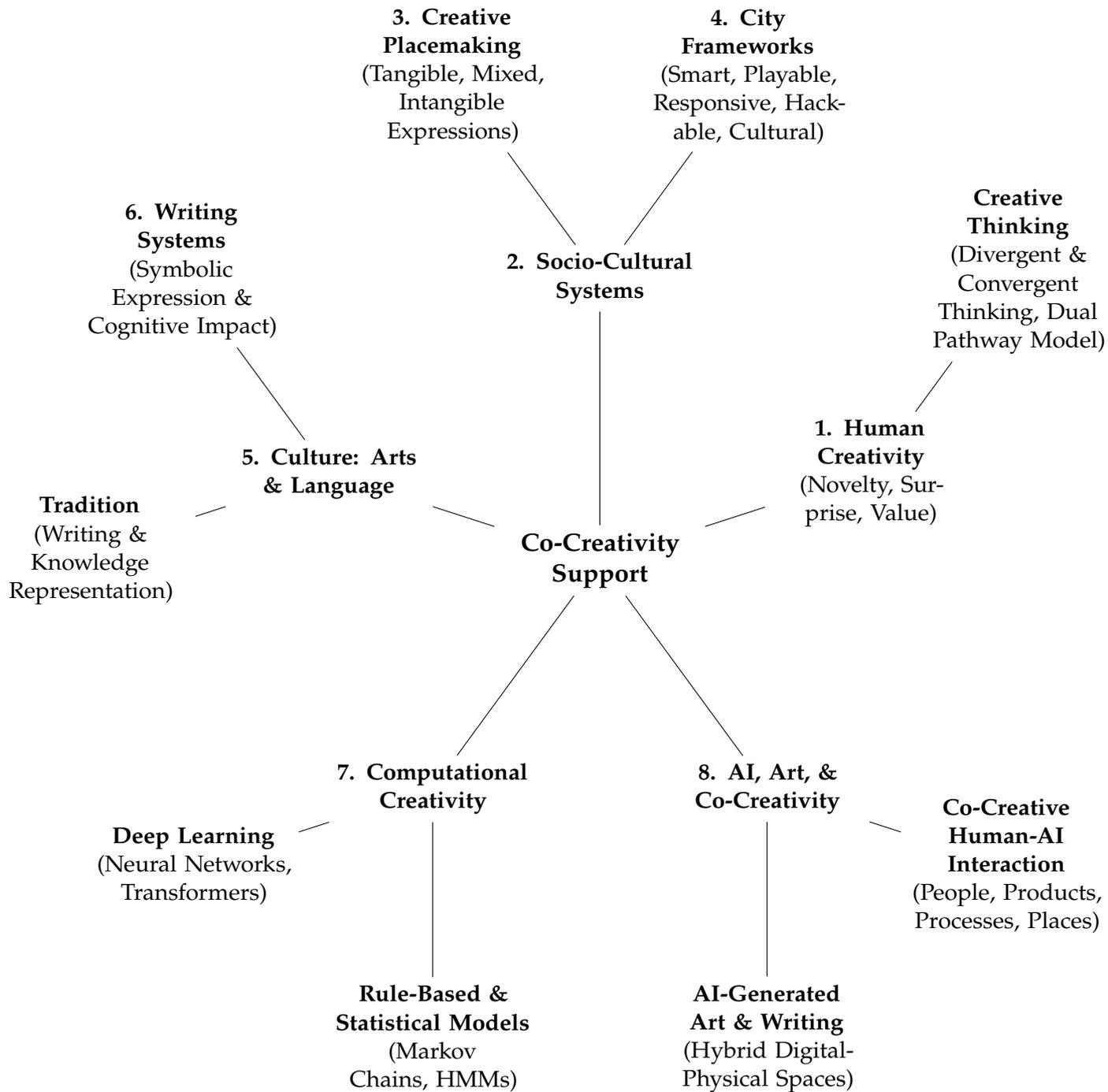


Figure 2.1: **Thematic Map of Key Literature:** A human-centred, interdisciplinary map linking human creativity, socio-cultural systems, culture and language, computational creativity, and AI-art co-creativity, with numbered topics linking to their respective sections.

## **Novelty, Surprise, and Value**

Boden (1990) distinguishes between H-creativity (“historical” creativity, ideas judged novel within a historical context) and p-creativity (“psychological” creativity, ideas that are new to the individual, even if not novel in the wider historical sense). Arguably, H-creativity is a special case of P-creativity. In social, economic and cultural terms, h-creativity is most important [24, 25, 26, 27]. However, within the psychology of creativity, p-creativity is more useful as it explores the processes of how individuals come up with ideas that are new to *them*.

There are three ways creativity is expressed through surprise (i) combinations of unfamiliar ideas ; (ii) exploration of conceptual space(s); and (iii) transformation of conceptual space(s). Creativity seen as a *conceptual space*, provides a bridge between psychology and computer science such as Natural Language Processing, Information Retrieval, and more latterly LLMs. These fields of computer science can be combined and labelled *Computational Creativity*.

Value is perhaps the most fluid of the dimensions of creativity in Boden’s model. It is necessary that some value is attributed to an idea for it to be creative. It is not clear however what kind of value - social, cultural, economic, or some other measure. Boden states that values are “difficult to recognize, more difficult to put into words, and even more difficult to state really clearly...because creativity by definition involves not only novelty but value, and because values are highly variable, it follows that many arguments about creativity are rooted in disagreements about value” [19].

## **The Psychology of Creative Thinking**

Creativity is thought to comprise two measurable cognitive ingredients that are crucial to the creative processes: *divergent* and *convergent* thinking. Divergent thinking represents a style of thinking that allows idea generation, in a context where the selection criteria are relatively vague and more than one solution is correct. Divergent thinking involves flexibility of the mind. In contrast, convergent thinking represents a style of thinking that allows finding single solutions to a well-defined problem, which requires more persistence and focus [20, 22]. It is suggested therefore, that divergent and convergent thinking are likely to serve different purposes, and to satisfy different task demands [28, 29].

## Divergent Thinking

Divergent thinking refers to the ability to generate multiple possible ideas, often in contexts where more than one solution may be correct [21, 28, 30]. A widely used measure is Guilford's Alternative Uses Task (AUT), in which participants are presented with a familiar object (e.g., a brick) and asked to produce as many different uses as possible [31].

Performance on the AUT is typically evaluated using four indicators: (1) *fluency*, the total number of ideas generated; (2) *flexibility*, the number of distinct conceptual categories represented; (3) *elaboration*, the amount of detail provided for each idea; and (4) *originality*, the rarity of responses compared to those of other participants. These measures capture different aspects of performance and vary in their theoretical transparency. Fluency and elaboration are not specific to creativity, as high scores may result from minor variations on the same idea and are sensitive to factors such as vocabulary and motivation. Originality is central to creativity, yet it is assessed relative to a given sample, meaning identical responses may be rated differently depending on the cohort's overall creativity level. Both originality and flexibility are also subject to contextual and subjective influences. Originality judgements can be shaped by cultural norms, and trends, so ideas considered novel in one setting may be commonplace in another. Flexibility, meanwhile, depends on how responses are categorised, a process that often involves subjective interpretation and may reflect prevailing conceptual frameworks rather than an objective measure of diversity [32, 33]. A recent evolution of the AUT is the Divergent Association Task (DAT) [34], which measures divergent thinking through the semantic distance between pairs of words rather than through the production of multiple ideas. The DAT was developed to address several limitations of the AUT, including its dependence on subjective scoring, sample-specific originality norms, and the time required for both task administration and evaluation. By relying on computational measures of semantic distance, the DAT provides a rapid, scalable, and theoretically grounded proxy for divergent thinking. While not a replacement for richer creative tasks, it offers an efficient baseline measure that can complement practice-based evaluations, an approach used in Study II of this thesis.

## Convergent Thinking

Convergent thinking is considered a process of generating one possible solution to a particular, well-defined problem [20, 22]. Mednick's Remote Associates Test (RAT) is commonly used in convergent thinking literature [35]. Within the RAT, participants are

presented with three unrelated words (e.g., *cocktail*, *dress*, and *birthday*) and identify the common associate (“party”). The total number of correct answers is used to evaluate their level of convergent thinking. Success within the task requires tight *top-down* constraints, as there is only one possible answer per item. Cognitively, this suggests a control state with a strong goal-directed bias toward persistence [29].

While the RAT is a core measure in convergent thinking research, it is not the only one. Other tasks explore similar cognitive processes, including colour categorisation, colour matching, word-ends, and mental arithmetic [36, 37, 38, 39].

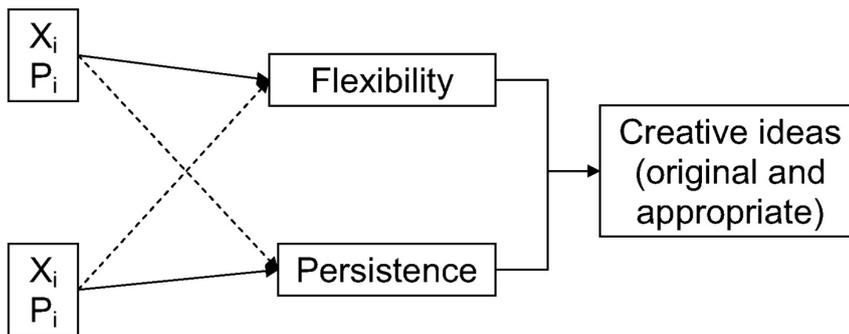
However, compared to these alternatives, the RAT is thought to involve more creativity-specific processes. This is because the RAT requires participants to hold one target word as a primary cue, and to think of other related associations. It is likely, therefore, that in contrast to what is commonly assumed, the RAT also involves aspects of divergent thinking and benefits from some degree of flexibility [40, 41]. A variation concerning the RAT was introduced by Kounios and Beeman, who distinguished between solutions found through analytical search and those found through spontaneous insight [42]. In contrast to analytical search, which can be thought of as hypothesis testing, insight solutions are experienced as sudden changes to knowledge representation leading to more ‘spontaneous’ solutions, often accompanied by a burst of emotion, including surprise [42, 43]. Given that finding solutions through insight can result from a defocused state, it is arguable that the processes underlying insight are closer to divergent thinking than analytical search [20].

### **Models of Combined Thinking**

The interplay between convergent and divergent processes suggested by RAT findings raises the question of whether these forms of thinking operate independently or in concert. The characterisations of divergent and convergent thinking, and the measures used to assess them, suggest that neither AUT nor RAT represent process-pure measures of divergent and convergent thinking or their underlying processes. In both tasks participants are required to hold one goal-related concept while switching between other possible or actual alternatives. It is likely, therefore, that the tasks require both divergent and convergent thinking, as well as both flexibility and persistence [44].

The Dual Pathway to Creativity (DPC) Model distinguishes between a flexibility route and a persistence route to creative performance, and it assumes that creative products rely on the two routes to different degrees. The model proposes that the routes collaborate

differently in divergent and convergent thinking [44]. The focus of attention varies as a function of task demands: when task ambiguity is high, attention is defocused, resulting in slower processing on the task; when task ambiguity is low, attention is focused, resulting in faster processing [45]. To create highly original ideas, flexibility is needed to switch between routes.

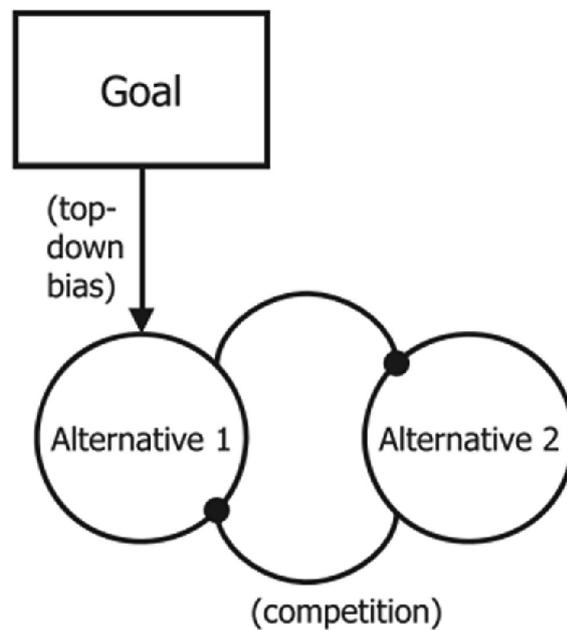


The Dual Pathway to Creativity Model.

Figure 2.2: **The dual pathway to creativity model** according to Nijstad et al. (2010). There are two pathways to achieve creativity: a flexibility and a persistence pathway, and both pathways are influenced by situational (denoted with  $X_i$ ) and dispositional (denoted with  $P_i$ ) factors. However, some situational and dispositional variables affect the flexibility pathway more strongly than the persistence pathway, and vice versa, which is indicated by solid (stronger relation) and dotted (weaker or negative relation) lines, respectively. The dual pathway to creativity model: Creative ideation as a function of flexibility and persistence, *European Review of Social Psychology*, 21:1, 34–77. *In the context of this review, the DPC clarifies how divergent thinking benefits from flexibility-oriented routes, while convergent thinking benefits from persistence-oriented routes, and how these routes may operate in combination depending on task demands.*

While the DPC model focuses on routes to creative performance, the Metacontrol State Model (MSM) offers a broader framework for understanding cognitive control in general, which can also be applied to creativity [20, 29]. The MSM describes the balance between persistence and flexibility in decision-making, characterised by strong top-down guidance by goal representations and mutual competition between alternative selection candidates [46]. The degree to which alternatives compete is determined by the present metacontrol state, which varies between persistence and flexibility. Extreme persistence consists of strong competition and strong top-down bias, whereas extreme flexibility consists of weak competition and weak top-down bias. Accordingly, convergent thinking benefits from a bias towards persistence, whereas divergent thinking benefits from a bias towards flexibility [20, 29]. A task such as the AUT would be expected to require weak top-down support, as the search criterion is vague and many answers are possible; the competition

between alternatives should therefore also be weak. This means that performance in this task strongly benefits from a control mode biased towards flexibility [29]. This is also the case for other divergent thinking tasks, such as creative story writing or thinking of unique consequences of hypothetical situations [47, 48]. The MSM can be understood as a general cognitive control framework that provides the underlying mechanism for the flexibility and persistence routes described in the DPC model, extending these concepts beyond creativity to decision-making and action control more broadly.



The Metacontrol State Model.

**Figure 2.3: Key assumptions of the Metacontrol State Model.** Metacontrol flexibility states are characterized by the strength of top-down guidance from goal representations for goal-consistent selection candidates and the strength of mutual competition for alternative selection candidates. Flexibility is implemented through weak top-down guidance and mutual competition, whereas persistence is characterized by strong top-down guidance and mutual competition. Based on Fig. 1a from Hommel (2015) *Between persistence and flexibility: The Yin and Yang of action control*. In: A.J. Elliot (ed.), *Advances in motivation science*, Vol. 2 (pp. 33–67). New York: Elsevier. *In the context of this review, the MSM aligns with the DPC in framing divergent thinking as benefiting from flexibility-biased states and convergent thinking as benefiting from persistence-biased states, highlighting creativity as the result of shifting between these modes.*

Although the DPC and MSM models differ in focus and emphasis, they are highly consistent. Empirical findings support the argument that divergent and convergent thinking are related to flexibility and persistence, respectively. Depending on the complexity and nature of the task, creatively engaging often involves switching between flexibility

and persistence [20, 39, 40, 46]. In summary, divergent thinking and insight solutions in convergent-thinking tasks benefit from metacontrol biases towards flexibility, whereas convergent, analytical thinking benefits from metacontrol biases towards persistence [20]. These models underline that creative performance emerges from dynamic shifts between flexibility and persistence. While both the DPC and MSM explain creativity in terms of cognitive control mechanisms, the following section broadens the perspective to consider creativity as a socio-cultural system, emphasising the environmental and cultural contexts in which these cognitive processes operate.

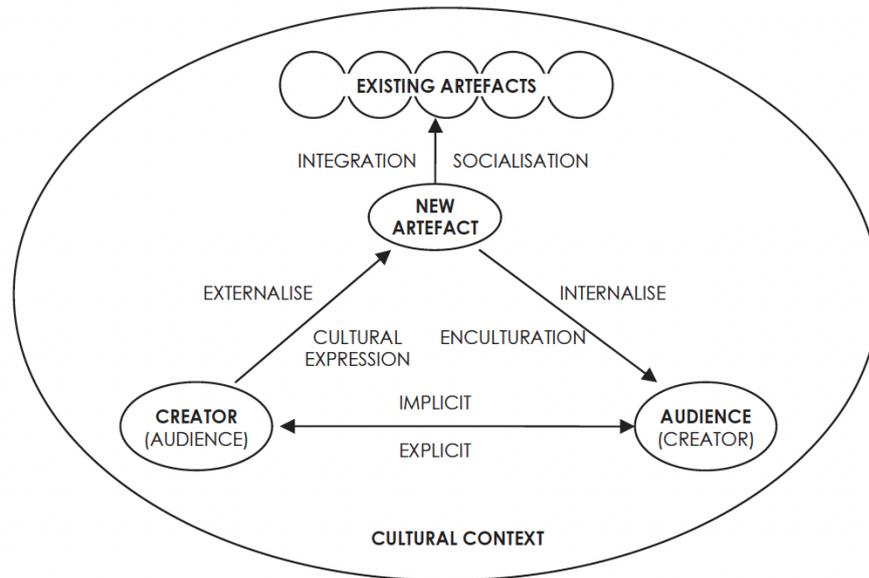
### 2.2 Creativity as Socio-Cultural System

The cognitive models discussed previously explain creative performance in terms of internal control mechanisms, specifically, the balance between flexibility and persistence. However, creativity is embedded within social, cultural, and environmental systems [12] that shape both the opportunities for and expressions of creative work.

Building on the four categories of creativity — *creative product*, *creative process*, *personal creativity*, and *environments that foster creativity* [49, 50], this section focuses on the latter, examining environments that foster creativity through the lens of cultural geography. The framework adopted here treats creativity as a complex system comprised of socio-cultural components [23, 51], in which creative expression is simultaneously an individual, social, and cultural act [51, 52].

In recent decades, creative cognition has shifted from a person-centred perspective to a socially dynamic one [53]. Within this framework, socio-cultural theories of creativity and learning have gained recognition, emphasizing that ‘culture clearly has a profound influence on the conceptualisation of creativity and on creative expression’ [54]. Creativity is socio-cultural because: a) the set of skills and types of knowledge that individual actors possess are developed through social interaction; b) creativity in itself is often the result of explicit moments of collaboration between individuals; c) creativity is largely defined by social judgement or validation; and d) creativity exists only in relation to an established ensemble of cultural norms and products that both aliment the creative process and integrate its *outcomes*.

Building on the shift towards a socially dynamic conceptions of creative cognition, knowledge can be understood not as isolated information, but as embedded within the practices of specific sociocultural communities, referred to as *communities of practice* [55].



Creativity as a socio-cultural-psychological process.

Figure 2.4: **Creativity as Socio-Cultural-Psychological process** Conception of creativity that brings together “self” and “others”, the “new” and the “old” and captures the relationships that tie them together in the form of creative activity. At the core of this diagram is the interplay between creator, creation and audience. Fig. 1 from Glăveanu (2011) *Creativity As Cultural Participation*. In: *Journal for the Theory of Social Behaviour* 41:1. pages 48-67 Oxford: Blackwell Publishing.

Learning is not merely the acquisition of facts but a process of becoming a member of these communities through progressive participation, a concept known as "legitimate peripheral participation." In this view, learning is fundamentally about identity formation, shaping individuals into particular kinds of persons within a given cultural and social context. The artifacts, symbols, language, and gestures that these communities produce are not just tools or representations; they are integral to the practice and meaning-making of the community. To grasp their cultural significance, termed *cultural transparency* one must gain access to the practices and discourses that give them meaning. Understanding is not a direct interaction between a person and an object but is mediated by the interpretative frameworks provided by their social and cultural environment. This perspective reframes knowing, learning, and intelligence not as isolated cognitive acts, but as socially situated activities embedded in lived experience. Furthermore, technology and material artifacts play a crucial role in shaping and encoding knowledge within these communities. The process of transforming understanding into artifacts and artifacts into understanding

is an ongoing exchange. However, as technological advancements have accelerated, communities of practice increasingly rely on complex systems that they do not fully comprehend, a phenomenon known as the "black-box syndrome." The growth of powerful AI such as LLMs are examples of this; it has been argued that black box systems threaten the ability of people to engage critically in an increasingly mediated and technologically driven society [55, 56, 57].

### 2.3 Creative Placemaking

The concept of *place-making* has a long history in geography and urban design research, offering foundational approaches for how space can be strategically changed [7]. The term itself can be traced back to the 1960s, when it was associated with efforts to transform public spaces and reshape the image of cities [7, 58].

*Creative placemaking* emerged more recently as a distinct policy term in the United States, popularised by Ann Markusen through her influential 2010 report for the National Endowment for the Arts. In this framing, creative placemaking

“animates. . . spaces, rejuvenates structures and streetscapes. . . and brings diverse people together to *celebrate, inspire, and be inspired*” [emphasis added] [59]. It uses arts and culture to advance quality of life within three interrelated domains: tangible, mixed, and intangible [60]. The National Consortium for Creative Placemaking (NCCP), a U.S.-based non-profit organisation, has further developed and promoted the practice, positioning it as a tool for community engagement and local economic development.

While widely embraced for its community-centred ethos, creative placemaking has also attracted critique. Scholars and practitioners have noted its potential for co-option by municipal or developer-led agendas, where it can serve as a relatively inexpensive means of urban regeneration that risks displacing the communities it seeks to empower [61, 62]. The tension between grass roots cultural development and top-down urban policy remains a central concern in the field. Placemaking in this thesis is not limited to transformation; it also encompasses the practices and policies through which cities such as Swansea and London sustain their distinctive identities. To operationalise the role of culture in placemaking, this thesis draws on the concept of the *playable city*; an approach that privileges low-barrier, temporary, and participatory interventions.

Expression Type	Definition	Key Aspects	Examples
<b>Tangible Expressions</b>	The <b>physical and material</b> features of a place	Buildings, landscapes, infrastructure, artifacts	Eiffel Tower, Great Wall, traditional houses
<b>Mixed Expressions</b>	A combination of <b>tangible and intangible</b> elements shaped by human interaction and cultural practices	Social spaces, lived experiences, evolving meanings	Street festivals, pilgrimage sites, street art
<b>Intangible Expressions</b>	The <b>symbolic, cultural, and experiential</b> aspects of a place	Traditions, stories, emotions, rituals, collective memory	Music, dance, oral history, religious rituals, social norms

Table 2.1: Summary of Expressions in Placemaking (Alan Lew, 2017)

**Tangible Expressions:** refer to the physical and material aspects of a place, including both natural landscapes and built environments. These are the visible and measurable components that define a location's structure, aesthetics, and physical presence. Key characteristics of Tangible Expressions include:

- *Physical Design and Architecture:* Buildings, streets, monuments, and infrastructure that shape the spatial layout of a place.
- *Natural Landscapes:* Mountains, rivers, forests, and beaches that contribute to the environmental identity of a location.
- *Urban Rural Structures:* Parks, plazas, historic sites, and transportation networks that organize how people move and interact.
- *Material Culture:* Objects, artifacts, and physical artworks that define the cultural identity of a place.

Examples of Tangible Expressions: The Eiffel Tower as a physical landmark of Paris; The Great Wall of China, representing cultural and historical identity; public squares and marketplaces that define the social structure of a city.

**Mixed Expressions:** exist on the continuum between Tangible and Intangible Expressions, incorporating both physical structures and the social, cultural, and symbolic meanings attached to them. These expressions are shaped by what Lew describes as ‘people practices’. These ‘people practices’ are often seen in the ‘everyday practices of people who occupy these planned places’ [63]. The key characteristics of Mixed Expressions are:

- *Tangible and Intangible:* Physical spaces gain meaning through human activity, traditions, and cultural interactions.
- *Lived and Experienced:* The way people use, interpret, and interact with a place shapes its identity.
- *Evolving Over Time:* Mixed expressions are dynamic, as people and cultures change the meanings attached to a place shift as well.

Examples of Mixed Expressions:

- *Street festivals and cultural parades.* A street (*tangible*) becomes significant when it hosts events such as The Notting Hill Carnival or Rio’s Carnival, turning it into a site of cultural expression.
- *Public parks with social significance:* A park may be a physical space, but its real value comes from how people gather, protest, or celebrate there (e.g., Trafagar Square in London).
- *Graffiti and street art:* While painted on a physical wall, street art gains significance through social and political narratives that define a place’s identity

**Intangible Expressions:** refer to the non-physical, symbolic, and experiential aspects of a place that shape its identity and meaning. Unlike Tangible Expressions, which focus on physical design and landscape, Intangible Expressions are formed through narratives, traditions, social interactions, and cultural practices. The characteristics of Intangible Expressions include the following:

- *Cultural and Symbolic Meaning:* The traditions, stories, and values associated with a place. For instance, folklore, myths, and historical narratives that give a location its identity.

- *Social and Emotional Connections*: The sense of belonging, identity, and nostalgia attached to a place. For example, the memories and collective experiences shared by a community.
- *Sensory and Experiential Qualities*: The atmosphere, mood, and emotions evoked by a space. These characteristics relate to the non-material experiences such as soundscapes, scents, and/or movement.
- *Dynamism and Evolution*: Intangible expressions change over time, and shaped by forces such as migration, technology, and generational shifts. Intangible Expressions are co-created through human activity rather than imposed through physical design.

Examples of Intangible Expressions:

- *Music, Dance, and Performance*: Cultural performances that embody local history and identity. The identity of a place shaped by traditional music and dance (e.g., Jazz in New Orleans, Tango in Buenos Aires).
- *Spiritual and Ritual Practices*: Pilgrimages, religious ceremonies, and indigenous knowledge systems tied to land and nature. Sacred spaces that hold meaning beyond their physical form.
- *Local Stories and Oral Histories*: Myths, legends, and community stories that pass down a place's significance. For example, Yoruba spiritual narratives tied to sacred forests or mountains.
- *Language and Dialects*: The way people speak, local slang, and linguistic identity that shape a region. Distinctive accents, proverbs, and linguistic traditions tied to place.
- *Social Norms and Everyday Life*: Unwritten rules that define how people behave in a space (e.g., café culture in Paris, street markets in Marrakech). Ways of gathering, greeting, and interacting in public spaces.
- *Festivals and Collective Memory*: Celebrations, protests, and communal events that define a place's identity. For example, the role of Carnival in Rio or Eid celebrations in Mecca as intangible expressions of place.

The elements of placemaking, Table 2.1 require various forms of (mainly) human creativity. Traditionally, this has been led by architects, urban designers, and artists. Over time, the creative leadership for placemaking has become more collaborative; the rationale for this is that 'creative placemaking by...one interest group is unsustainable' and 'collaboration should be shared among as many groups as possible within a community.' [60, 64].

## 2.4 Creative, Playable, and Smart Cities

Humans are motivated by pleasure, social and emotional connection, agency, and joy [65]. If cities fail to afford these experiences, they can become unwelcoming. Fun is not an optional quality when it comes to urban spaces, it is deeply necessary. The work of Jane Jacobs' and others, posit that fun is necessary for urban spaces. They argue that cities are made up of moment-to- moment interactions between people, and as such they should cater to socio-emotional needs [6, 66, 67].

Play is a fundamental human need that enriches the experience of the world [11, 67, 68, 69]; it is socio-emotionally productive even if it does not produce materially productive outcomes [69, 70, 71]. Play is also a way to bring creativity to the urban space. Inspired by existing studies on how play and games can support day-to-day *ludic* experiences, researchers argue for the potential of play in urban design. In particular, there are three qualities of play that are considered socio-emotionally important [6, 72].

1. Play brings joy to situations that might otherwise be mundane: it generates positive emotions because it speaks to the inner desire for joy and stimulation [69].
2. Play can help people to experience feelings of agency and act in ways that feel meaningful.
3. Play empowers people to act upon the world around them, and cultivates creative ideation [71].

HCI research has explored the intersection of play and smart cities, through the 'playable city' [6, 73, 74]. The term *Playable City* originated at the Pervasive Media Studio in Bristol initiated through an annual award to develop projects that playfully situate people in relation to smart cities infrastructure [11, 75]. Building on this, the Playable City methodology re-appropriates urban infrastructure to make "playable places... encouraging citizens to reimagine and reconfigure their city's services, places and stories" [76]. The aim of this approach to placemaking is to find new ways of connecting city dwellers both

with each other, and with the urban environments in which they live [73, 76]. Playful street furniture, for example, aims to engage communities, stimulate public behaviour change, and demonstrate the advantages of playable cities [74, 77]. More importantly, urban interventions form a stronger connection between people and place, with greater emphasis placed on lived experience through active participation rather than the mere provision of utility. However, in designing playful street furniture, the subjectivity of human perceptions and situated conditions of the urban space heavily influence the process of interaction, and resulting user experience. Play is typically viewed as a childish activity. It is often deemed as a luxury to adults, and as a result non-essential [78]. Given this, designing compelling invitations to play can be just as challenging as ensuring that play aligns with its cultural and environmental context [79]. Donoff categorized playable experiences by play type, design feature, and implementation styles. The resulting typology (table 2.2) represents a framework for appreciation of ways to heighten urban play opportunities for adults (as well as children), while not privileging or prioritizing one type over another [73].

<b>Play type</b>	<b>Design</b>	<b>Approach to implementation</b>
Acting contrary to social convention	Attractive colours	Busy location
Adaptation of a well-known game type	Bicycle infrastructure	Guerilla installation
Auditory stimulation	Biomimicry	Infill in underused space
Chance	Imageability	Opportunity to people watch
Cognitive games	Multigenerational appeal	Route choice or environmental mastery
Competition	Pedestrian lighting	Sense of belonging and community connection
Cooperation	Scientific design	Temporary, pop-up or seasonal
Creative play	Unique paths	Uses existing infrastructure
Opportunity to escape	Use of common, everyday materials or objects	Use of social media or pop culture
Opportunity to increase social contact	Use of props to alter movement	

Table 2.2: A Typology of Urban Play (adapted from Donoff, 2014, p. 138 [73]).

Table 2.2 presents a typology of urban play, adapted from Donoff, categorising different forms of play based on their design characteristics and approaches to implementation [73]. The typology identifies various play types, including spontaneous actions that challenge social conventions, structured adaptations of existing games, sensory experiences, cognitive engagement, and competitive dynamics. Each type is linked to specific design elements, such as the use of colour, infrastructure modifications (e.g., bicycle paths and pedestrian lighting), biomimicry, and the incorporation of everyday materials. The

implementation strategies highlight spatial and contextual considerations, including location-based factors (e.g., busy areas, underused spaces), behavioural affordances (e.g., temporary installations, social media integration), and environmental interactions (e.g., imageability, water features).

In relation to playable cities, this framework underscores how urban environments can be designed to encourage creative expression, agency, and social connectivity. However, the notion of ‘participation’ in such contexts warrants critical examination. Participation is not equally accessible or appealing to all groups, playful design can unintentionally privilege certain demographics (e.g., younger, mobile, tech-literate residents) while excluding others due to cultural norms, mobility constraints, or lack of interest [80]. While interventions are often framed as making cities more ‘dynamic’ or ‘adaptive’, these effects may be temporary and geographically limited, with installations functioning more as spectacles than as catalysts for lasting community transformation [81]. The functions of participation also vary: some projects aim to foster genuine community building, while others primarily serve commercial, touristic, or branding objectives [82]. This raises the question of whether playful urban interventions produce deep social engagement or aestheticised, surface-level experiences that align with broader urban marketing strategies [83]. In creative terms, the typology highlights the potential of urban play to stimulate novel interactions with the built environment. By enabling improvisation, experimentation, and reinterpretation of space, play can contribute to a more adaptive urban landscape. Yet the incorporation of non-traditional design elements, such as guerrilla installations and temporary pop-ups, also raises questions about sustainability, inclusivity, and who benefits from the disruption of conventional spatial hierarchies. These tensions complicate the narrative of playful cities as universally positive, situating urban play within wider economic, cultural, and political dynamics. The Stop, Smile, Stroll intervention exemplifies the principles outlined in the typology by integrating playful elements into everyday urban experiences. In terms of play type, it incorporates elements of acting contrary to social convention by inviting passers by to pause and engage in lighthearted actions that contrast with the typical rhythms of urban mobility. It also involves aspects of auditory stimulation, chance encounters, and cognitive engagement, as participants encounter unexpected prompts encouraging movement and interaction with their environment.

From a design perspective, Stop, Smile, Stroll uses attractive visual cues, pedestrian-friendly infrastructure, and imageability, a concept that enhances wayfinding and

emotional connections to space. The reuse of common materials, temporary signage, and interactive prompts aligns with the typology's emphasis on scientific design and unique paths, reinforcing urban spaces as sites for creative exploration. Augmented reality overlays, paired with short music snippets, enable participants to capture and share moments of mood and expression, generating an evolving emotional 'map' of the location [67, 84, 85]. While the intervention demonstrates how playful design can momentarily disrupt routine and create shared experiences, its impact is necessarily brief and spatially constrained. Participation may be self-selecting, shaped by social comfort in public spaces, mobility, or familiarity with digital tools [80]. Although it fosters micro-interactions, these connections are typically transient, and their contribution to sustained community building remains uncertain. As with other temporary interventions, the project risks serving as an aesthetic or branding device as much as, if not more than, a vehicle for deep social change [81, 82]. *Stop, Smile, Stroll* therefore illustrates both the potential and the limitations of small-scale playful interventions. It supports curiosity, agency, and momentary connection, but also embodies the tensions identified in the typology: inclusivity, temporality, and the interplay between genuine social engagement and the political economy of urban placemaking. Following *Stop, Smile, Stroll*, this thesis treats AI-enabled interventions, such as those trialled via *Arts ARKADE*, as cultural probes: situated deployments used to assess how generative tools are taken up within local creative ecologies. The playful city thus offers a lens for understanding how computational tools interact with public cultural practices and spatial constraints



Stop, Smile, Stroll

Figure 2.5: **Stop Smile Stroll** breaks the mundane routine of commuting, providing a brief sense of excitement and connection between strangers. Stop Smile Stroll is playful intervention at pedestrian crossings that acts as a 'mood detector and reactor'. The experience invites passers-by to show how they're feeling through their expressions, and captures these before revealing the collective 'mood of the moment'. The public art intervention was originally commissioned in 2018 as part of the Playable City Award by Watershed Arts Trust.

### Smart Cities and Beyond

Competing understandings of cities have co-existed and co-evolved over the past decade [68, 78]. Some of the frameworks developed to respond to urban design challenges include: *smart cities*, *playable cities*, *responsive cities*, *algorithmic cities*, *hackable cities*, and *creative cities*. Technology innovation is the driving force behind the various frameworks.

City Type	Key Features	Focus
<b>Smart City</b>	Data-driven governance, IoT, AI, real-time optimization	Efficiency, automation, and predictive analytics; risk of surveillance and corporate control
<b>Playable City</b>	Interactive, gamified, and engaging public spaces	Encourages creativity and social interaction; prioritizes emotional and participatory experiences over efficiency
<b>Responsive City</b>	Real-time adaptation based on human and environmental feedback	Dynamic urban infrastructure that adjusts to citizen behavior and environmental conditions
<b>Algorithmic City</b>	AI-driven decision-making, predictive models, and computational logic	Optimization of urban life but risks bias, opacity, and lack of citizen control over decision processes
<b>Hackable City</b>	Open-source, citizen-led innovation, DIY urbanism	Emphasizes adaptability and grassroots urban activism; challenges top-down governance
<b>Creative City</b>	Focus on cultural production, artistic expression, and creative industries	Encourages economic and cultural innovation but may risk gentrification and exclusion of marginalized communities

Table 2.3: Overview of City Frameworks

The notion of a *creative city* investigates how citizens think and act creatively, positioning creativity as a strategic driver for urban planning and development [86, 87]. Popularised by Charles Landry and later reinforced by Richard Florida's *creative class* thesis, this discourse has been influential globally, shaping municipal policies from the late 1990s onwards. A creative city promotes creativity among its citizens, aims to provide fulfilling lived experiences, and aspires to leverage cultural and creative capital for urban regeneration [6, 88, 89]. However, Florida's work in particular has been heavily critiqued for overstating the economic benefits of attracting a 'creative class' while neglecting the gentrification, exclusion, and inequality such strategies can exacerbate [61, 62]. Related terminology, such as the *cultural city*, is sometimes used interchangeably or as an extension of the creative city framework, though it may place stronger emphasis on heritage, the arts, and cultural policy.

In contrast, the *smart city* emerged in the late 2000s as a dominant framing of urban futures. A smart city is generally defined as one that uses digital technologies, data-driven systems, and networked infrastructures to enhance efficiency, governance, and quality of life [90]. While creative and cultural city discourses captured global policy attention in the 1990s and early 2000s, the smart city rose to prominence in subsequent years, not necessarily because it displaced the earlier frameworks, but because it aligned with broader shifts toward digitalisation, urban informatics, and technology-led development. As Nijholt observes, "every city wants to be a smart city" [90], a sentiment exemplified by initiatives such as 'Smarter London Together — The Mayor's roadmap to transform London into the smartest city in the world' [91].

The smart city paradigm, however, has been questioned for its technological determinism. Researchers argue that if technology is positioned as the primary determinant of urban improvement, social and cultural value can be eroded [78, 90, 92]. Mattern's provocation, *A City Is Not a Computer* argues that "this seems an obvious truth, but we need to say it loud and clear. Urban intelligence is more than information processing" [93]. Mattern captures a growing concern: while promising efficiency and improved governance through data-driven technologies, the smart city often reduces the complexity of urban life to computational problems to be optimised. This reductionist lens risks overlooking cities as rich, social, and historically embedded environments [78, 79, 93]. Overemphasis on digital infrastructure, algorithmic governance, and sensor-based monitoring can obscure the messy, unpredictable, and inherently human dimensions of city life. While smart technologies can enhance certain efficiencies, they risk quantifying social interaction,

community resilience, and cultural expression in ways that prioritise optimisation over lived experience. Moreover, reliance on big data and predictive algorithms introduces biases that may reproduce or exacerbate existing inequalities. Rather than reducing cities to computational models, many researchers advocate for holistic, participatory, and context-aware approaches to urban planning that foreground local knowledge and social infrastructure [77, 90]. The resulting vision is for cities that are adaptable, inclusive, and human-centred, yet also able to leverage the benefits of appropriate technologies.

The first generation of smart cities were largely focused on infrastructure, efficiency, and data-driven governance, leveraging digital technologies to optimize urban management [94]. This approach often resulted in technocratic models that prioritized automation and surveillance while neglecting broader social and cultural concerns [93]. In recent decades, smart cities have evolved to a second and third generation [94]. Second-generation smart cities began incorporating elements of citizen participation and social innovation. These cities moved beyond mere efficiency and embraced human-centred urbanism, integrating public engagement, open data, and digital inclusion as key principles [95]. The emergence of third-generation smart cities marks a shift toward creativity, playability, and cultural production. Concepts such as Playable Cities and Creative Cities integrate urban spaces with artistic expression, interactive design, and storytelling [89]. The third-generation smart city models recognize that cities as cultural ecosystems where human experiences, emotions, and social interactions play a crucial role [96]. Building on this evolution, the Cultural City emerges as a more comprehensive and holistic framework. Unlike previous models that primarily focused on digital infrastructure or civic participation; the Cultural City positions arts, language, and cultural production as central to urban development. The Cultural City model aligns technological progress with artistic creativity, reinforcing the idea that cities thrive not just through efficiency but through the stories, symbols, and shared experiences that define them. As Florida argues, “technology alone isn’t enough”; long-term economic growth depends on “a much more basic factor: human creativity.” The places that succeed are those that harness creativity in all its forms — technological, economic, artistic, aesthetic, civic, and political — and “when all those forms of creativity come together in a place, that place begins to grow propulsively” [97].

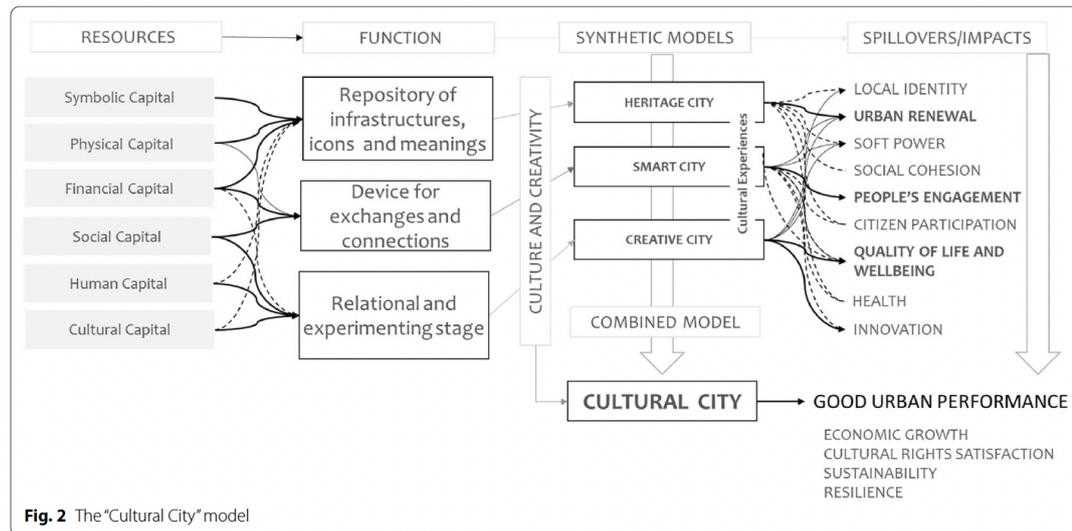


Figure 2.6: The integration of the dimensions of the Heritage City, the Creative City, and the Smart City in an enabling context is the core proposal of the Cultural City (Rausell-Köster et al., 2022).

Rausell-Köster et al. propose the *Cultural City* as a holistic framework that explains urban success through the integration of cultural experiences. They emphasize the role of culture as a driving force for urban development. Culture is defined in terms of the accumulation of cultural experiences that take place within a given territory over time, and that generate value in multiple ways. The model synthesizes key elements from alternative urban models: *The Heritage City*, *The Smart City*, and *The Creative City* [98].

*The Heritage City* encompasses both tangible urban assets (streets, buildings, and monuments) and the intangible narratives and meanings associated with them. Beyond physical structures, the city serves as a container of cultural significance, where historical and social narratives shape identity and place-making. The narratives extend beyond branding, influencing socio-cultural and economic processes. In many cases, the intangible aspects of heritage cities contribute as much as, or more than, material assets in generating collective value and a shared sense of place [63, 98]

*The Smart City* leverages digital technology, data analytics, and networked infrastructure to enhance urban efficiency, optimize city functions, and improve residents' quality of life. It utilizes real-time data, automation, and interconnected systems to manages key

resources (e.g., transportation, energy, and public services). A key characteristic of the smart city is its ability to concentrate human capital, and offer cultural experiences through density, agglomeration, and idea exchange. This enables spontaneous interactions, cross-disciplinary collaboration, and the emergence of new opportunities that might not arise in less connected environments.

*The Creative City* emphasizes the attractiveness and competitiveness of urban environments driven by cognitive and symbolic elements, where creativity fuels market, aesthetic, and social innovation. A key indicator of a Creative City is the availability of professional opportunities within the creative sector. However, criticism of the Creative City include (i) trivialization of symbolic meaning and the commodification of cultural experiences, even those with significant social value and (ii) potential for social polarization, as the dominance of the creative class in certain urban spaces contributes to gentrification and social segmentation. Despite these challenges, the Creative City remains a desirable model, catering to an increasingly educated and culturally engaged population seeking enriched expressive, social, and professional experiences [86].

The Cultural City is interpreted through multiple perspectives (i) a repository of resources (*The Heritage City*); (ii) a connective interface (*The Smart City*); and (iii) the setting for citizens' life and social and professional experiences (*The Creative City*). Each perspective incorporates culture in a different way, enabling different models of value creation and processes of value (re)production. In each of the urban models, production processes combine symbolic, physical, financial, social, human and cultural capital in multiple ways. This is supported by urban strategies that provide cultural experiences; the result is that culture, in its different dimensions, becomes the foundation for urban development [64, 97, 98].

### **2.5 Culture: Arts and Language**

The Cultural City serves as a bridge between urban development and the broader cultural landscape, emphasizing the importance of arts and language in shaping human experience. Culture is not only embedded in physical spaces but is also constructed and expressed through artistic and linguistic practices that define collective identity and creativity. The role of arts and language in shaping cultural identity is central; both act as mediums

for storytelling, historical continuity, and human expression. Cities, as sites of cultural production, enable the development and dissemination of artistic and linguistic forms that contribute to collective meaning-making and innovation [98]. As cities evolve, the integration of cultural production into urban life becomes increasingly vital, with writing and storytelling emerging as key mediums through which culture is preserved, transformed, and reimagined. The resulting vision is for cities that are adaptable, inclusive, and human-centred, yet also able to leverage the benefits of appropriate technologies. Within this trajectory, cultural expression becomes a natural bridge to the creative practices examined in this thesis, particularly those grounded in performance, language, and urban participation. These debates also intersect with the playable city, where urban design incorporates playful, participatory elements. Like creative and smart city strategies, such initiatives need critical scrutiny for inclusivity, longevity, and the political and economic agendas behind them.

### **Oral Tradition and the Evolution of Writing**

This section situates the research within a longer history of communication shifts, from oral storytelling to symbolic systems and the invention of the alphabet. While writing has transformed knowledge storage and dissemination, speech remains the primary means by which most people communicate. While some languages are both written and spoken, most of the world's languages exist only in spoken form, making speech the primary mode of human communication [99]. This claim supports the argument that oral communication, rooted in the oral tradition, often conveys layers of information, including rhythm, tone, and emotional nuance, that text alone cannot fully capture. Throughout history, societies have preserved and shared knowledge through oral traditions, symbolic forms, and written language. These forms of expression are deeply interconnected, shaping how societies create, store, and share meaning across generations. As Courlander observes,

“The oral tradition . . . reflects ideas, themes, superstitions and truths . . . A village may reshape, to its own liking, a widespread tale. A narrator may embellish, recast and refine stories known elsewhere . . . or compose new narratives out of day-to-day living. The process of creating literature goes on side by side with the process of preserving what is generalised and old . . . for the sake of convenience we call folklore, but what in reality is nothing less than

an enormous residue of human experience and a treasury of social values and literary creations” (pp. 1–8) [100].

Before the advent of writing, oral storytelling served as the primary method for preserving and transmitting collective knowledge. Folklore, myths, and communal narratives were continuously reshaped to reflect local values, experiences, and beliefs. This adaptive process ensured that cultural memory remained relevant to each generation while maintaining continuity with the past. The fluidity of oral traditions also connects to the way symbols and religious beliefs developed over time. The diversity of gods, spirits, and symbolic systems across cultures suggests an underlying human need to encode meaning into recognizable forms [101, 102]. Symbols, much like oral traditions, serve as vehicles for transmitting shared values and worldviews.

*Cultural Transmission.* The idea of cultural transmission extends beyond spoken and religious traditions to a broader concept introduced by Richard Dawkins: the meme. Just as genes carry biological traits across generations, memes whether ideas, motifs, or symbols are transmitted through teaching, imitation, and shared cultural practices. If widely adopted, memes can persist beyond individual societies, forming a collective cultural memory that outlives its originators [103, 104]. The concept highlights how oral, religious, and written traditions function as interconnected systems for encoding and transmitting human knowledge.

“What made the alphabet so revolutionary was the ease with which people could learn to use it...the scribes of Egypt and Mesopotamia guarded the secrets of the written word. One who was literate had power over those who were not. . . the alphabet ended the hegemony of the literate elite. Instead of a complex syllabary of six hundred uniform characters, or six thousand hieroglyphs...an alphabet contained twenty-odd letters” [105].

*The alphabet as catalyst for knowledge democratization.* While oral traditions and symbolic representations sustained cultures for millennia, the development of writing systems introduced a new paradigm for storing and disseminating information. The invention of the alphabet was particularly revolutionary, as it significantly lowered the barriers to literacy. Unlike the intricate logographic systems of Mesopotamia and Egypt, which

were controlled by a literate elite, alphabetic writing reduced language to a simple set of symbols, making it easier to learn and use. This shift disrupted the exclusivity of written knowledge, enabling broader participation in cultural and intellectual life. The alphabet did more than democratize literacy though; it transformed the way knowledge was structured and stored. Unlike pictographic writing, which visually represented objects and concepts, alphabetic scripts relied on abstract symbols that carried no inherent meaning until arranged in a specific order. This abstraction not only allowed for more flexible and expressive language use but also encouraged systematic thought and classification. The ability to efficiently store and retrieve information led to the development of organized knowledge systems, laying the foundation for philosophical inquiry, scientific reasoning, and historical record-keeping [105, 106].

*The evolution of cultural memory.* Despite its revolutionary impact, early writing was not initially a comprehensive means of recording spoken language [107]. The earliest Sumerian scripts functioned as partial writing systems, capable of documenting only specific types of information, such as economic transactions and administrative records [105, 106]. Over time, writing systems evolved into full scripts, that could represent language more completely, enabling the recording of complex narratives, abstract ideas, and personal expression [108, 109]. This transition marked the final step in the transformation of human communication, from oral tradition to symbolic representation to fully developed written systems [109]. The progression from oral storytelling to symbolic systems to full writing illustrates a continuous refinement of how humans encode and transmit knowledge. Oral traditions ensured cultural continuity, symbolic representations enriched religious and artistic expression, and writing systems democratized knowledge and enabled complex intellectual traditions [108, 109]. The concept of the meme ties these developments together, showing how cultural information - whether spoken, written, or visual, persists across generations, shaping human thought and creativity [103, 104]. Language, writing, and symbolic representation are not isolated phenomena but deeply interconnected tools for preserving human experience. The invention of the alphabet, and the evolution of writing, did more than expand access to knowledge; they also redefined how societies store, retrieve, and engage with their collective memory [106, 107, 109]. This evolution of language and writing also marks the foundation for artistic expression in verse and song. As oral forms adapted to written media, the rhythmic, mnemonic, and performance

## *2. Literature Review*

---

aspects of language persisted, particularly in traditions where sound, repetition, and structure carried both meaning and emotion. Poetry and rap exemplify this continuity: each transforms linguistic pattern into expressive rhythm, uniting memory, embodiment, and interpretation. The following section examines how rhythm, rhyme and cadence operate as structural features linking poetic craft to musical performance.

## 2.6 Poetry, Hip Hop, and Lyrical Expression

Poetry has long been recognised as a demanding creative form that requires rich vocabulary, phonemic awareness, and world knowledge <sup>1</sup>[114, 115, 34]. Its compression of meaning, layered use of metaphor, rhythm, and structure make it well suited for examining how computational systems shape creativity [9, 116]. Poetry has been described as “language with a shape” (p. 8) [10], as well as “a contract between writer and reader to explore the ambiguity of language” [110], underscoring its dual function as both constrained and interpretive practice. These same qualities, brevity, rhythm, and emotional charge, align poetry with contemporary digital communication formats such as social media and messaging platforms [117, 118], reinforcing its relevance in the present era. Rap continues this poetic lineage in combining storytelling with metre, rhyme, and rhythm, and demanding high levels of lyrical skill and cultural fluency [9, 119]. While Chapter 4 focuses primarily on written poetry, rap is considered as a contemporary extension of poetic expression, and its role in AI-assisted co-creativity is explored more fully in Chapter 5.

Spoken word poetry is an art form concerned with *oracy* [120]. It is part of the *oral tradition* which in pre-literate societies included oral poems, chants, songs, riddles, proverbs, as well as nursery rhymes, genealogies, incantations, and funeral rites. As Bearder states, “poets of oral cultures are bearers of knowledge (w)rapped in rhyme, story, and song ” [120]. The oral tradition, as well as being extensive, is highly sophisticated. As Zumthor argues, “oral poetry generally has more rules, and more complex ones than the written form.” [121]. Rap can be situated as an evolution of the poetic tradition, a view supported by prominent poets and MCs [119, 122, 123]. Describing the evolution of poetry in material and technical terms, Schmidt suggests that “poets of the fourteenth century, dreaming of their work passing from hand to hand, have a difference sense of their destiny from poets hammering away at word processors or *rapping under strobes*” [emphasis added] (p. 8) [10]. Bradley makes the more direct claim that “ every rap song is a poem waiting to be performed ” (p. xi) [9]. Rap is a part of the broader cultural movement known as hip-hop. Hip-hop traditions include breakdancing, beatboxing, DJing, and music [124]. As music, hip hop has emerged as one of the most popular genres globally [125, 126]. Hip-hop music has two distinct components: the musical beats

---

<sup>1</sup>Poetry is among the oldest known literary forms, with examples dating back over 4,000 years [110, 111]. In ancient Greece, it was prescribed alongside medicine for its therapeutic qualities [112, 113], highlighting its longstanding cultural role in both personal expression and collective healing.

created by DJs, and the rapped vocals performed by MCs. Rap contains evolving lexicons based around (i) numerous genres; and, (ii) vocal delivery based around syncopation, pitch, intonation, and cadence [127, 128, 129]. Additionally, rap’s short, well documented history allows observation of the genre from its genesis to the present [127, 129]. Rap’s key components are presented in Table 2.4.

### Poetry, Song and Hip Hop: A Shared History

Poetry and rap are both examples of heightened speech. In heightened speech, the structure of speech is elaborated or exaggerated to create an emotional or aesthetic impact. Words, rhythm, pitch and timbre all contribute to speech heightening. In ancient Greece poems, through their *musical performance*, were a means for a permanent recreation of a cultural memory, between myth and history. [130]. During this period, lyrical poetry was referred to as *ta mele*, meaning “poems to be sung.” [9] Greek song (melic, iambic and elegiac poetry) between the seventh to fifth centuries gradually saw a conversion of primarily *musical performances* into written texts lacking any form of musical notation. [emphasis added] [131]. Other cultures, such as the Yorùbá or Hausa of Nigeria, use identical words for poetry and song; *orin* in Yorùbá and *wak’a* in Hausa [132, 133, 134]

Component	Linguistic Features	Musical Features
Content	Storytelling, metaphor, word-play	Phonetic manipulation, assonance, consonance
Flow	Rhyme schemes, syntactic complexity	Rhythm, tempo, syncopation
Delivery	Prosody, emphasis, vocal texture	Timbre, breath control, cadence

Table 2.4: Core structural and expressive components relevant to both poetic and musical analysis.

### Differences between Poetry and Rap

This section also provides the stylistic grounding for the rhyme-focused analyses undertaken in Studies II and III, where rhymed verse becomes the primary domain for evaluating human–AI creative interactions. The concept of *flow* is central to understanding rap as both a linguistic and musical form. The use of flow here is separate to that by

Csikszentmihalyi<sup>2</sup> which refers to a mental state. The rhythm of rap flow differs from rhythm in poetry as rap is set against music; the music forms the metrical hierarchy. In performance terms, flow includes the rapper's rhythmic and articulatory style, the way syllables are timed, stressed, and delivered in synchrony with a beat. It encompasses cadence, rhyme placement, line breaks, and the subtle variations in emphasis and tempo that give a performance its signature *feel*. Unlike traditional poetry, where rhythm is often metrical and abstracted from performance, rap flow is inherently tied to time, sound, and embodied delivery. Rhythm in rap flow is largely musical in nature and can therefore be conceptualised in frameworks of traditional music theory to some extent (e.g., metre, syncopation, rhythmic motives). There are complicating factors however such as; (i) natural rhythms of speech influence the rhythms of rap, and (ii) rap contains *micro-timings* [135] and other rhythmic nuances not captured in traditional music notation.

In simple terms it has been suggested that, " the flow of a hip-hop song is...the rhythms and rhymes it contains. " [119] Music and literary scholars however, consider flow a more complex topic. To date, there is not a widely accepted method to describes rap flow. Notable attempts to create methods (and tools to implement them) include work by Adams [136], Condit-Schulz [127] and Ohriner [137]. Of these, Ohriner's approach is perhaps the most comprehensive and articulated in the book, *Flow: The Rhythmic Voice in Rap Music*. Ohriner articulates the complexity of flow as follows (p. 3-5) [137].

Scholars differ mainly on the proper scope a definition of flow should address. What is flow a feature of? A line? A verse? An artist? An era? A region? What musical parameters are involved? Emcees are said to have their own distinctive flow, each different from another but applicable to many of an artist's verses. And emcees are often accused of stealing or "biting" the flow of another emcee, so "flow" must have some reality beyond the rhythm of the words in a particular verse. The broad view, that flow is made up of all the rhythmic, articulative, and prosodic features of a verse, is equally debilitating. Under this view, are there any aspects of rapping that do not pertain to flow? Is "the flow" any different from "the music"? In sum, how can one represent the flow of a verse such that one can compare it to that of another verse by the

---

<sup>2</sup>Though separate concepts it is likely that MCs are in a flow state [23] when freestyling as confirmed by Liu et al [115]

same artist, in the same subgenre, or originating in the same geographical or historical locale?

A high level model by Ohriner articulates the relationships between the various aspects of flow (e.g., text, rhythm, rhyme). The *global features of flow* are quantitative descriptors that summarize aspects of an artist's or verse's delivery as single numerical values. Although operationally beyond the scope of this study, the features are useful with respect to future work. Further detail is in Appendix Section D

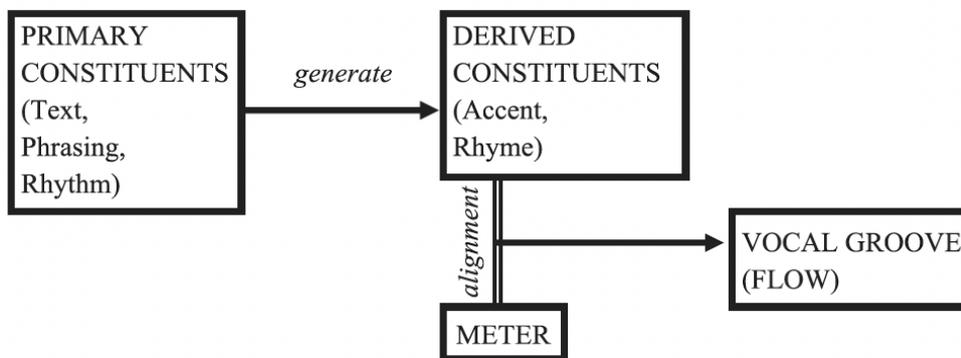


Figure 2.7: Study III: Ohringer's Model of Flow: A schematic illustrating how *primary constituents* (text, phrasing, rhythm) generate *derived constituents* (accent, rhyme), which align with meter to produce the emergent quality of *vocal groove* or flow.

The model contains *primary constituents* (e.g., words, phrasing, and rhythm) that generate other *derived constituents*. The derived constituents (e.g., rhyme, accent) are phenomena that are less precise in definitional and measurement terms.

## Cadence

Cadence, the rhythmic structure of syllables and stress, is essential to both spoken word poetry and rap. Flow combines cadence with rhyme and timing to form a rapper's stylistic fingerprint. In musical terms, the majority of rap can be counted in  $\frac{4}{4}$  time with MCs typically lagging a beat by between 40-180 milliseconds [127, 138]. Another feature of speech rhythm that influences rap flow is speed. Rap is generally faster than other forms of song with few sustained syllables or rests of significant duration. The vocal delivery speed of rap has system design implications with respect to Automatic Speech Recognition

(ASR) and Word Error Rate (WER). These are explored further in Section 5.2. Although the terms *flow* and *cadence* are often used interchangeably, they carry different emphases: MCs typically invoke flow as a holistic quality of delivery, whereas poets frame cadence in terms of rhythm and phrasing. In this study, cadence is adopted as a limited proxy to partially operationalise the broader and less precisely defined notion of flow, with recognition of its caveats. For clarity, the terms are used interchangeably throughout (for example, in questionnaires), even though their conceptual distinctions are acknowledged.

## Rhyme

Rhyme is a *perceptual phenomenon* which is evoked by *phonemic parallelism* [emphasis added] [127]. Due to their special phenomenological features, rhymed syllables articulate the highest rhythmic layer in rap flow. Unstressed syllables form the lowest layer.

Work by Hirjee and Brown [139] proposed metrics such as rhyme density and internal rhyme frequency as ways to consider rhyme patterns and perceived lyrical complexity. While rhyme and flow are distinct, they are deeply interdependent: flow determines where and how rhymes fall within a beat, while rhyme structures often reinforce or disrupt rhythmic regularity. Within a study of rap songs Condit-Schulz found variance between songs is greater than variance between MCs with respect to rhyme density and rhythm [140]. Hirjee and Brown [139] found that MCs can be accurately distinguished by based their use of rhyme and other vocal features. The interpretation of these studies is that the rhyme aspect of rap flow is relatively consistent. Flow is seen as an interaction between several components: rhyme schemes, rhythmic density, syntactic phrasing, and beat alignment. These elements combine to form what Condit-Schultz [127] refers to as a rapper's "rhythmic fingerprint." Flow is also dynamic, in that it shifts in response to instrumental changes, performance context, or adversarial constraints (as in battle rap). At a high-level flow is marked by syllable placement, controlled syncopation, and consistent alignment with musical bars. Appendix D contains a basic worked example of flow taken from *How to Rap* and an online lyric education module based on its insights [119, 141].

The *McFlow* system introduced by Condit-Schultz [127] provides one of the first large-scale, data-driven analyses of flow using timestamped syllable alignments. *McFlow* enables statistical comparison of rhythmic density, syllabic placement, and syntactic phrasing across a corpus of recorded rap performances. It shows how different artists deploy rhythmic structures over time, offering a foundation for computational modelling of stylistic variation. Table 2.5 provides a brief summary of approaches to modelling flow. Appendix D.1 contains a more detailed evaluation.

Feature	Adams’s Model	Condit-Schultz’s Model
<b>Rhythm</b>	Focuses on alignment of syllables with metrical structure. Analyses rhythmic density (syllables per beat) and placement of stressed syllables. Highlights syncopation and flow consistency across measures.	Uses timestamped syllables to measure rhythmic density (e.g., 4.5 syllables/sec). Captures onset timing and compares rhythmic patterns statistically across a large corpus.
<b>Rhyme</b>	Rhyme is a central structuring device. Emphasizes placement of rhymes at bar ends and their role in defining phrasing. Also considers multisyllabic and internal rhymes.	Quantifies rhyme frequency and placement. Measures rhyme density, showing that end rhymes on beat 4 are most common. Focuses on corpus-level patterns of rhyme.
<b>Syntax</b>	Analyses alignment between syntactic and metrical units. Considers enjambment (overflow of syntactic units across bars) as a stylistic marker.	Aligns lyrics to bar divisions and examines frequency of phrase boundaries matching bar lines. Models phrasing statistically based on corpus transcriptions.
<b>Articulation</b>	Includes qualitative features like staccato/legato phrasing, consonantal clarity, and expressive timing (swing, layback). Treated as integral to flow.	Does not model articulation explicitly. Focuses on quantifiable timing and phonetic features; performance nuances are not encoded in the data-set.

Table 2.5: **Rap Flow Model Comparison.** Comparison of rap flow models relevant to Study III. The table highlights linguistic and rhythmic features used in flow analysis.

## Pitch

Komaniecki highlights pitch as a key parameter in the structure and delivery of rap flow [142]. In English prosody, pitch typically rises at the onset of a prosodic unit and gradually

declines as breath dissipates. This overarching pattern, average pitch and variability decreasing toward utterance endings, establishes prosodic boundaries. Deviations from this pattern may indicate sub-boundaries or positional cues within larger structures [127]. Most pitch movements occur on high-pitched accented syllables, though everyday speech tends to occupy the lower third of the vocal range [143]. By contrast, MCs often rap at higher registers, exploiting unused vocal range to shape emotional tone or structural emphasis—for instance, adopting a natural register for introspection, a lower one for anger, or a higher one for comedic effect [127].

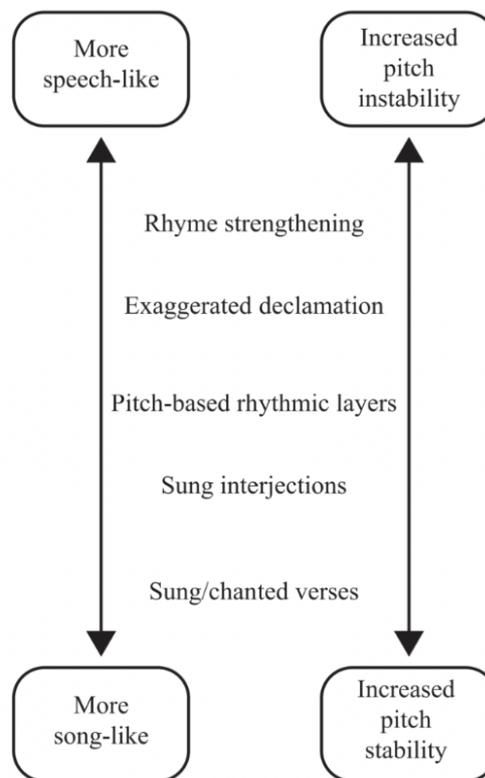


Figure 2.8: **Speech–Song Continuum in Rap Delivery.** A conceptual model mapping transitions in rap performance from more *speech-like* to more *song-like* expression, and from *increased pitch instability* to *increased pitch stability*. Intermediate steps include rhyme strengthening, declamation, pitch-based rhythmic phrasing, and sung interjections, culminating in fully sung or chanted verses.

Techniques closer to the top of the diagram are imprecisely pitched as in typical speech, while techniques towards the bottom are more precisely pitched as in typical song.

### Rhyme Strengthening and Pitch

Rhyme strengthening describes a technique in which MCs manipulate pitch to reinforce aural connections between rhymed syllables or groups. In the context of rap analysis, the term ‘rhyme’ extends beyond its traditional poetic definition to include any perceptual linkage through repeated forms or patterns. Thus, in rap, features such as rhythm, timbre, or pitch may function analogously, enabling listeners to perceive structural coherence even without perfect phonetic rhyme.

An example of rhyme strengthening through pitch can be found in Eminem’s *White America* [144], where the MC’s delivery is notated across multiple pitch register. As figure 2.9 shows, most of the verse is performed in Eminem’s mid-range voice. A closer pitch-based analysis reveals that different rhyme clusters are articulated at distinct relative pitch levels. For instance, syllables that rhyme with “blue” (highlighted in gray) are consistently delivered at a lower pitch, whereas rhymes associated with “baby” or “just like yourself” remain closer to Eminem’s vocal centre. The vertical separation of rhyme streams, assigning different rhyme families to different pitch levels, enhances the clarity of complex rhyme patterns and supports listener perception of structure within rapid, polysyllabic sequences. The technique becomes particularly salient in lines that feature intricate internal rhyming, such as “sits on the shelf” and “dimples would help,” which exhibit Eminem’s characteristic use of dense polysyllabic rhyme. Moreover, the pitch contour helps reveal deeper sonic relationships between seemingly unrelated lines—such as a potential cross-phrase, multi-pitch rhyme between “blue baby” and “lose Shady”, illustrating how pitch is deployed as an expressive and structural device in the construction of vocal flow. Appendix D provides a more detailed example.

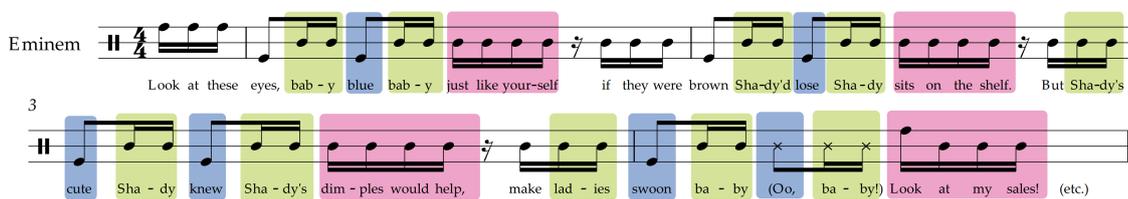


Figure 2.9: Eminem Example: A colour-coded transcription illustrating how Eminem’s delivery blends rhythmic speech and melodic phrasing. Highlighted segments show transitions from speech-like to more song-like forms through pitch-based rhythmic grouping, rhyme clustering, and sung interjections. The example exemplifies the continuum described in the Rap Pitch Model (Figure 2.8).

These distinctions between rhyme, cadence, and flow provide the analytical foundation for the computational measures introduced in Chapters 4 and 5, where they are operationalised through rhyme density, phonetic similarity, and temporal alignment. Rhyme density offers a means of quantifying structural and stylistic complexity across poetic and musical forms. It is typically expressed as the ratio of rhymed to total phonetic tokens and may be normalised by line or token count to account for differences in verse length or rhythmic pacing. This approach facilitates comparison between poetic and rap-based works without privileging a single formal model, instead acknowledging variation in how rhyme operates, externally at line endings in poetry, and internally across multi-syllabic sequences in rap. These normalisation conventions serve primarily to enhance interpretability rather than define the phenomenon itself, ensuring continuity between the written verse analysis in Study II and the performance-oriented evaluation in Study III. Understanding these as both aesthetic and quantifiable phenomena allows poetic and musical creativity to be analysed through algorithmic means. This convergence of literary form and technical method leads directly into the field of computational creativity—where models of generation, evaluation, and interaction aim to formalise how creative processes might be simulated or extended by machines.

## 2.7 Computational Creativity

'I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted.'

Alan Turing, 1950 [145]

Scientists, authors and inventors have long explored the possibilities of machines that can write. One of the earliest successful attempts was by John Peter in the 17 century. Peter published a booklet entitled *Artificial Versifying, or the School-boy's Recreation: A New Way to Make Latin Verses* [146]. The booklet contained six grids, each containing letters. By choosing any six-digit number, the reader could interrogate the tables and generate a line of six Latin words in hexameter rhythm or meter. The tables, ordered according to the rules of Latin grammar, produced grammatically correct sentences of the form: Adjective Noun Adverb Verb Noun Adjective. The words for each table conformed to the hexameter rhythm of classical Latin [110]. By the early 19th century there was discussion among scientists on whether computation machines, such as Leibniz's calculator and Babbage's Difference Engine, could be said to emulate mental processes [147]. However, until John Clark no one had demonstrated a mechanical device to automate the *creative process* of versification. In 1832 Clark invented the Eureka machine; by using a schema for a line of verse, and selecting words for each category with the appropriate meter, the machine could compose at random over 26 million different lines of well-formed Latin verse. Clark described the verse output as 'conceived in the mind of the machine' [147, 148].

### Structuralist Approaches to Storytelling

Long before computational creativity became a field of study, structuralist theorists explored the underlying patterns governing narrative construction. The work of two key figures, Vladimir Propp and Lev Vygotsky, were particularly influential.

*Morphology of the Folktale*: Propp analysed Russian folktales and identified a fixed pattern narrative structures underlying traditional storytelling. Propp's seminal study, *The Morphology of the Folktale*, introduced thirty-one *narrative functions* (predefined plot events

that occur in a predictable sequence) and seven *actants* (abstract character functions such as the hero, villain, helper etc) [149]. The functions occur in a specific order, although not all functions are present within each story. This type of structural analysis of folklore is referred to as *syntagmatic* where the focus is on the events of a story and the order in which they occur [150, 149]. Propp's formalised approach later influenced computational storytelling systems such as Tale-Spin and Minstrel that treated narrative as a rule-based sequence of plot elements. These systems attempted to replicate Propp's structures to generate machine authored stories [110, 151, 152].

However, a criticism of Propp is that his work "did not pay enough attention to cultural and historical aspects of the Russian fairy tale and studied only the general without regard to the particular"; it was asserted that "no attempt to relate his extraordinary morphology to Russian...culture as a whole...structural analysis is not an end in itself! Rather it is a beginning...It is a powerful technique of descriptive ethnography inasmuch as it lays bare the essential form of the folkloristic text. But the form must ultimately be related to the culture or cultures in which it is found" [149, 153]. This criticism of Propp is essentially concerned with *meaning* and the claim that contextual detail and nuance are lost when using a generic analytic framework. Similar criticisms are also made about rules-based and generative AI systems [56, 154].

*Theory of Narrative and Cognitive Development:* Vygotsky ideas emphasized narrative as a 'tool for thought', suggesting that storytelling is not just a literary device but a fundamental mechanism of cognitive development [155]. The Zone of Proximal Development (ZPD) concept highlighted the role of support in learning, whereby individuals build on prior knowledge through interaction and guided learning [114, 156]. Vygotsky's insights influenced modern AI creativity research, particularly in how human-AI collaboration can scaffold creative writing. Recent studies explore AI as 'cognitive assistants' that supports writers in expanding their creative capacity [157, 158, 159].

### **Artistic Approaches to Generative Language**

The generative potential of language has long been a source of interest within artistic and literary communities. One of the most well-known movements exploring this idea was Ouvroir de Littérature Potentielle (Oulipo), or the Workshop of Potential Literature. Oulipo embraced constraints as part of a deliberate approach to creativity. Founded in 1960

by writers and mathematicians in France, Oulipo sought to experiment with constraints and formal structures as a means of expanding creative possibilities. Instead of viewing rules as limitations, they embraced them as tools for unlocking new forms of expression [26, 110]. An example of their work is Raymond Queneau's *Cent mille milliards de poèmes* (One Hundred Million Million Poems), which consists of ten sonnets, each printed on a card with individual lines on separate strips. Since all sonnets share the same rhyme scheme and sounds, each line can be interchanged with the corresponding lines from the other sonnets, generating  $10^{14}$  unique poems - 100 trillion variations [160, 161]. The work exemplifies how combinatorial techniques can extend literary creativity beyond traditional authorial control, introducing the idea of algorithmic text generation in advance of computational methods.

The theme of automating the writing process also appeared in mainstream literature. In Roald Dahl's short story *The Great Automatic Grammatizator*, Adolf Knipe (a frustrated writer) builds a machine capable of producing literary works based on formulas that represent the rules of grammar and narrative. Unlike early 20th-century concepts of 'automatic writing', which emphasized spontaneity and the subconscious, Dahl's imagined system operates like a mechanical organ, where an operator manipulates dials and settings to produce stories with predefined emotional tones and narrative structures [162]. The story presents a satirical critique of industrialized creativity, and foreshadows contemporary debates around machine-generated literature. Alongside earlier experiments (e.g., Peter [146], Clark [147]), artistic exploration of generative constraints and automation contributed to the development of computational storytelling models, and sought to formalize and narrative text generation [110, 148].

### **More Poetry is Needed**

Poetry generation has been a foundation of computational creativity research over the past 50 years. Over this period, dozens of systems have produced poetry of varying sophistication [163]. Poetry is particularly suitable relative to other literary genre due to a number of factors. (i) Prose stories, however small, must have a structure such as characters, dialogues, plot etc. (ii) Poetry, on the other hand, can be abstract. Despite being one of the most difficult literary genres to understand (for humans) due to its multitude of language levels (vocabulary, grammar, syntax, semantics, phonemics, metrics, figures of speech, etc.), poetry does not necessitate the use of all of them at the same time. In other words, it is not mandatory for all these elements to be present in a single poem.

Writing poetic text does not have to be extremely precise, as several rules of language that hold in the generation of natural language can be broken in poetry [164]. The above characteristics make poetry a fertile ground for experimentation. Chapters 4 and 5 explore computational poetry in more detail.

### Rule-Based AI Story Generators

In response to Turing's challenge, "Can machines think?", early computational storytelling systems were grounded in symbolic AI and rule-based logic. In 1952, Turing's colleague, Christopher Strachey wrote a combinatorial algorithm to create electronic love letters. The poems it generated are some of the first works of electronic literature. Strachey wrote that his 'rather simple trick' could produce 'the illusion that the computer is thinking.' [165, 166]. A defining feature of early story generators was their reliance on hand-coded knowledge structures [167]. These systems typically employed one of two approaches.

*Story grammars*, inspired by linguistic grammar, define the structural components of a narrative, including plot elements, character roles, and causal relationships, through rule-based or hierarchical models. Story grammars can be used analytically to deconstruct existing narratives or generatively to create structured stories based on predefined rules.

*Story planners* operate on the premise that story generation is a goal-driven process. They use symbolic planning techniques to generate a sequence of events that unfold between the beginning and end of a story.

Following are notable examples of rule-based story generators.

- *ELIZA (1966)*: A natural language processing program designed to simulate human-like conversation through pattern matching and scripted responses. Despite lacking true understanding, ELIZA created the illusion of meaningful interaction, leading to the term the 'ELIZA Effect.' It demonstrated how simple rule-based systems could evoke emotional responses. However limitations such as the absence of reasoning, or contextual awareness, highlighted the gap between syntactic processing and true semantic understanding [168, 169].

## 2. Literature Review

---

- *Tale-Spin (1977)*: One of the first AI-driven story generators, Tale-Spin used symbolic AI to model character behaviors and decision-making within a simulated environment. Characters pursued goals based on logical constraints, leading to emergent narratives [152].
- *Narrative Grammars (1977)*: Proposed that stories could be understood as cognitive scripts, where meaning emerged from predefined templates of human experience. This approach influenced later computational models of narrative structure [170].
- *Universe (1984)*: the system had a character generator, a story planner, and used a library of schemas to indicate how scenes develop. The schema could in turn reference sub-schemas. Stories were constructed by iteratively decomposing schemas into goals and sub-goals [171]
- *Minstrel (1993)*: Introduced case-based reasoning (CBR), a method where AI 'remembers' past plot structures and recombines them in new ways. In theory, this allowed for greater flexibility in story generation compared to rigid rule-based systems [151].

Rule-based systems provide structure, but suffer from rigidity and predictability. Stories often follow repetitive patterns and lack the variation and adaptability seen in human-authored narratives. As a result, researchers sought alternative methods generate dynamic and unpredictable text.

## Statistical and Probabilistic Approaches

With advancements in computational modelling, rule-based approaches gave way to statistical and probabilistic models for text generation. Unlike rule-based approaches, these methods leveraged patterns in large datasets to generate narratives probabilistically [110, 167].

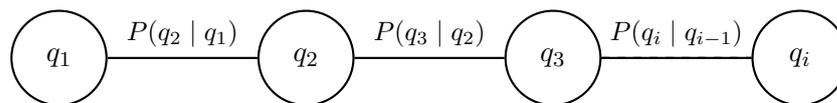
### Markov Chains

Markov chains and N-gram models are probabilistic model that predict the next word or phrase based on observed word probabilities within a given corpus. The models assume that that likelihood of a word occurring depends only on a limited number of preceding words, rather than an entire sentence or document [172].

More formally, consider a sequence of state variables  $q_1, q_2, \dots, q_i$ . A Markov model embodies the Markov assumption, which states that when predicting the future, only the present state matters, and the past has no direct influence beyond the current state. This can be expressed mathematically as:

$$P(q_i | q_1, q_2, \dots, q_{i-1}) = P(q_i | q_{i-1}) \quad (2.1)$$

where the probability of the current state  $q_i$  depends only on the immediately preceding state  $q_{i-1}$ , rather than the full history of previous states [172].



**Markov chains:** In the context of language modelling, a *first-order Markov model* assumes that the next word depends only on the current word, while a *higher-order Markov model* incorporates more prior words into the prediction. However, Markov models are limited by their inability to retain long-term dependencies, making them ineffective for capturing complex linguistic structures.

**N-gram models** extend the Markov assumption by conditioning word predictions on a fixed number  $N$  of preceding words. For instance, a *bigram model* predicts the next word based on the previous word, while a *trigram model* conditions predictions on the last two words. By approximating linguistic history using only a limited window of context, N-gram models offer computational efficiency but struggle with sparsity, requiring smoothing techniques to handle unseen word sequences [172].

Key Features	Strengths	Weaknesses
Simple and efficient for short text generation	Good at mimicking style of training data	Lacks long-term coherence
Requires minimal computational power	Effective for chatbots and text completion	Ignores grammatical and semantic structures
Predicts words probabilistically	Can produce locally fluent text	Repetitive and/or nonsensical outputs in longer passages

Table 2.6: Comparison of Strengths and Weaknesses of Markov Chains and N-gram Models

Typically, stories require coherence in order to be satisfying [173, 174]. Coherence is a multifaceted construct with varying definitions; there is some consensus that context, structure, and theme are important features of narrative coherence [173]. The requirement for narrative coherence limits the effectiveness of Markov models in generating satisfying rich narratives (e.g., plays or novels). However, Markov models have been used with some success for poetry and lyric generation [175].

### Hidden Markov Models (HMMs)

A Markov chain is useful for computing probabilities of sequences of observable events. However, in many text generation tasks, the underlying structure governing these events is *hidden*. In story or poetry generation, for instance, we observe words but do not directly observe the latent narrative structure, themes, or emotional progression influencing the text. A Hidden Markov Model (HMM) allows modelling of both the *observed* text (e.g., words or phrases) and the *hidden* elements that shape the generated text (e.g., plot

structure, semantic categories, or poetic metre). In the case of story generation, the hidden states might correspond to:

- **Plot Structure:** Given a set of story states (e.g., introduction, rising action, climax, falling action, conclusion), an HMM can generate narrative sequences by probabilistically selecting the next state based on the current one.
- **Character Development:** HMMs can be used to track a character’s emotional progression across a story, ensuring coherent emotional arcs.
- **Metre and Rhythm:** In poetry, hidden states could correspond to metrical patterns (e.g., iambic pentameter, free verse), ensuring syntactically and rhythmically structured output.

HMMs provide a structured, probabilistic approach to generating coherent stories, poetry, and other forms of text. By modelling hidden states such as narrative roles, character emotions, or metrical constraints, they enable data-driven text generation while maintaining logical progression [175, 176, 177]. Deep learning models have largely replaced Hidden Markov Models (HMMs) in large-scale text generation due to their superior ability to capture context and generate coherent text. However, HMMs remain a lightweight, interpretable choice for structured, rule-based text generation tasks.

Key Features	Strengths	Weaknesses
Uses hidden states to model unseen aspects of text generation	More structured than Markov Chains	Lacks long-term dependencies
Commonly used in speech recognition and part-of-speech tagging	Infers underlying themes in stories	Fixed transition probabilities limit creativity
Introduces probability distributions for word sequences	Improves structured storytelling over Markov models	Formulaic and predictable storytelling

Table 2.7: Strengths and Weaknesses of Hidden Markov Models (HMMs)

## Neural Networks and Text Generation

By the 2010s, the field moved toward deep learning, leveraging neural networks for text generation. Unlike statistical models, neural networks learn from large datasets to model text generation in a more context-aware way [178].

### Recurrent Neural Networks and Long Short-Term Memory Networks

Recurrent Neural Networks (RNNs) are sequence-based models trained on a corpus of stories to predict the probability of the next character, word, or sentence based on prior context. By sequentially processing *tokens* the smallest meaningful units of text, which can be words, subwords, or characters—RNNs capture dependencies, allowing earlier tokens to influence later ones [172, 177, 179]. However, standard RNNs struggle with long-range dependencies due to vanishing gradients, limiting their ability to maintain coherent narrative structures over extended passages. Long Short-Term Memory networks (LSTMs), an advanced variant of RNNs, address this limitation by incorporating memory cells and gating mechanisms to selectively retain or forget information [180]. This allows LSTMs to capture long-distance relationships in text, improving coherence in generated stories [179]. Text generation is typically performed by sampling from the learned probability distribution, often using techniques such as temperature scaling (to control randomness) or beam search (to optimize sequence selection) [181, 182].

---

Key Features	Strengths	Weaknesses
Maintains context better than statistical models	Generates more fluent and coherent text than probabilistic models	Struggles with very long-range dependencies
Learns from large text datasets rather than using predefined rules	Can adapt to different writing styles	Prone to repetitive loops in text generation
Uses hidden layers and backpropagation to learn relationships between words	Produces better grammar than Markov models	Lacks understanding of deeper narrative structures

---

Table 2.8: Strengths and Weaknesses of Recurrent Neural Networks (RNNs) and LSTMs

## Transformer Models

Transformer models represent a significant breakthrough in computational storytelling by overcoming the limitations of earlier neural network architectures. They have become synonymous in mainstream culture via LLMs such as ChatGPT or Gemini [183, 184]. Unlike RNNs and LSTMs, which process text sequentially, Transformers leverage self-attention mechanisms to capture long-range dependencies and contextual relationships more effectively [185, 186]. Transformer-based models introduced the self-attention mechanism, allowing text to be generated by attending to all words simultaneously instead of sequentially [185]. This enables them to generate highly coherent, stylistically adaptive, and contextually relevant narratives [187, 188, 189].

### The Attention Mechanism in Transformer Models

The attention mechanism is the core of Transformer models, enabling them to process and generate text with *long-range dependencies* and *context awareness*. It assigns weights to different words in an input sequence, allowing the model to focus on relevant information.

#### Self-Attention

Self-attention, also known as *scaled dot-product attention*, allows a model to consider different parts of a sequence simultaneously, unlike recurrent models (RNNs, LSTMs), which process text sequentially. Given an input sequence  $X$ , self-attention computes the relevance of each word to all others.

*Advantages of Multi-Head Attention:*

- Captures different contextual meanings at multiple levels.
- Enables the model to focus on multiple words at once.
- Improves parallelization (faster than sequential processing in RNNs).

In sequence-to-sequence tasks (e.g., translation), Transformers use cross-attention between the encoder and decoder:

- *Encoder self-attention*: The encoder learns contextual representations.
- *Decoder self-attention*: The decoder attends to previously generated words.
- *Encoder-Decoder attention*: The decoder focuses on relevant parts of the input sequence.

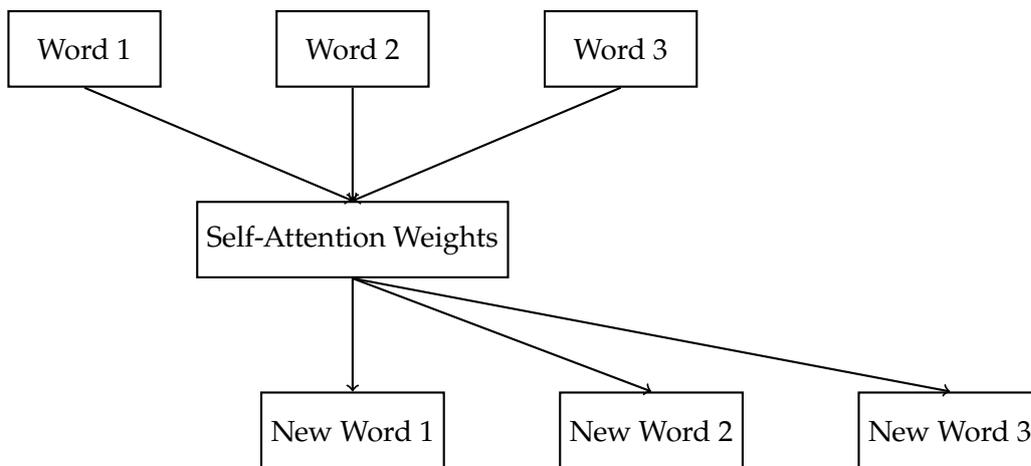


Figure 2.10: Self-Attention Mechanism: The model assigns weights to words in the sequence, determining how much influence each has.

The attention mechanism enables Transformer models to dynamically weigh the importance of words, making them highly effective for text generation, translation, and summarization. The introduction of multi-head attention and self-attention has been a major breakthrough in Natural Language Processing (NLP). (See Table 2.9 for an overview of the key features of Transformer-based models).

Key Features	Strengths	Weaknesses
Uses self-attention to process text non-sequentially	Maintains context across long passages	Computationally expensive
Few-shot learning allows adaptation with minimal examples	Generates highly coherent and realistic narratives	Lacks deep semantic understanding
Can mimic different writing styles and adapt to context	Enables interactive chat-based storytelling (AI Dungeon, GPT-based writing)	Prone to hallucinations and factual inconsistencies

Table 2.9: Strengths and Weaknesses of Transformer Models

The evolution of probabilistic text generation has progressed from early probabilistic models to advanced deep learning techniques. Initially, statistical approaches such as Markov Chains and Hidden Markov Models provided structured yet limited text generation capabilities. These were later replaced by neural network-based models, which improved fluency and coherence. RNNs and LSTM models introduced memory-based improvements, while Transformer architectures revolutionized natural language generation through self-attention mechanisms. Table 2.10 below provides a comparative overview of probabilistic text generation approaches, highlighting their features, strengths, and limitations.

## 2. Literature Review

Model Type	Key Features	Strengths	Weaknesses
Markov Chains & N-grams	Predicts next words based on probability of previous words	Fast, mimics style well	Lacks coherence, struggles with longer sequences
Hidden Markov Models (HMMs)	Uses hidden states for structured storytelling	Adds structure to generated text	Limited adaptability, still lacks deep understanding
Recurrent Neural Networks (RNNs) & LSTMs	Uses memory cells to retain context over longer sequences	Improves fluency and coherence	Prone to repetition, forgets long-term dependencies
Transformer Models (GPT, BERT)	Self-attention enables contextual awareness	High fluency, capable of long-form generation	Computationally expensive, lacks true semantic understanding

Table 2.10: Comparison of Probabilistic Text Generation Approaches

### Ongoing Challenges

Despite their impressive capabilities, LLMs have raised concerns about authenticity, bias, and creative agency.

1. **Lack of Deep Understanding:** LLMs generate text by analyzing vast amounts of statistical patterns, but they do not comprehend meaning in the way humans do. Their narratives, while fluent, often lack genuine intent or emotional depth [56].

2. **Bias in Storytelling:** Since LLMs are trained on historical and cultural texts, they could replicate and amplify biases present in those sources. Researchers are actively working on methods to reduce bias and improve model fairness [56, 57, 190].

3. **Over-Reliance on Probability:** Unlike human authors, who integrate originality and lived experience, LLMs generate text based on statistical likelihoods, which can sometimes result in generic or formulaic outputs [56, 191].

Description	Era	Technical Approach	Creativity Type
<b>Rule-Based Systems</b>	1970s–1980s	Symbolic AI, logical inference, hand-crafted rules	Combinatorial (fixed narrative templates)
<i>Tale-Spin</i> (Meehan, 1977)	1970s	AI-driven character decision-making within logical constraints	Exploratory (simulated world with bounded choices)
<i>Narrative Grammars</i> (Schank and Abelson, 1977)	1970s	Script-based models defining predefined plot structures	Combinatorial (reusing existing patterns)
<i>Minstrel</i> (Turner, 1993)	1990s	Case-based reasoning (CBR), adaptation of stored plot elements	Exploratory (recombinations of learned narratives)
<b>Statistical Models</b>	1990s–2000s	Probability-driven models, predicting next words based on corpus	Combinatorial (pattern-based variation)
Markov Chains	1990s	Word prediction based on n-gram probabilities	Combinatorial (structured randomness)
Hidden Markov Models (HMMs)	2000s	Probability distributions for sequential text generation	Exploratory (predictive text structures)
<b>Deep Learning (Neural Networks)</b>	2010s	Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTMs)	Exploratory (context-aware sequences)
<b>Transformer Models and LLMs</b>	2017–present	Transformer-based architectures (GPT series, BERT)	Transformational (adaptive and style-aware creativity)
GPT-4, Gemini, Claude	2019–present	Large-scale training, attention mechanisms, prompt-based learning	Transformational (zero-shot, few-shot adaptability)

Table 2.11: Evolution of Computational Story and Poem Generation, categorized by technical approach and Boden’s creativity types.

## 2.8 Co-Creative Spatiality

Co-creativity is an interactive and iterative process where humans and AI engage in mutual creative influence, shaping and refining each other's outputs [8]. Rather than being a static tool, AI functions as a dynamic agent within creative ecosystems, capable of suggesting, adapting, and co-constructing artistic, linguistic, and design-based outputs alongside human collaborators. From a spatial perspective, computational creativity can be understood as embedded within material and relational networks that structure human-machine collaboration. Drawing on Jordanous [12], this relationship can be framed through four interconnected dimensions:

- **People:** The individuals and communities engaged in co-creative activities, from artists, writers, and designers to researchers and general users.
- **Products:** The creative artifacts generated through human-AI collaboration, including text, music, visual art, architecture, and interactive media.
- **Processes:** The structured interactions that enable co-creativity, such as prompt-response cycles, feedback loops, iterative refinement, and generative constraints.
- **Places:** The physical and digital environments where co-creativity unfolds, ranging from AI-powered design studios and hybrid digital-physical spaces to collaborative online platforms.

By positioning AI within this broader socio-spatial framework, we move beyond viewing AI as an isolated computational system and instead recognize it as part of a co-creative ecology that is distributed across people, products, processes, and places.

### *The Role of Interfaces and Platforms in AI-Human Co-Creativity*

A central locus of human-AI co-creativity is the interface—a spatially structured environment where human inputs and AI-generated outputs converge. These interfaces range from text-based AI models like GPT, where writers iteratively refine AI-generated text, to AI-assisted music composition tools, generative design software, and interactive creative coding platforms.

Co-creative platforms, such as collaborative AI-driven art tools, automated storytelling environments, and AI-assisted writing assistants, act as mediators that shape the nature

of creative agency, collaboration, and interaction. The affordances and constraints of these platforms define who or what participates, how creative outputs emerge, and how agency is distributed between human and machine actors. Thus, co-creativity is not only a cognitive process but a spatial one, shaped by the technological architectures, social environments, and material conditions in which it occurs. Understanding co-creativity through spatiality allows for deeper insights into the ways AI integrates into creative workflows and how it expands or constrains human creative expression.

Dimension	Definition	Examples in Human-AI Co-Creativity
<b>People</b>	Individuals and communities engaging in AI-assisted creative tasks	Writers collaborating with GPT, artists using generative tools, AI-assisted architects and designers
<b>Products</b>	Creative artifacts generated through human-AI interaction	AI-generated poetry, AI-assisted music, algorithmically designed urban spaces
<b>Processes</b>	The iterative and structured interactions that enable co-creativity	Prompt-response cycles, reinforcement learning in art generation, real-time AI feedback systems
<b>Places</b>	The digital and physical environments where co-creativity occurs	AI-driven creative platforms, hybrid virtual/physical co-creation spaces, interactive digital art installations

Table 2.12: Dimensions of Co-Creativity in Human-AI Interaction (adapted from Jordanous, 2016)

### Hip hop as a Lens for Co-Creative AI

Hip hop emerged in the 1970s as a cultural movement rooted in the urban margins of the Bronx, New York, combining music, dance, visual art, and lyrical performance as forms of social commentary and creative expression [192]. Among its foundational elements, DJing, MCing, breakdancing, and graffiti, rap music (or MCing) rapidly evolved into a global art form centred on rhythmic, poetic language delivered over beats. From its inception, hip hop has been deeply entangled with technology not as an accessory but as a core driver of innovation, identity formation, and stylistic experimentation. Themes of place and space have also been important within hip hop. Early descriptions of hip

hop identify the public sphere as significant whether through visual expression (e.g., graffiti), or as urban geography rendered through lyrics [192]. Rose introduced a spatial analysis in detailing the ways that hip hop displays a 'transformative capacity endlessly capable of altering the use of space and technology [193].' Rose stresses the importance of the 'postindustrial' city as the central urban influence, 'which provided the context for creative development among hip hop's earliest innovators, shaped their their cultural terrain, access to space, materials, and education (p. 34)[193]. Since its beginnings, hip hop culture has also maintained an inbuilt element of competition waged through the hip hop's cultural forms (e.g., breakdancing, graffiti, rap). This completion has traditionally been staged within geographical boundaries [192]. The most visible form of hip hop's competition is though music and lyrics through the MC. As Jay-Z describes in *Decoded*

Hip-hop is a perfect mix between poetry and boxing. Of course, most artists are competitive, but hip-hop is the only art that I know that's built on direct confrontation. . . one thing you'll hardly ever find in hip-hop is rappers harmonizing on the mic. In the earliest days of hip-hop, MCs had to prove themselves to DJs before they could rock a party. The competition grew from there. . . it was about who could rep the hood, the borough, the city. battles are a hip-hop phenomenon that you just don't see played out in the same explicit, public way in other genres of music. Everywhere we look, we see competition (p. 48-49) [194]

Rap music takes the city and its multiple spaces as for foundation for its cultural production. In the music and lyrics, the city is an audible presence, explicit cited and digitally sampled (p. 201) [192]

### **Hip Hop as a Technological Culture**

From the 1970s, hip hop has evolved through successive waves of innovation, each defined not only by lyrical ingenuity and social commentary but by the adoption and repurposing of new technologies. At its core, hip hop music, emerged through technology. DJ Kool Herc's innovations with the "breakbeat" in the early 1970s relied on extending the percussive section of funk records using two turntables and a mixer. This technique of break looping marked the beginning of turntablism, transforming the record player into an instrument, and opening the door to the sampling aesthetic that would define hip hop production.

Grandmaster Flash and Afrika Bambaataa expanded this with scratch techniques and rhythmic manipulation, turning the DJ setup into a live performance interface [195, 196]. This was not simply about sound reproduction, it was about transformation: isolating, looping, and recombining fragments of recorded music to create something entirely new. In this environment, technological adaptation became a creative norm: artists reimagined commercial tools like drum machines (Roland TR-808), samplers (Akai MPC), and 4-track recorders as vessels for sonic experimentation [193, 197].

### **From the Turntable to the Transformer**

With the arrival of drum machines like the Roland TR-808 and samplers such as the E-mu SP-1200 and Akai MPC60, producers gained greater control over rhythm and texture. These machines allowed for granular manipulation of sonic material, enabling producers like J Dilla, DJ Premier, and RZA to develop distinctive sonic signatures [198]. As hip hop developed, so did its technological dependencies. The cassette tape enabled the spread of DIY mixtapes and underground battles, while the emergence of digital audio workstations (DAWs) in the 1990s brought studio-level production to home setups [199].

Sampling, both an artistic and legal battleground, allowed artists to \*embed fragments of history\* into their music, creating layered sonic texts that remixed the past into the present. It was both a technique and a politics: a way to quote, critique, and extend Black musical traditions through the layering of historical fragments. Far from being static, the sonic qualities of rap are shaped by their media environment: lo-fi hiss of cassette bootlegs, the digital clipping of early MP3s, or the algorithmic compression of today's streaming platforms. The UK equivalent of rap music, Grime, shares a close relationship with technology. Grime was shaped by young producers with minimal formal training, working with freely available or pirated software such as FruityLoops Studio. Lacking access to traditional synthesizers or samplers, they constructed distinctive, often futuristic soundscapes using the limited tools at hand. The resulting aesthetic was not merely a product of creative intent, but also of technological constraint [200, 201].

### **Lyricism, Flow, and Computational Practice**

While beat production evolved alongside hardware, the lyrical dimension of hip hop, rapping, developed its own technical grammars. Flow, cadence, and rhyme density became crucial stylistic markers, often signalling regional styles [9]. MCs pushed language to rhythmic and semantic extremes, producing what Alim describes as “lyrical acrobatics” dense, metrically complex verses that stretched the limits of syntax and prosody [202].

Lyrical, hip hop has also embraced technological mediation. From SMS battle rhymes to online cyphers, artists have used emerging platforms to write, rehearse, and circulate their work. Autotune, initially dismissed as artificial, became a vocal aesthetic that reshaped the emotional range of rap [203]. More recently, voice synthesis, TTS, and AI-generated lyrics are being integrated into production pipelines, suggesting that generative AI tools are not an external disruption, but rather the latest stage in hip hop’s ongoing dialogue with machine systems. Across decades, hip hop has treated technology not as neutral infrastructure but as a site of cultural play, political commentary, and identity work. As Moten and Harney argue, black musical traditions often subvert and repurpose available technologies to create counter-publics and alternative archives [204]. In this light, the use of AI in lyric writing, voice synthesis, or performance can be understood not as a break from tradition, but as an extension of hip hop’s ethos of *remix*, *reappropriation*, and *reinvention*.

Recent computational studies have treated flow and rhyme as measurable, algorithmically determined phenomena [139, 205]. In this context, AI systems that can generate, evaluate, or remix rap lyrics represent a continuation of the genre’s long-standing interest in constraint, formalism, and virtuosity. Generative AI systems are the latest in a line of technologies, from 4-track recorders to DAWs, that have changed how artists think, write, and perform.

### **Artistic Voice**

In written literature, *voice* refers to an author’s unique style or perspective, encompassing tone, diction, and narrative stance. This sense of voice is connected to *originality*. In *How to Write* it is described in terms of an artist who “ encounters the origin of their work within the *individual* experience of imagination.” It is the authentic experience that affords the work authority and therefore originality (p. 78) [206]. Within spoken word poetry and rap, voice has an artistic and physical dimension. It is not only a marker of identity but also the literal sound of the performer: their accent, pitch, cadence, timing,

and breath. Voice in this context is simultaneously a creative signature and an expressive medium. Freestyle, the spontaneous composition of lyrics, is widely considered the most challenging skill to master in rap [119]. An fMRI study of freestyle rap revealed that it requires rapid linguistic processing, as MCs must generate meaningful, rhyming phrases in real-time while adhering to tempo and rhythm constraints [115]. Freestyle is a particularly demanding domain for the dual conception of voice. Success depends on the linguistic properties of the verse, such as rhyme and wordplay, and also on delivery. Rhythm, vocal texture, and interaction with a beat are all critical to audience reception. Research confirms the complex cognitive processes involved in freestyle rapping such as improvisation, rhythmic control, and performance adaptation [115, 127].

Improvisation is an important topic for studying creativity more generally [20, 23, 207]. Specifically, as a creative act, verse improvisation requires extensive vocabulary, mastery of complex rhyme patterns, and broad subject knowledge across diverse topics [115, 9, 208]. The process of verse improvisation shares key characteristics with the *flow state*, where cognitive effort, heightened focus, and automaticity merge to enable fluid creative expression [23]. Freestyle rap also relies on phonological awareness, lexical retrieval, and verbal dexterity [115]. The ability to manipulate sound, word meanings, and rhythm in real-time is fundamental to the art form [119]. Battle rap, a competitive type of freestyle, introduces additional constraints that amplify both its creative challenge and its suitability for computational modelling [209]. Battle rap requires MCs to create responses that directly engage with an opponent while incorporating rhetorical strategies such as humour, wordplay, and personal rebuttals [120, 209, 210]. These constraints result in two main properties:

1. *A High-Level Creativity Challenge for Human MCs* Battle rap demands that performers adapt their performance on-the-fly while maintaining coherence, linguistic complexity, and audience engagement. The interplay between structured rhythmic demands and unpredictable opponent-generated content make it a cognitively difficult form of improvisational creativity [115, 211].

2. *A Structured Framework for Computational Approaches* Battle rap's constraints and evaluative criteria (e.g., audience reaction, rules) offer the potential for AI systems that can generate, evaluate, and refine battle rap performances using natural language processing, machine learning, and adversarial feedback loops.

Research shows that expert freestyle rappers engage in highly complex processes of linguistic improvisation, rhythmic control, and performance adaptation [115, 127]. Modelling these human attributes remains a challenge for current AI systems. At present, even state-of-the-art language models are not designed for real-time, dynamic, embodied interaction. In performance contexts such as rap, breakdowns in timing, style, or cadence are immediately perceptible; an AI that misaligns syllables or mismatches tone cannot pass as a co-performer.

Freestyle and battle rap together provide a compelling framework for examining how AI might perform real-time improvisation, respond to human input, and navigate the tension between structure and creative freedom. These practices highlight the importance of synchrony, adaptation, and stylistic coherence—qualities that remain under-explored in current models of computational creativity. The same principles extend to other time-based and embodied art forms such as dance and spoken word [212, 213]. These insights establish a conceptual bridge between the cultural study of hip hop and the design of AI systems capable of augmenting, evaluating, or emulating human creative performance.

### **Experimentation as Norm**

Hip hop's embrace of technology is not incidental, it is constitutive. As Moten suggests, Black musical traditions have long negotiated with technological mediation as a mode of survival and innovation [204]. Hip hop inherits this tradition, treating technology not as external but as internal to cultural production. From Auto-Tune (in T-Pain or Kanye West's work) to vocal chopping and remix culture on platforms like SoundCloud and TikTok, the voice itself becomes a modifiable object, a programmable instrument. This experimental ethos aligns directly with current AI applications in music and text generation. Tools like ChatGPT, Jukebox, and custom transformer models can now assist or even co-create with artists, offering prompts, generating punchlines, or mimicking flows. For some, this raises concerns about automation and authorship; for others, it extends the remix logic embedded in hip hop's DNA. Just as sampling redefined what it meant to compose, generative AI is reshaping what it means to write, perform, and collaborate. As Rose and Schloss argue, hip hop has continually absorbed, repurposed, and recontextualised technological systems as expressive resources [193, 197]. This practice that persists today in the engagement with AI and machine learning tools. This thesis builds on that premise. By framing AI tools as part of a lineage of technological mediation in hip hop, it becomes possible to explore

how co-creative systems can be designed to support cultural context and new forms of expression, without erasing the human artist. The question is not whether AI belongs in hip hop, but how it is used, and to what ends. Taken together, the theoretical perspectives on creativity, spatiality, and co-creation reviewed in this chapter find a living analogue in hip hop, a cultural practice that has always treated technology as a creative partner. From turntables to transformers, hip hop artists have not only adopted but redefined technological tools as expressive media. This thesis follows in that lineage. As table 2.13 illustrates, hip hop and generative AI are not isolated cultural phenomena, but overlapping systems of remix, performance, and authorship. These parallels justify treating hip hop not just as a cultural product, but as a framework for understanding AI-enabled creativity.

Theme	Hip Hop Evolution	Generative AI Evolution
<b>Sampling</b>	Turntablism and beat sampling (1970s–80s) created new compositions from existing records.	Pretrained models (e.g. GPT, MusicLM) draw from large corpora to generate new text, audio, and images, often recombining existing material.
<b>Authorship</b>	Questions of originality vs. derivation sparked debates (e.g., over ghostwriting or recycled bars). Sampling lawsuits (e.g., <i>Biz Markie</i> , 1991) reframed legal authorship.	Copyright concerns about model outputs and training data usage. Questions of whether outputs are derivative, transformative, or original.
<b>Remix Culture</b>	Remixes and mixtapes played a critical role in underground and mainstream scenes. Mashups and reinterpretations blurred authorial boundaries.	Fine-tuning, prompt engineering, and transfer learning mirror remix logic—altering foundational models to fit new creative goals.
<b>Performance</b>	Freestyle battles and live shows foreground improvisation, agility, and audience interaction. Flow and cadence define skill.	Real-time text and voice generation allow responsive AI outputs. Use cases include live AI DJs,
<b>Distribution</b>	Cassettes, radio, and later digital platforms (e.g., SoundCloud, YouTube) democratised production and distribution.	Open-source models, Colab notebooks, and API platforms enable global access and creative experimentation.
<b>Ethics &amp; Resistance</b>	Hip hop emerged as a voice for marginalised communities. Themes of resistance, surveillance, and appropriation recur.	Generative AI raises concerns over data bias, and cultural appropriation.

Table 2.13: **Parallels between the evolution of hip hop and generative AI**, focusing on sampling, authorship, remix culture, and creative agency.

This chapter has traced the interdisciplinary evolution of creativity, from psychological and cognitive models to socio-cultural perspectives in which creativity emerges through interaction, collaboration, and environment. This situated framing underpins placemaking, where creative practice is embedded in physical, social, and cultural infrastructures.

Building on the role of place in creative practice, the review examined how arts and language function as vehicles for cultural expression, with writing serving as a medium for storytelling, identity-making, and artistic production. The historical development of writing technologies provided a bridge to artificial writing systems—from rule-based methods to statistical models and, ultimately, to contemporary deep learning architectures such as GPT and BERT that reshape how creative agency is distributed across human and machine actors.

The chapter then situated AI-driven creative processes within cultural geographies, showing how generative systems operate across digital, physical, and hybrid spaces. By framing AI as an embedded component of socio-cultural environments, the review highlighted how such systems influence artistic production, mediate collaboration, and reconfigure creative agency. These perspectives establish the theoretical and technical foundations for the investigations that follow, where AI-assisted co-creativity is explored in the contexts of placemaking, poetry, and performance.

### **Research Gap and Key Areas of Exploration**

Despite advances across urban interaction design, digital placemaking, and computational creativity, three gaps remain:

1. **Situated AI within Creative Practice.** Existing research rarely investigates how AI participates in community-based, place-embedded creative activity. There is limited empirical evidence on how digital cultural infrastructure integrates AI into local artistic ecosystems. This gap aligns with the thesis claim that place acts as an active variable in creative production.
2. **Interpretable Proxies for Poetic Creativity.** While computational metrics for text exist, few provide interpretable, task-aware measures that meaningfully relate to poetic form, rhyme, cadence, and practitioner judgement. This gap motivates the development of the interpretable proxies central to the thesis.

3. **Co-Creative and Adversarial Human–AI Performance.** Human–AI collaboration has typically been explored through constrained, imitation-oriented text generation. Little work examines adversarial or performance-based interaction, especially in spoken word or freestyle rap— where co-agency, stimulation, and responsiveness matter. This aligns with the thesis focus on more-than-human creative ensembles.

These gaps directly motivate the three empirical studies: Study I investigates situated, place-based creative practice; Study II develops interpretable measures of creativity through constrained poetic tasks; and Study III extends these mechanisms into adversarial performance with Digital Creative Twins. As a combination, the studies provide a situated, interpretable, and stimulation-focused framework for understanding human–AI co-creativity across place, poetry, and performance.

## Chapter 3

# Study I: Place as a *Hip Hop State of Mind*



Figure 3.1: Exterior of Arts ARKADE, Piccadilly, London. The launch reimagined the former Clydesdale Bank as a platform for artistic practice, community memory, and digitally mediated expression, drawing inspiration from a hip hop state of mind.

“Our cultural strategy will meet head-on our city’s challenges and embrace its many opportunities through the ability of culture to engage and inspire - connecting the ordinary and the extraordinary. Culture will play a transformational role, unifying People and Place, mobilised by stronger, richer and deeper participation. This collaboration and dialogue will be facilitated by digital technology, as well as place based activity - demonstrating how culture is the ‘bridge’ to tackling the inequality in Swansea but which is also so prevalent in cities and towns in Wales and across the UK.” [2, p. 8]

*Swansea UK City of Culture 2021*

### 3.1 Introduction

Arts ARKADE was a practitioner-led hybrid cultural and technical platform deployed in Swansea and later London, combining venue design, digital projection and AR systems, and artist-led participatory programming. It operated as an experimental model of digital cultural infrastructure embedded within specific urban contexts. Study I examines how this configuration functioned across the two locations (Swansea and London), addressing:

**RQ1: Place and Practice. How does digital cultural infrastructure influence creative practice in place-based contexts?**

The study investigates how creative behaviour emerges in situ, how artists engage with the spatial and technological affordances of each site, and how practitioner-led design shaped these interactions. The analysis is supported by a set of subsidiary guiding questions (Table 3.1), which operationalise RQ1 across creative infrastructure, participatory design, cultural practice, and site integration. The vision underlying both deployments is described in Section 3.4, where stakeholder aims are also set out.

### 3.2 Context and Related Work

The UK is heavily urbanised, with most of its population living in towns or cities, consistent with global trends in urban population growth [214]. Towns and cities form the functional building blocks of contemporary society but face significant challenges in catering to the

physical needs of their populations, but also to provide experiences that enhance well-being and quality of life. As technology platforms proliferate, cities must foster a positive interplay between people, place, and technology. Understanding the evolving relationship between physical and digital space is key to imagining, designing, and sustaining cities that promote belonging, interaction, and public engagement. Designing technologies for art, culture, and learning offers a compelling pathway toward reimagining cities. Smart cities are often framed as opportunities to optimise infrastructure and enhance economic performance [68, 215]. However, they also provide openings for civic engagement by reframing urban space as playful, participatory, and emotionally resonant [68]. While urban design has traditionally been the domain of planners, architects, and geographers [216], interaction design now plays an increasing role as cities become digitally networked through public Wi-Fi, digital displays, and the widespread adoption of smartphones [217, 95]. Within HCI, researchers explore the intersection of play and urban technologies [68, 79], while visual artists introduce playfulness into the public realm through street art, interactive furniture, and dynamic installations [218, 219].

Urban technologies must be adapted to local contexts to be effective. This principle underpins the concept of playable cities and aligns closely with localism, understood here as decentralisation or devolution to citizens and communities [220]. Localism also addresses skills gaps and reduces dependency on national or international artists by building local creative capacity. This resonates with creative placemaking, which “envisions a more decentralised portfolio of spaces acting as creative crucibles” [59]. This chapter introduces Arts ARKADE, a hybrid cultural-technical platform developed in Swansea and later deployed in London. Each site functions as a space for creative infrastructure, digital capacity-building, and community engagement. Arts ARKADE operates at the intersection of urban culture, participatory design, and emerging technology, offering a model of how digital tools can enhance artistic practice, reframe spatial narratives, and foster co-creation. Study I directly addresses RQ1, examining how the configuration of Arts ARKADE as digital cultural infrastructure shaped creative practice in Swansea and London. No additional research questions were introduced; all subsidiary observations are interpreted in relation to the primary focus on place-based creative engagement.

## A Hip Hop State of Mind

The term *hip hop state of mind* is used to describe hip hop as cultural framework to “positively transform...lives and achieve a formidable sense of empowerment, street knowledge, resilience, and self-healing” [221]. While generative AI is not directly embedded in this phase, the methods and infrastructure established through Arts ARKADE, particularly around accessibility, iterative design, and community involvement, lay essential groundwork for the creative systems evaluated in Study II. Arts ARKADE draws on hip hop as a cultural frame that links movement (dance), voice (spoken word, rap), imagery (street art), and collaboration (cyphers, jams). This framework connects poetic and embodied creativity across urban and digital contexts, linking the core themes of this research: place, poetry, and generative AI.

Figure 3.2 introduces the thesis schema—**Place, Poetry, Generative AI, and Practice**—with **Place** foregrounded to reflect Arts ARKADE as an active driver of creativity in a hip-hop ecology. Minimal tags mark the situated overlaps (public realm, ARKADE site, community-led practice), while external callouts summarise the Study I argument and a right-hand note previews the word-oriented tools taken up in Study II. The same Venn will reappear with shifted emphasis in Studies II–III and the Discussion to make the trajectory—*place* → *word* → *tools* → *performance*—explicit.

## Hip Hop as Place

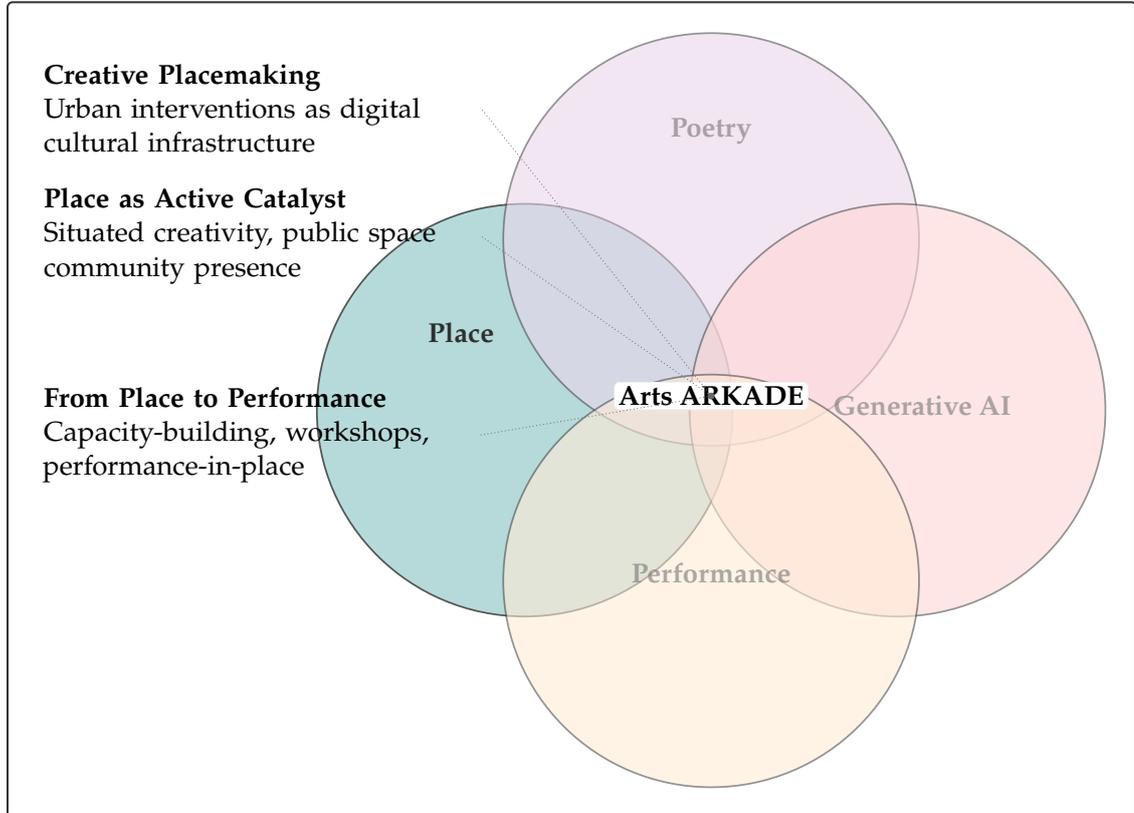


Figure 3.2: **Study I framework: Place** foregrounded within the place–poetry–performance–AI schema. This positions Arts ARKADE as a site where public realm, community presence, and performance-in-place shape creative work. The model aligns with *creative systems* approaches; see Jordanous [12] and Literature Review Section 2.8. Later chapters return to **Poetry** and **Generative AI** to examine how computational tools mediate style and performance.

### 3.3 Research Aim and Study Focus

Study I addresses RQ1 by examining how creative practice develops within Arts ARKADE as a form of digital cultural infrastructure. To support this analysis, a set of subsidiary guiding questions was developed. These questions do not constitute additional research questions; rather, they operationalise RQ1 across four domains identified in the literature on placemaking, urban interaction design, and performance-based media. The questions and associated considerations are summarised in Table 3.1.

### 3. Study I: Place as a Hip Hop State of Mind

Research Questions	Key Considerations
<b>Creative Infrastructure</b>	
How can physical environments support inclusive, digitally enabled creative activity?	Spatial configuration, embedded interaction, sensory feedback
What infrastructural models enable low-barrier access to computational tools?	Affordability, modularity, standardised toolkits
<b>Participatory Design and Skill Development</b>	
How do local artists engage with unfamiliar technologies through prototyping?	Onboarding, collaborative scaffolding, socio-technical literacy
What is the value of MVAs in supporting rapid experimentation?	Iteration, reflection, fast-feedback prototyping
<b>Cultural Practice and Expression</b>	
How do hip hop principles (voice, movement, rhythm) inform interactive space design?	Multimodality, expressive range, culturally grounded design
What forms of co-creation emerge in hybrid artistic-technological spaces?	Collaboration, authorship, responsiveness
<b>Case Study Integration</b>	
How does the Hip Hop for Health project exemplify the ARKADE vision?	Multimodal integration, creative health, freestyle practice in context

Table 3.1: Research questions guiding Study I, with associated key considerations.

## 3.4 Method

This study adopts an embedded case-study methodology across two sites (Swansea and London), combining participatory design activities, situated observation, workshop facilitation, and documentation of creative engagement. The evaluation approach is practice-led: creative outputs, participant interactions, and site-specific behaviours were analysed to identify patterns of authorship, participation, and use of digital affordances. This iterative structure aligns with the study's aim of understanding how creative infrastructure shapes place-based practice.

### **Practitioner Framing**

The methodological approach for Study I was shaped by practitioner knowledge. The design and facilitation of the placemaking interventions drew on practice-led research principles in which creative activity functions both as an investigative mode and as a means of eliciting situated, community-rooted insight. The author's background in urban culture, creative writing, and performance provided an interpretive lens for understanding how participants engaged with narrative, spatial storytelling, and collaborative artistic production in place-based settings. This positioning aligns with practice-led research as an iterative movement between creative practice and reflective analysis, where practitioner expertise guides methodological choices and enables richer interpretation of creative processes [16, 222].

### **Analysis Approach**

The analysis followed a practice-led interpretive approach consistent with the practitioner framing outlined above. As the primary aim of Study I was to understand how creative practice unfolded within the configured spaces of Arts ARKADE, the analytic process drew on the researcher's professional expertise in performance, poetry, and community-based creative production. The materials generated across the deployments, including observational notes, photographs, workshop artefacts, participant comments, and on-site documentation—were reviewed iteratively to identify patterns in participation, authorship, and engagement with the digital affordances of the venue.

This interpretive review emphasised how creative behaviours emerged in context rather than reducing them to coded categories. A light cross-site comparison (Swansea and London) was undertaken to examine how different infrastructure conditions shaped modes of expression and public interaction.

### **Arts ARKADE Vision**

Across both Swansea and London, Arts ARKADE was conceived as a practitioner-led model of digital cultural infrastructure designed to support creative practice in place-based contexts. In Swansea, the vision was co-developed with Swansea Council's Cultural Services team as part of a town-centre regeneration pilot [223]. In London, the vision was extended through a partnership with The Crown Estate, who framed the flagship

### 3. *Study I: Place as a Hip Hop State of Mind*

---

Piccadilly site as a “centre for urban excellence” intended to inspire the next generation of Londoners [224, 225]. In Swansea [223] the physical spaces were renovated through a participatory design process involving paper sketches and 3D visualisations [226], which informed the placement of projectors, lighting, screens, and other equipment. The design phase was followed by technical implementation and artist engagement. In contrast, the London deployment operated in a high-footfall urban setting with minimal installed technology; instead, activity centred on AR in the surrounding public realm, weekly cyphers, DJing, dance events, and street-facing creative practice.

To contextualise the Swansea and London case studies, a comparative framework was developed to identify the structural and operational dimensions most relevant to Arts ARKADE’s role as digitally enabled creative placemaking. Both locations share a core curatorial vision yet operate within markedly different spatial, cultural, and technological environments. The framework distils these into five analytical dimensions: spatial context, artistic ecosystem, technological infrastructure, modes of participation, and digital–physical integration. These dimensions foreground both the continuities and divergences that shape participant experience, community engagement, and the potential for technology-mediated creative practice.

Study I therefore investigates **RQ1** by examining how Arts ARKADE, configured as digital cultural infrastructure, shaped participation, authorship, and creative behaviour across two contrasting locations. No additional research questions were introduced; all subsidiary observations are interpreted in relation to the primary focus on place-based creative engagement.

Arts ARKADE explored how digital cultural infrastructure can act as a catalyst for regeneration. By focusing interventions within a single building, the project tested approaches that might later be scaled to the city as a whole. The venue, a former music shop, was renovated through a participatory design process which included paper sketches and 3D visualisations [226]. The 3D models informed placement of equipment such as projectors, lighting, and display screens. The design phase was followed by technical implementation and artist training.

Study I frames Arts ARKADE within a hip hop ethos that treats place as a catalyst for poetry and rap, where visual arts, movement, and music shape site-specific creative practice.

Dimension	Description
<b>Swansea</b>	
Spatial context	Post-industrial waterfront; adaptive reuse of warehouse
Artistic ecosystem	Emerging local artists; strong university links
Technological infrastructure	<b>Large immersive screens</b> ; on-site AR installation
Modes of participation / culture	<b>Limited events</b> : cyphers, installations, workshops
Digital–physical integration	Screen-based immersion plus AR; episodic programming
<b>London</b>	
Spatial context	Urban high street; Piccadilly flagship site
Artistic ecosystem	Established networks; high public footfall
Technological infrastructure	<b>AR outside the location</b> ; no screen technology on site
Modes of participation / culture	<b>Regular cyphers</b> , DJing, dance, workshops, installations
Digital–physical integration	Street-facing AR; recurring live culture and community practice

Table 3.2: **Framework for Arts ARKADE Studies.** Overview of the components of the Arts ARKADE deployments, including spatial design, community engagement, and digital interventions. This table situates Study I within the broader logic of digital cultural infrastructure introduced in Chapter 1.

## Swansea and London Case Studies

Arts ARKADE, operated across Swansea and London as a catalyst for cultural exchange and technological experimentation. In each location, the intervention explored how physical and digital environments integrate meaningfully through arts-led placemaking. Its adaptive reuse of underutilised spaces, particularly its flagship site in Piccadilly, London, demonstrates how creative infrastructure can transform exclusive or neglected areas into inclusive, multi-sensory environments [7, 227]. These transformations incorporated tools such as augmented reality (AR), generative AI, and spatial computing to support new forms of expression, interaction, and storytelling. The Arts ARKADE case studies were exploratory, place-based interventions that mapped the creative landscape and surfaced practice-led hypotheses, constraints, and candidate measures from real contexts. The

### *3. Study I: Place as a Hip Hop State of Mind*

---

insights guided Studies II and III, where they were formalised and tested under more controlled conditions.

Swansea and London shared strategies and distinct local adaptations. Swansea functioned as a test bed for experimentation, building digital fluency and integrating emerging technologies into education and public programming. London focused on amplifying existing grass roots creative practices through cultural ambassadors. These differences underscored the need for placemaking strategies to remain sensitive to local ecosystems, technological resources, and socio-cultural dynamics [59]. Beyond digital augmentation, Arts ARKADE embedded movement, music, and embodied creativity into its placemaking ethos. Workshops, jams, and performances foreground the body not just as a participant but as a medium of spatial inscription. This aligned with a performance understanding of placemaking as an iterative, co-created process; performed rather than imposed. Central to this methodology is hip hop, not only as a cultural reference point but as an integrative framework that links movement (dance), voice (rap, spoken word), imagery (graffiti), and collaborative process (cypher, remix). Hip hop functions as a socio-cultural and epistemological bridge that binds together the themes of this thesis: place, poetry, and generative AI. Its principles informed the design of Arts ARKADE, and the methods of creative experimentation. This approach raised a set of critical questions: How do digital tools shape access to space, visibility, and voice? Who benefits from digitally augmented placemaking, and who might be excluded or underserved? Can these practices challenge dominant narratives of urban development, or do they risk reproducing spatial inequalities under new technological guises? [7, 59]

By situating Arts ARKADE within debates on digital geographies, cultural infrastructure, and participatory design, the study contributes to understanding how technology can support equitable, culturally rooted urban futures.

## Arts ARKADE Swansea (2022-2023)

**Vision:** Arts ARKADE Swansea was co-created with Swansea Council’s Cultural Services team as part of a pilot programme support town-centre regeneration [223]. Based in the former Cranes music shop [228], the space was refitted with digital technologies to enable collaboration, prototyping, and digital artworks for public platforms [226]. It also acted as a creative hub for local artists to acquire computational skills and experiment with interactive digital works. A structured workshop programme introduced participants to interaction design, machine learning, and creative prototyping, combining practical training with community-based capacity building. Arts ARKADE also connected with Swansea’s wider cultural venues (e.g., The Volcano Theatre, the Dylan Thomas Centre) as part of a creative development network [2].



Figure 3.3: **Former Cranes Music Shop.** The former Cranes music shop prior to redevelopment. The unit was transformed into Arts ARKADE between 2022 and 2023 as part of Swansea’s cultural regeneration programme.

## Design and Development

The mood board (Figure 3.4) served as an early conceptual anchor for the project’s creative direction, particularly in relation to screen-based interaction and spatial design. Drawing on visual references from street culture, speculative design, and digital performance environments, the board helped articulate a visual language that balanced raw, grassroots

### 3. Study I: Place as a Hip Hop State of Mind

---

aesthetics with emergent forms of technological immersion. These references informed not only the graphic design of the screens themselves but also broader questions of how public realm interfaces could support experimental modes of engagement, storytelling, and participation.

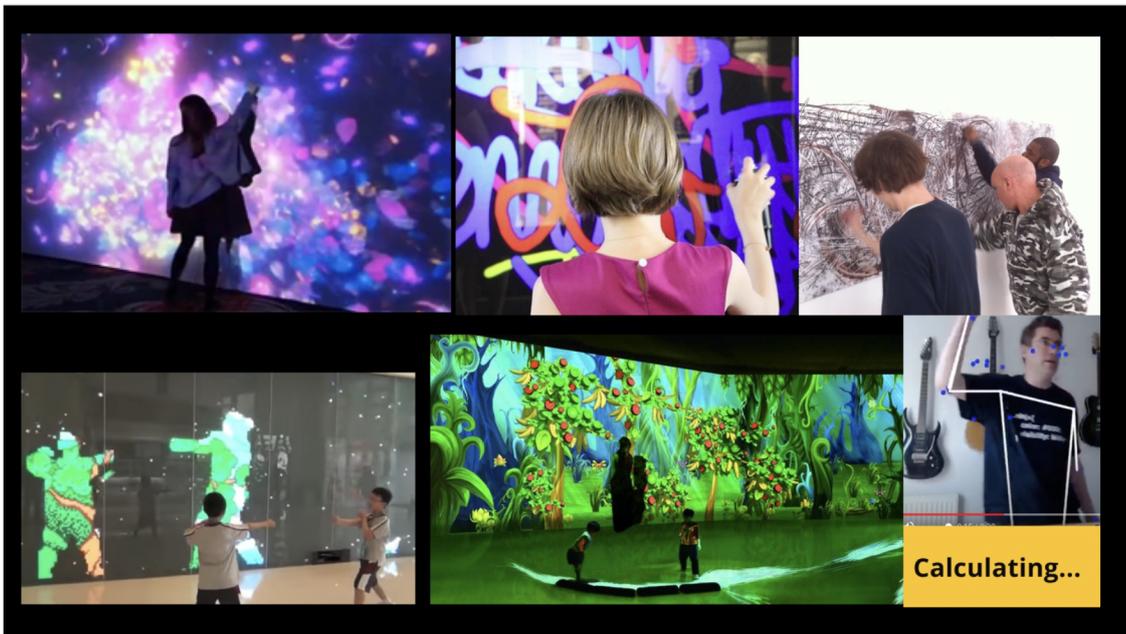


Figure 3.4: **Mood board.** These visuals informed immersive screen layouts and the public realm experiments that shaped the project's initial approach to technology-enabled placemaking. The large screens were designed to immerse visitors and offer multi-sensory such as the ones experiences.

In particular, the imagery influenced the prototyping of large-format screen layouts that aimed to go beyond passive display. The screens were envisioned as portals for mood, rhythm, and atmosphere, generating emotional resonance through layered visuals, ambient motion, and responsive audio. These experiments supported the project's approach to technology-enabled placemaking by inviting participants into multisensory experiences that were culturally grounded and spatially transformative. The examples in Figure 3.6 illustrate how this design ethos was translated into physical installations and digital interfaces.

## What

Activities (music, dance etc) as icons. The icons feature inside the space.

## Where

Swansea. Represent with graphic rather than words. Pixelated swan on the water.

## Quote

Powerful words to inspire action. Simple and memorable. The words tell us what we need to do.



Figure 3.5: **Window Design Concept.** Swansea Arts ARKADE window design concept and final execution (bottom photos). The design included (i) **Lyrics** by the MC *Ice Cube* who was also represented internally via a mural blended with graffiti. The design goal of the window was to (A) act as a prompt to enter the space and (B) provide a call-to-action for visitors that "Express Yourself" was the key purpose of building. (ii) **Pixelated graphics:** these were design to provide a sense of nostalgic history but also provide an accessible way to include QR codes as part of the aesthetic. The QR codes were intended to act as both information gateways and a delivery platform for work by local artists, following participation in Artcamps. The basic proof of concept was tested. Due to lighting issues as well as software constraints the concept was not further developed.

### 3. Study I: Place as a Hip Hop State of Mind

---

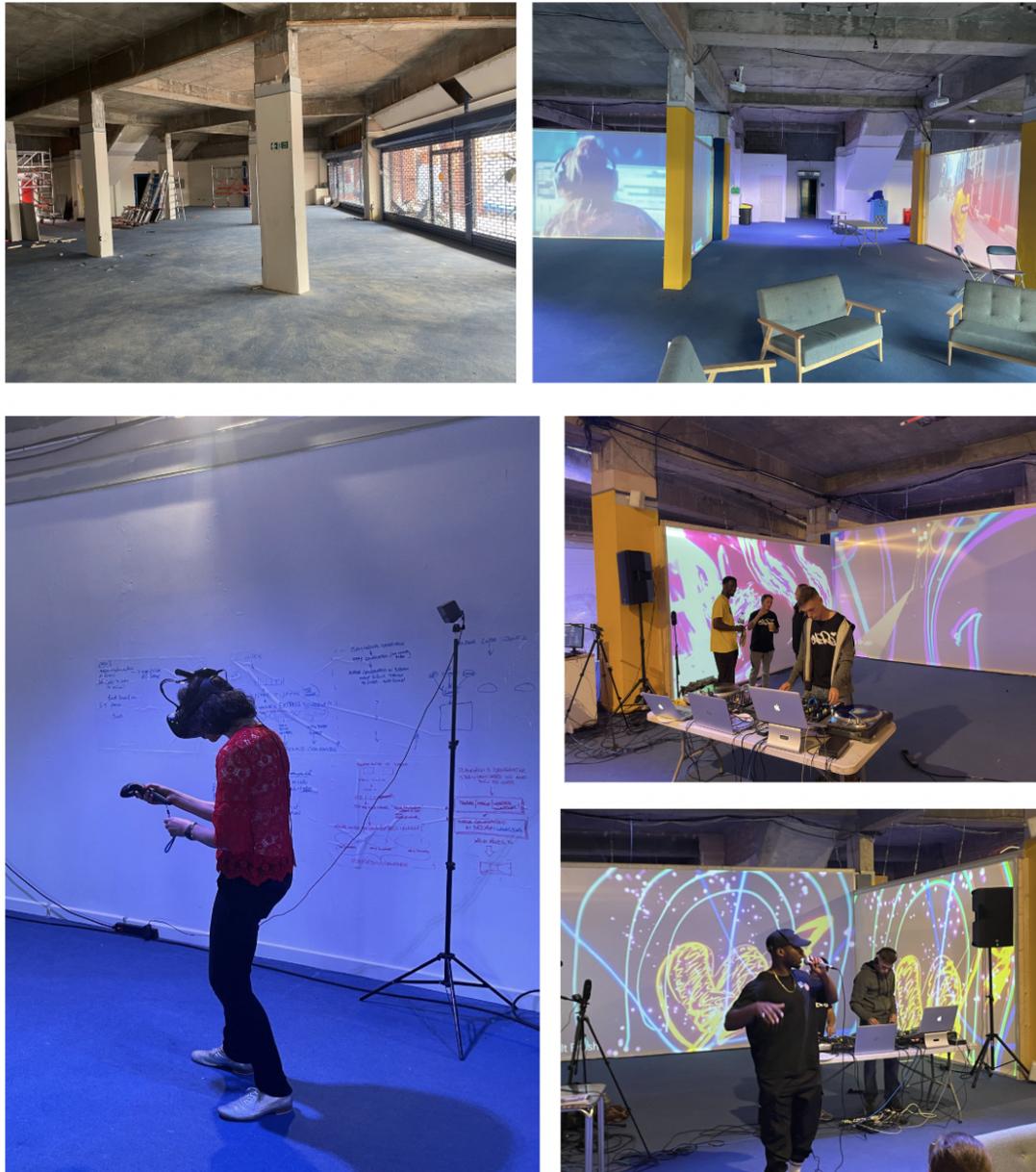


Figure 3.6: **Digital Screens.** Photo (top left) shows the building prior to the project starting. A key part of the design concept was digital screens. The implementation used standard technologies such as projectors, PCs and theatrical canvas. The screens provided experiences such as **(i) Multiple simultaneous viewing** (top right photos) in which visitors could see images from multiple screens in various positions. This could be used to build intrigue and surprise depending on the type of installation. **(ii) Large 3-sided immersive screen** (bottom right) on which to project images. In the photos, the screen displays images from the output of a 3D virtual reality painting application, Tilt Brush [229]. The video was created by an artist who moved around a part of the building (bottom left). The Tilt Brush output (positional and other data such as colour, brush type, stroke thickness etc.) was used to create a video which formed the backdrop for a live MC freestyle (bottom right). The Tilt Brush output included words and phrases, allowing the movement of the artist to connect with the lyrics and wordplay of the MC.

## Technology: Set-Up and Equipment

To support accessible creative experimentation, Arts ARKADE Swansea was configured using a low-cost, modular technology stack that leveraged commodity hardware and widely available software. Figure 3.7 presents the building layout, showing how the Physical Making Space, Digital Making Space, Immersion Space, Dance Space, and Exhibition Walls were arranged across the venue. This layout enabled multiple parallel modes of engagement, from structured workshops to informal experimentation.

Across these zones, the technical infrastructure consisted of the following components:

- **Standard PCs with consumer-grade GPUs.** These machines were used to run browser-based machine learning tools, projection systems, and audio–visual applications. The choice of commodity hardware ensured that all creative processes could be replicated outside specialist labs.
- **Large-format digital screens and short-throw projectors.** These provided full-wall or floor projection surfaces and were used extensively in the Immersion Space, Dance Space, and Exhibition Walls. Their purpose was to create responsive visual environments for interactive performances, digital art prototypes, and participatory activities.
- **Webcams and USB microphones.** These devices supported motion detection, real-time body tracking, spoken-word capture, and audio-reactive experiments. Their plug-and-play nature minimised technical overhead for artists participating in workshops or Artcamp activities.
- **Browser-based pose-estimation frameworks.** Tools such as **PoseNet** and **BlazePose** were used for lightweight, real-time body tracking. Because these systems run directly in the browser using TensorFlow.js, they removed the need for specialist installations or high-end computing, aligning with the project’s emphasis on accessibility.
- **Open-source creative toolkits.** JavaScript libraries (e.g., p5.js, WebGL-based environments) and freely available audio–visual packages enabled rapid prototyping. These were used by artists to explore projection mapping, motion-responsive sketches, and interactive poetry or performance work.
- **Limited proprietary software.** Where appropriate, Resolume Arena [230] was used for projection blending, video layering, and advanced media manipulation during

### 3. Study I: Place as a Hip Hop State of Mind

showcases or exhibition events. However, the primary workflow remained centred on open, replicable tools.

Where possible, the technology stack prioritised minimal cost and maximal accessibility. This meant favouring consumer hardware, web-based tools, and open-source frameworks over specialist research software. This design choice aligned with the practice-led ethos of Arts ARKADE, in which technology functioned as a configurable layer of the creative environment rather than a barrier requiring technical expertise. It also ensured that the interventions could be reproduced in community contexts, making the infrastructure suitable for public-realm experimentation and artist-led exploration.

Overall, this configuration provided a flexible and robust foundation for the creative processes examined in Study I. The equipment, spatial layout, and software tools allowed participants to engage with digital making, interactive performance, and AI-assisted experimentation in ways that were situated, culturally grounded, and technically accessible.

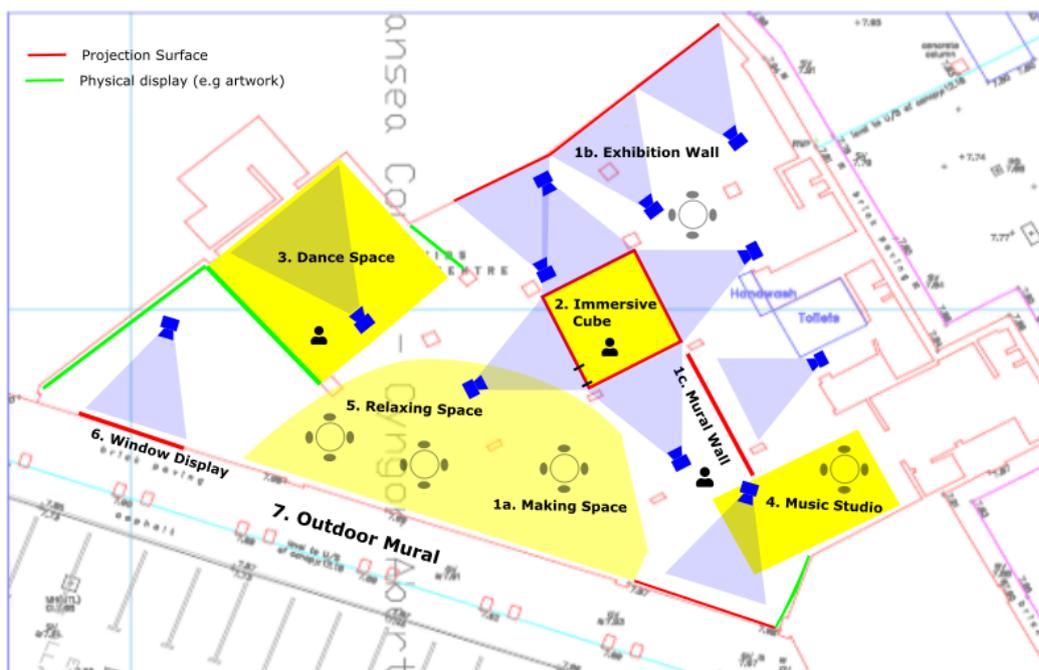


Figure 3.7: **Arts ARKADE Swansea: Building Layout.** The diagram illustrates the spatial configuration used during the project, including the Physical Making Space (1a), Exhibition Wall (1c), Mural Wall (1b), Digital Making Space, Immersion Space, Relaxing Space, Dance Space, and associated projection surfaces. This layout provided multiple zones for creative workshops, Artcamp sessions, interactive installations, movement-based experimentation, and public-facing digital artworks.

### Activity 1: Artcamp

Artcamp was developed in response to the barriers that many artists face when approaching emerging technologies. Advanced digital tools are often expensive, complex, or require specialised training [231]. Public-realm interaction design also demands iterative experimentation, modelling of dynamic situations, and the ability to reflect on the role of algorithms in shaping experience [232]. The pedagogical model drew directly from rapid prototyping practices in the technology industry [231, 233, 234] and was aligned with an iterative cyclic framework, in which practitioner knowledge, reflection, and creative experimentation continuously feed one another. Within this framework, the researcher's own creative practice<sup>1</sup> played a central role in shaping workshop flow, demonstrations, and critique. Artcamp introduced participants to computational frameworks through live demonstrations and guided exercises. Artists were shown how to work with the building's digital affordances, including projection surfaces, the immersive screen cube, the digital studio, webcams, microphones, and browser-based creative software. The emphasis throughout was on reconfigurable, accessible tools that supported rapid experimentation.

Artcamp was an intensive, two-day introduction to computational creativity and participatory design within the Arts ARKADE environment. Participants were divided into small teams and tasked with developing Minimum Viable Artworks (MVAs) using the creative infrastructure available on site. The workshops focused on embodied modalities such as movement and voice, with the goal of introducing computational thinking through hands-on interaction and rapid prototyping.

### Participants

Participants were practising artists and creative practitioners recruited through Swansea Council's Cultural Services team, local arts organisations, and the researcher's existing community networks. All participants responded to an open call circulated via mailing lists and partner organisations<sup>1</sup>. The cohort represented a broad spread of creative disciplines, early-career local practitioners, and experienced national and international artists.

Because Study I functioned as a practice-led placemaking deployment rather than a controlled experiment, demographic variables such as age were not analytically critical. What mattered for RQ1 was participant engagement with the space, their creative

---

<sup>1</sup>including professional experience as arts workshop leader with theatre, dance and other performance arts. See Section 1.3 for further detail

### 3. Study I: Place as a Hip Hop State of Mind

backgrounds, and how they interacted with the digital affordances of Arts ARKADE. A summary of participant backgrounds and experience levels is provided in Table 3.3.

ID	Gender	Role / Background	Experience / Locality
P01	F	Dance	Highly experienced, international
P02	F	Music / Drums	Experienced, international
P03	M	Writer	Early career, local
P04	M	Visual arts / Graffiti	Early career, local
P05	F	Visual arts / Animation	Mid-career, national
P06	M	Graffiti	Experienced, national
P07	M	Music	Experienced, local
P08	F	Dance	Early career, local
P09	F	Visual arts	Early career, local
P10	F	Music	Early career, local
P11	M	Visual arts	Early career, local
P12	F	Visual arts	Early career, local
P13	M	Interactive / Digital	Early career, local
P14	M	Music	Early career, local
P15	M	Dance	Highly experienced, international
P16	F	Visual art / Interactive	National
P17	M	Visual arts / Interactive	National
P18	F	Visual arts / Projection	Early career

Table 3.3: Summary of participants involved in Arts ARKADE Swansea activities. Study I was a practice-led placemaking deployment rather than a controlled experiment; therefore, only role, expertise level, and locality are reported. Age data was not collected, as it was not required for evaluating RQ1.

The groups worked with standard PCs equipped with consumer-grade GPUs, large digital screens, and short-throw projectors, alongside widely available open-source software. Practical sessions introduced lightweight machine-learning models such as PoseNet and BlazePose for webcam-based movement tracking. These tools supported experiments in interactive dance, gesture-responsive visuals, and spoken-word performance. Because the models ran directly in web browsers, no specialised installations were required, lowering the barrier to entry. Consistent with practice-led research principles, technical onboarding was facilitated through the researcher’s own creative and technical expertise, enabling participants to engage with unfamiliar tools through guided, hands-on exploration [16].

The workshop also introduced artists to interaction-design methods such as storyboarding, rapid sketching, high-level system diagrams, and *Wizard of Oz* techniques for simulating system behaviour [235]. Practical tasks included setting up projection-mapped

scenes, creating gesture-triggered visual changes using JavaScript libraries (e.g., p5.js), and testing interaction concepts using the building’s sensor and display infrastructure.

Across the two days, the researcher provided both technical support and creative guidance – from configuring pose-estimation demos to advising on movement- and voice-led interaction concepts. This role aligns with practice-led approaches, where practitioner expertise provides tacit knowledge that helps participants navigate the creative–computational space.



Figure 3.8: Artcamp session at Swansea, bringing together local and visiting artists in a repurposed creative space. Participants engaged in spoken word, movement, and visual experimentation, using projection surfaces, webcams, and browser-based tools to prototype interactive ideas.

### Minimum Viable Artworks

Adapted from the software industry’s Minimum Viable Product (MVP), Minimum Viable Artworks (MVAs) were deliberately constrained prototypes intended to test early concepts quickly [233, 234]. MVAs drew upon the building’s equipment—projectors, digital screens, microphones, sensors, and web-based ML libraries—and typically combined simple interaction logic with supporting sketches or system diagrams.

Each team produced an MVA within the two-day structure. These prototypes included movement-responsive sketches using PoseNet, interactive projection scenes, audio-reactive

visualisations, and browser-based mini-experiences. Qualitative reflections indicated that participants had developed confidence with computational concepts and were able to articulate technical decisions within their creative practice. Appendix A.1 provides an illustrative MVA process and output.

### **MVA Toolkit**

The MVA toolkit comprised both hardware and software resources configured to be accessible to artists with limited technical experience. The hardware included consumer projectors, webcams, microphones, and large-format screens; the software environment included TensorFlow.js, Phaser, p5.js, BlazePose and PoseNet demos, and standard web browsers. The toolkit was intentionally lightweight and reconfigurable, allowing participants to assemble interactive experiences from readily available components. This approach reflected the broader ethos of Arts ARKADE: to cultivate digital literacy and creative agency using tools that could be easily replicated beyond the venue.

All teams successfully created MVAs within the two-day Artcamp structure, demonstrating rapid uptake of computational thinking concepts. Participants presented prototypes to one another in a concluding critique session, supported by sketches, system diagrams, and recorded demonstrations. These artefacts informed later stages of the Arts ARKADE programme and design decisions.

### Activity 2: Hip Hop Case Study

The Hip Hop for Health (HH4H) project provides a concrete example of the integrative vision behind Arts ARKADE. Developed in collaboration with external partners, HH4H combined freestyle MCing, Tilt Brush movement data, and immersive projection to create an interactive performance space [236]. In this system, gestural input defined the spatial dynamics of the environment, while live vocal performance animated its narrative content. HH4H demonstrates three core contributions of the Swansea study: (i) the multimodal integration of movement, music, and voice; (ii) the fostering of a creative community around a shared platform; and (iii) the situating of freestyle practice within a technologically mediated spatial context. The programme drew directly from the cultural frame of the *hip hop state of mind*, remixing and reimagining the environment as both a physical and poetic canvas [236]



Figure 3.9: Swansea Hip Hop Project workshop session as part of Arts ARKADE. The event transformed a repurposed campus space into a collaborative hub for spoken word, breakdance, and visual design, engaging both local and visiting artists in the spirit of the *hip hop state of mind*.

### 3. Study I: Place as a Hip Hop State of Mind

Category	Description
Artist workshops	Two artist-led sessions and five Artcamp induction workshops introducing computational tools and rapid prototyping.
Hip Hop 4 Health	Two-day programme on hip hop, health, and technology; included talks, performances, and participatory activities.
Volcano Project	Cross-generational performance and digital work for World Cities Day, led by Catharine Bennett.
Rebooked Exhibition	Exhibition on creative technology, reuse, and sustainability using repurposed books.
<b>Overall activity</b>	Combined delivery of workshops, public events, and exhibitions across the Swansea programme.

Table 3.4: Summary of Arts ARKADE Swansea outputs, including public events, artist development activities, and project outcomes (2022–2023).

## Outcomes

The Swansea phase of Arts ARKADE combined artist-led workshops, rapid prototyping sessions, and public-facing events to explore place-based creativity. Central to this was the Artcamp induction programme, inspired by technology-industry models of accelerated skills development. Across its first year, the programme engaged seventeen artists through a mix of development workshops, community projects, and exhibitions, delivering over twenty-four activities in total (Table 3.4). Highlights included the two-day *Hip Hop 4 Health* programme, the cross-generational *Volcano Project* performance, and the *Rebooked* exhibition on reuse and sustainability. These activities established a model for integrating physical placemaking with computational creative practice while highlighting both the opportunities and constraints of participatory infrastructure. Participants reported increased confidence in engaging with computational tools, and several expressed interest in further experimentation.

While the evaluation remained largely qualitative, the Swansea programme involved a mix of artist development workshops, public events, and exhibition activities. Across the first year, these activities engaged local and visiting practitioners and provided opportunities for community participation through performances, talks, and hands-on creative sessions. Together, these outputs positioned Swansea as a grassroots test bed for

technology-enabled placemaking, generating insights about accessibility, skill-building, and community-led experimentation.

### **Limitations**

While the Swansea case study offers valuable insights into participatory infrastructure and community-led experimentation, several limitations should be acknowledged. The sample size was relatively small, and the study did not include formal pre- and post-intervention assessments to measure skill acquisition. The duration of the interventions was limited, which may have constrained the depth of engagement and the sustainability of outcomes. Furthermore, the evaluation relied primarily on qualitative reflections from participants and facilitators, rather than on systematic or longitudinal analysis, which limits the generalisability of the findings.

## **Arts ARKADE London (2023-2025)**

**Vision:** Following the Swansea pilot, Arts ARKADE expanded its presence to London with a flagship site at 33 Regent Street in Piccadilly. Co-developed with The Crown Estate, the project was envisaged as a “centre for urban excellence” designed to inspire the next generation of Londoners via hip hop culture and technology [224]. Situated at the junction of cultural tourism, retail, and nightlife, the West End location gave the project symbolic relevance: extending Swansea’s experimental model of technology-enabled arts into one of the world’s most visible cultural districts [237]. The transformation of a retail site into an urban culture hub reflected both the ambition to stand out in one of London’s busiest pedestrian areas, and the intent to support hip hop–inspired forms of creativity. The vision was described as “to use technology to build a new generation of urban artists - using Augmented Reality, generative AI and spatial computing technologies...whilst supporting the vision of London as the leading global destination for urban culture” [224]. While Swansea focused on prototyping immersive AR and community-led events, the London phase drew on the framework of playable cities [68]. This positioned Piccadilly as a test bed for scaling hip hop–driven placemaking. The London case study explores how this vision was put into practice, and the opportunities and challenges that emerged from its implementation.

## Design and Development

The redesign of the London site followed a practice-led process in which spatial, visual, and cultural elements were co-developed with practitioners from the hip hop community.

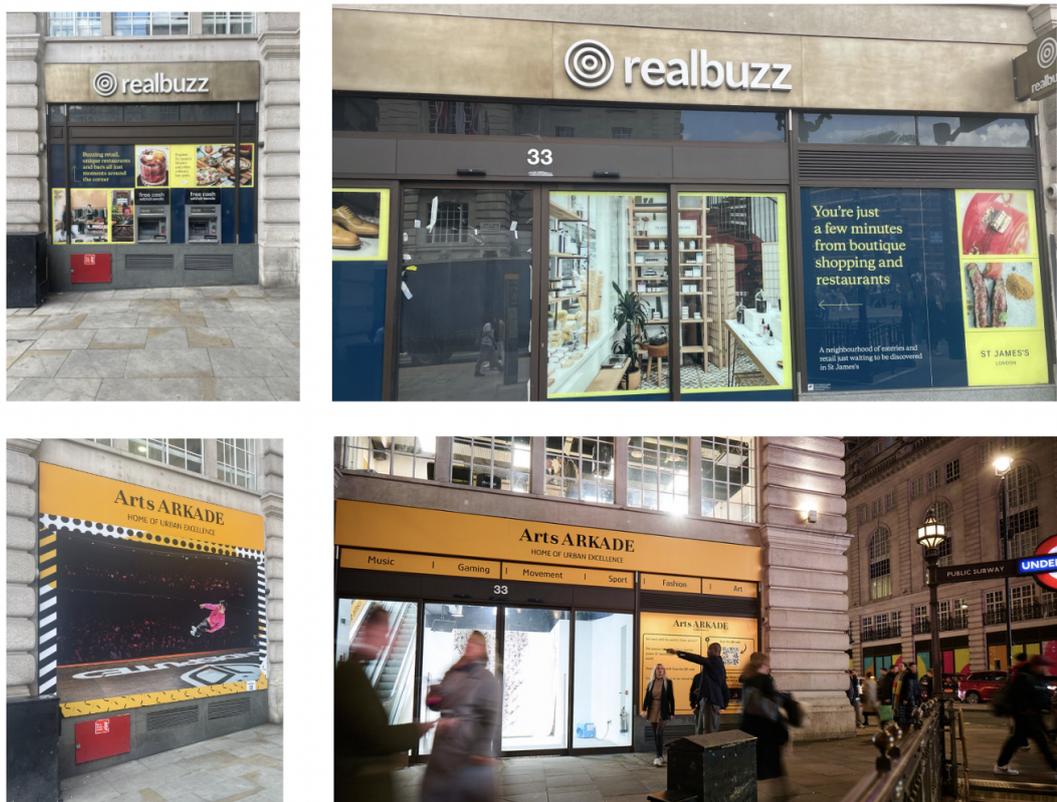


Figure 3.10: Exterior design development for Arts ARKADE London at 33 Regent Street, Piccadilly. Top row: original high-street retail frontage of the unit prior to redevelopment. Bottom row: completed exterior branding, featuring bold yellow-and-black colour scheme and large-format cultural imagery to signal the venue's urban focus. The transformation aimed to position the site as both a recognisable destination for urban excellence and a point of connection between London's global cultural visitors and the local creative community.

Community expertise informed iterative decisions about layout, aesthetic language, and the integration of digital and analogue tools. Figure 3.10 shows the transformation of the building from a disused retail unit.

### 3. Study I: Place as a Hip Hop State of Mind

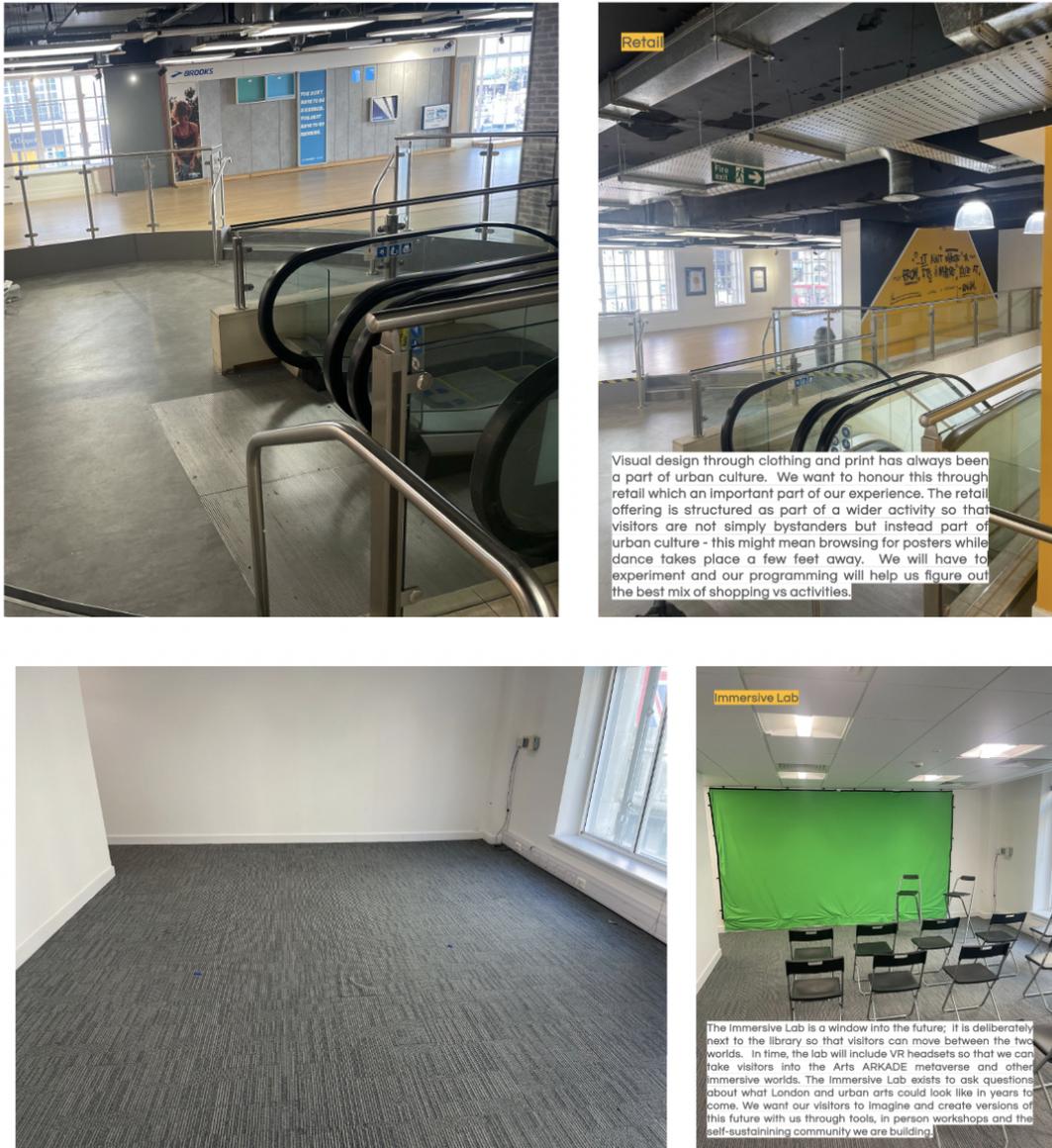


Figure 3.11: **Main and Immersive Spaces.** Top left: main interior prior to redesign. Top right: redesigned space incorporating hip hop visual language and wall text developed collaboratively with community practitioners. Bottom left: immersive space prior to redesign. Bottom right: immersive space after redesign, intended for green-screen work, AR demonstrations, and 3D visualisation. The design work followed a practice-led approach in which spatial form evolved through iterative collaboration with artists and facilitators.

In figure 3.13 the large vinyl poster (top left photo) acted as a trigger for an AR overlay (top right photo). the outside of the building was use as a “digital canvas” that visitors

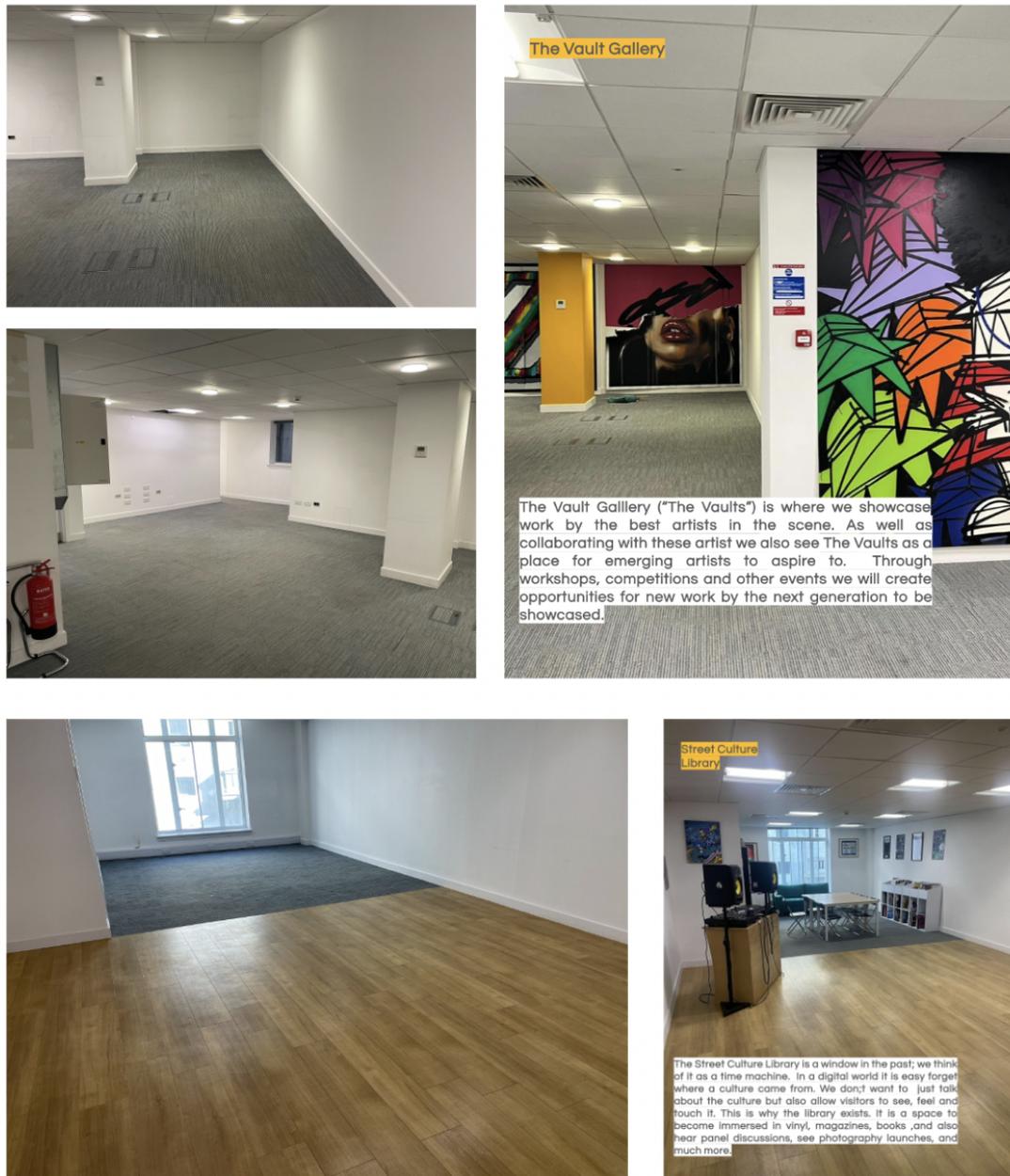


Figure 3.12: **Vault Gallery and Library.** Top row: gallery space before and after redesign. The final design showcased large-scale graffiti and street art, produced in collaboration with leading UK artists, with selective AR elements triggered by QR codes. Bottom row: the Street Culture Library before and after installation, featuring analogue media (magazines, vinyl, turntables) curated with practitioners to reflect hip hop's archival and material culture.

### 3. Study I: Place as a Hip Hop State of Mind

---

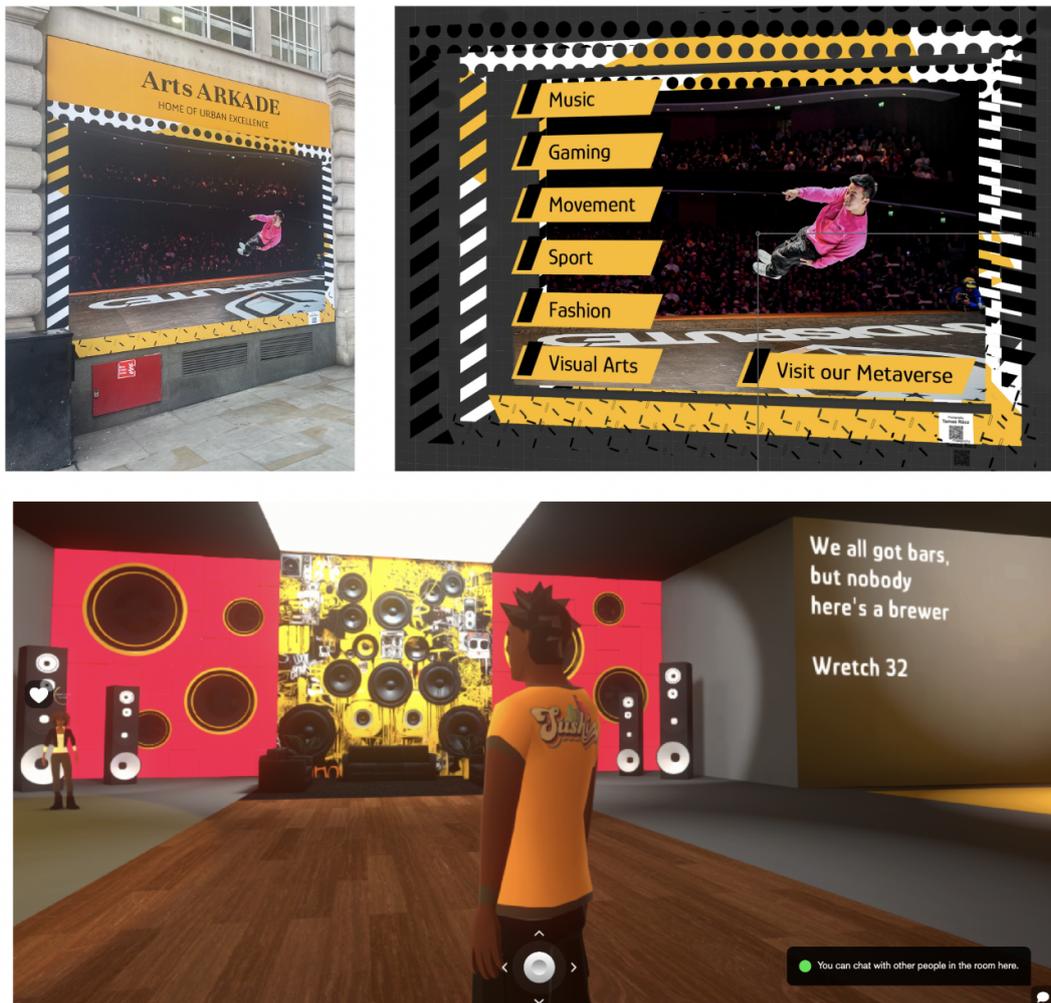


Figure 3.13: **Augmented Reality and Metaverse.** Images show the evolving frontage designs, AR activations, and prototype metaverse environment. These elements were developed iteratively with hip hop creators and technologists to test how digital overlays and virtual environments might extend placemaking into hybrid physical–digital domains. These exploratory prototypes were not implemented within the study but served to scope potential future extensions of the work, particularly in relation to hybrid physical–digital interactions and emerging agent-based systems such as Digital Creative Twins.

could interact with via their phones. Alongside this, a spatial website offered an immersive experience via desktop PCs and VR headsets. The spatial website included artificial characters that visitors to the metaverse could engage with in real time using their voices.

The guiding metaphor of Arts ARCADE was the idea of *place as canvas* not just a site to be occupied, but a layered, adaptable surface for cultural inscription. Through

the reuse of physical infrastructure (e.g., 33 Regent Street), the incorporation of digital overlays (e.g., augmented reality, metaverse integration), and the cultivated embodied practices (e.g., street dance, graffiti, turntablism), Arts ARKADE reconfigured space as an active participant in meaning-making. Design choices were not imposed top-down but evolved through use, creating an iterative feedback loop between community practice and spatial form.

### **Activity 3: Long Dark Tunnel Exhibition**

As part of Study I's investigation into the role of digital and street-level interventions in creative placemaking, the *Long Dark Tunnel* exhibition served as a pivotal example of temporary urban transformation through graffiti, sound, and immersive installation. Hosted at Arts ARKADE's flagship site in Piccadilly, London, from March to April 2025, the exhibition featured renowned graffiti artists 10Foot, TOX, and FUME, whose decades-long contributions to London's visual culture has often crossed the line between illegality and public recognition [238, 239]. The show reconstructed the sensory environment of the London Underground, including a replica graffiti-covered train carriage, motorway signs, archival material, and sound design referencing jungle music and hip hop. The exhibition drew thousands of visitors, and was paired with a special edition of *The Big Issue* guest-edited by 10Foot, which included interviews with TOX and Banksy, further embedding the project in the space of 'grassroots artistic resistance' [240, 241]. The early closure of the show in April, amid controversy unrelated to the internal operation of the exhibition, highlighted the tensions between sanctioned and unsanctioned expression in the heart of London's commercial district [242, 243].

The interior of Arts ARKADE London was extensively refitted to host The Long Dark Tunnel exhibition, transforming the building into a large-scale urban art environment. The space was configured as a "building-as-canvas," with walls, floors, stairwells, and transitional areas treated as surfaces for creative intervention. A full-scale replica of a London Underground train carriage was constructed and installed as the exhibition's central feature; this was subsequently painted live by the lead exhibition artists and a small group of prominent London graffiti writers. Surrounding areas—including mezzanine walkways, escalator wells, and perimeter walls—were covered in custom murals, tags, character work, and installations, creating an immersive multi-level environment reminiscent of metropolitan transit spaces and graffiti subculture.

### 3. Study I: Place as a Hip Hop State of Mind

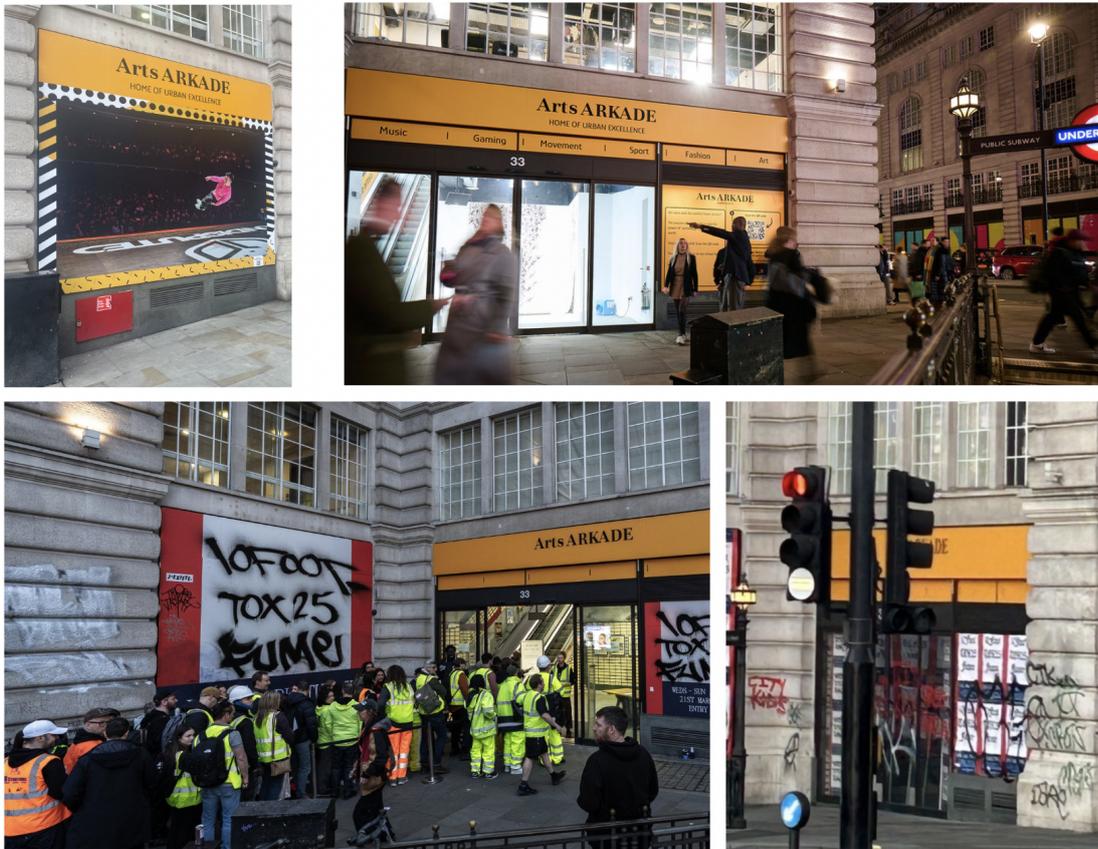


Figure 3.14: **Exterior Branding and Long Dark Tunnel Exhibition.** Modified exterior for event-specific collaboration. Top row: standard yellow-and-black exterior branding during daytime and evening operation. Bottom left: modified frontage for the Long Dark Tunnel exhibition, produced with participating graffiti artists. Bottom right: additional graffiti applied during the exhibition period, which contributed to the early closure of the show. Together these images illustrate the dynamic interaction between designed environment, community practice, and public response.

This refit served two purposes. First, it provided a practice-led example of how urban creative languages can be embedded directly into the fabric of a cultural venue, blurring boundaries between gallery, street culture, and public space. Second, it functioned as an engagement mechanism for urban audiences: visitors navigated an environment that visually and spatially reflected the cultural origins of the artworks, encouraging embodied interaction and reinforcing the exhibition's thematic focus on movement, transgression, and underground creative networks. This approach positioned the building not merely as a container for the exhibition but as an active participant in the creative experience.

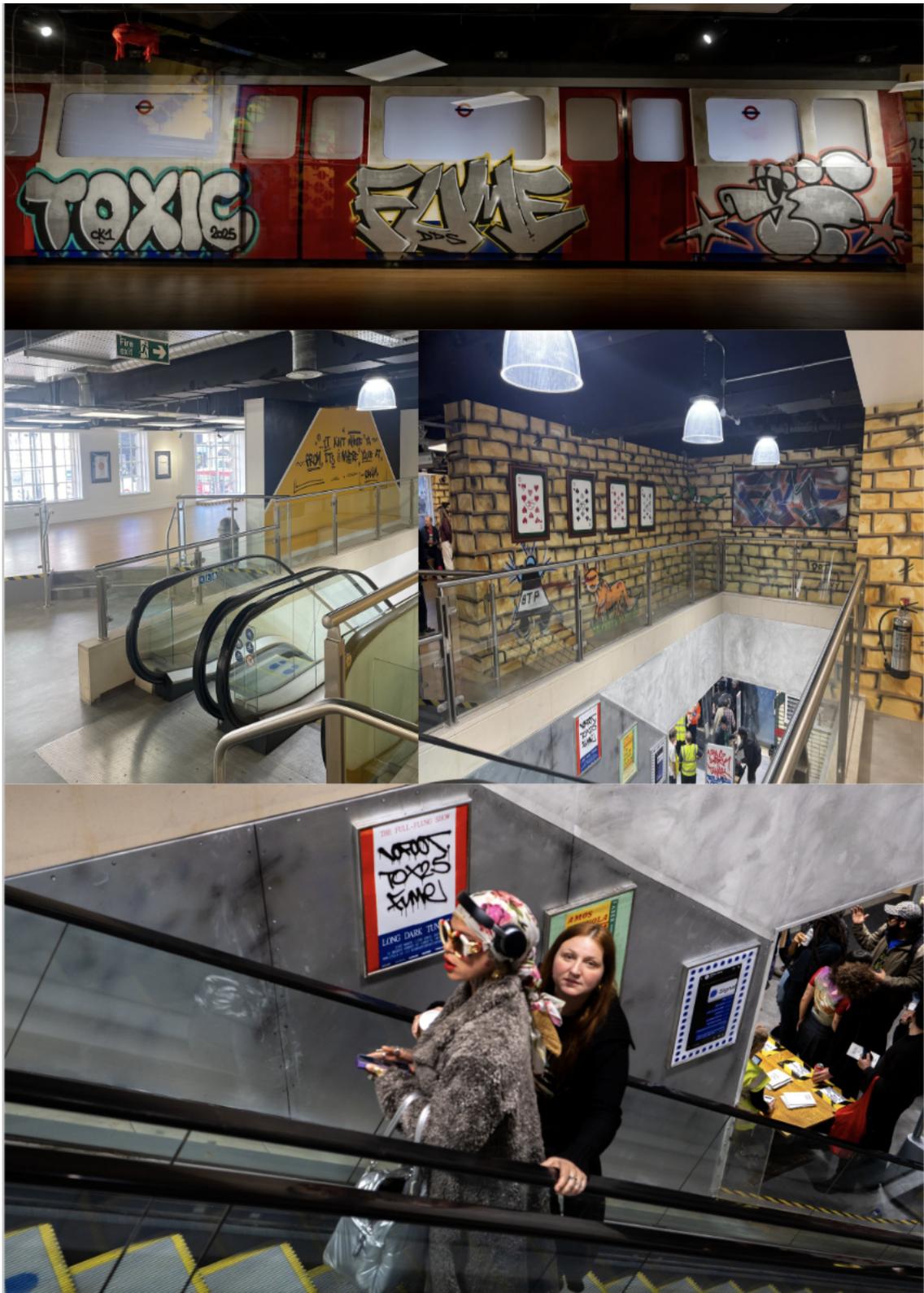


Figure 3.15: **Interior refit for the Long Dark Tunnel exhibition.** Top: full-scale replica London Underground carriage constructed for the show. Middle: (L) original and (R) murals spanning mezzanine walkways. Bottom: escalator redesign illustrating the “building-as-canvas” design approach.

#### Activity 4: Beatsora Rap Cypher

The *Beatsora Cypher*, hosted at Arts ARKADE London, showcased freestyle hip hop as a socially embedded, co-creative practice. The cypher<sup>2</sup> extended the project's use of hip hop as both methodology and community tool. Though not AI-mediated, it embodied key HH4H values—improvisation, shared space, and mutual elevation. It also served as a live counterpoint to the more structured, computational approaches in Studies II and III.



Figure 3.16: Beatsora Cypher at Arts ARKADE London. Top image. Freestyling MC during the Beatsora Cypher at Arts ARKADE London. Bottom image. Close-up of mid-delivery flow. A visual record of the affective intensity central to live hip hop performance.

---

<sup>2</sup>A cypher is a gathering of rappers who come together to freestyle lyrics and feed off each other's energy [244].

## 3.5 Findings

The evaluation of Study I draws on qualitative data including participant feedback, observation, photographic documentation, and artefact analysis. While Swansea provided a richer data set due to the presence of functional infrastructure, London offered insights into the limitations and affordances of urban deployment without full technical support. This section identifies key indicators of success and proposes metrics for future iterations.

The London phase built on the Swansea foundations but operated at a much larger scale. Situated in Piccadilly Circus, Arts ARKADE engaged a far broader public, recording over 6,000 in-person engagements across multiple events and drop-in activities, alongside continuous interaction with Regent Street retail visitors. Programming included regular cyphers, DJ sets, graffiti sessions, book launches, and panel discussions such as *Original London Style*, supported by cultural ambassadors and established hip hop networks. In addition to its physical presence, the London site generated significant media and digital reach, with an estimated combined social media audience of 650,000 through partners and ambassador networks. Thematic highlights included the *Long Dark Tunnel* exhibition, which drew thousands of visitors and demonstrated both the opportunities and tensions of urban placemaking in high-profile settings.

Overall, London scaled up the ARKADE model from a regionally focused, event-based platform to a city-centre cultural hub with sustained programming, international artist collaborations, and daily audience turnover. The location's visibility and retail-level foot traffic created new opportunities for daily interaction. The comparison in Table 3.5 highlights how the project expanded from a regionally focused, event-based model to a city-centre cultural hub.

3. Study I: Place as a Hip Hop State of Mind

Theme	Hypothesis / Assumption	Findings
<b>Scale and infrastructure</b>	Small, low-cost interventions in Swansea would act as a catalyst for local participation.	Confirmed: lightweight interventions (e.g., Artcamp, HH4H) successfully activated participation, but momentum was dependent on local networks. London demonstrated the ability to scale up, with high-footfall visibility and continuous programming.
<b>Technology vs community</b>	Advanced digital infrastructure (screens, AR) would be sufficient to drive engagement.	Partially supported: while immersive screens and AR attracted interest in Swansea, sustained engagement in both sites was shaped more by community networks, ambassadors, and cultural practices than by technology alone.
<b>Cultural framing (hip hop)</b>	Hip hop principles could act as a shared framework for participation across contexts.	Supported: freestyle, cypher, and graffiti practices provided culturally resonant entry points in both Swansea and London, though the expressions varied (grassroots in Swansea; established communities in London).
<b>Portability vs specificity</b>	The ARKADE model would translate consistently across sites.	Mixed: while the curatorial vision travelled, local context (scale, demographics, networks) strongly shaped uptake. Swansea emphasised experimentation and skills development; London emphasised visibility and cultural capital.

Table 3.5: **Hypotheses and Findings.** Summary of the central hypotheses guiding Study I and the empirical observations from the Swansea and London deployments. The table clarifies how each hypothesis links to the research question on place and creative practice.

## **Illustrative Outputs: Arts ARKADE Public Reception (London)**

During the Arts ARKADE London initiative, public feedback highlighted the cultural legitimacy, collaborative potential, and relevance of the space. Comments for the project included:

“This is revolutionary. Another glass ceiling broken for hip hop. Respect!”

“Amazing scope. Let’s bridge the gap between hip hop, street culture, and brands in order to push the culture forward.”

“A unique place. . . this is real hip hop.”

“I am so inspired by what can be done in this space to change our relationship with public spaces.”

“Love the passion, vision, and mission. We are happy to collaborate on making dance and the arts accessible to all.”

Visitors included members of Mo Flow Crew, Hip Hop To Help, and House of Create, as well as prominent figures from UK hip hop and music scenes such as The Ragga Twins, Trevor Nelson, Danny “D” Poku, Normski, Killa Kela, Harry Shotta, and Manga Saint Hilare. These reflections indicate that Arts ARKADE succeeded not only as a site of creative expression but also as a cultural infrastructure that audiences recognised as authentic and overdue. The frequent emphasis on collaboration, flexibility, and ‘real hip hop’ suggests that the model resonated across both artistic and institutional boundaries. This reception aligns with Study I’s broader claim that creative placemaking, when designed through culturally embedded grammars, can reshape how public spaces are valued and imagined. To conclude the chapter, Table 3.6 explicitly maps the reduced set of research questions to the findings from both Swansea and London. This synthesis serves two purposes. First, it demonstrates how each question was addressed in practice, drawing on evidence from across the two case studies. Second, it anchors the comparative dimension of the research, making clear where approaches, strategies, and outcomes aligned or diverged. In doing so, the table functions as a summary of key contributions and a bridge to subsequent chapters, which extend these insights into the exploration of AI-supported practices.

3. Study I: Place as a Hip Hop State of Mind

Research Question	Findings: Swansea Study	Findings: London Study
<b>RQ1-1: How can creative infrastructure support inclusive, digitally enabled cultural practice?</b>	Renovation of a single building into a flexible media arts hub; installation of large immersive screens and AR; open workshops lowering technical barriers.	AR in exterior public space without large screens; regular cyphers, DJing, and dance events; integration of digital tools into established hip hop community spaces.
<b>RQ1-2: What strategies enable artists to engage with unfamiliar technologies through participatory design?</b>	Artcamp induction; rapid prototyping; cross-disciplinary artist collaborations; scaffolding technical skills through hands-on installation work.	Iterative co-design with local artists; translation of Swansea-developed AR concepts into site-specific London contexts; emphasis on adapting tools to outdoor use.
<b>RQ1-3: How do hip hop principles inform placemaking and interactive space design?</b>	Hip Hop 4 Health: cyphers, freestyle rap, dance, and AR installations; integration of movement, voice, and public participation in a health-themed creative framework.	Weekly cyphers and performance-based programming; use of graffiti, dance, and spoken word as co-creative public interventions.
<b>RQ1-4: How does the Arts ARKADE model scale across different cultural and spatial contexts?</b>	Proof-of-concept in Swansea with short-term events and artist residencies; emphasis on building initial community engagement.	Scale-up in London with sustained programming, broader audience reach (6,000+ visitors), and adaptation to high-traffic urban setting.

Table 3.6: **Research Questions to Findings Mapping.** Mapping of Study I's research questions to key findings, showing how the Arts ARKADE deployments generated evidence for understanding place as an active variable in creative practice.

### 3.6 Discussion and Future Work

By designing place as an open, collaborative canvas, Arts ARKADE illustrates how participatory placemaking enables cultural expression across boundaries of age, geography, and medium. This foundation in site-specific creativity serves as a conceptual and methodological springboard for the next study, which shifts focus to linguistic creativity—where writing tools and poetic forms serve as similarly structured canvases of co-creation. This section synthesises the lessons from both sites and articulates the contribution of Study I to the overall thesis. It discusses the feasibility of participatory digital infrastructure, the role of creative practice in skill-building, and the importance of context-specific adaptation. It also reflects on how different urban settings shape what kinds of creativity can emerge.

Since the early 2000s, digital technologies have challenged traditional understandings of public space [215, 245]. The rise of advanced virtual environments has further blurred boundaries between the physical and digital [246]. As computational methods become embedded in the design of urban environments, disciplines such as architecture, media art, and computer science increasingly overlap [6, 217, 247]. Arts ARKADE demonstrates how infrastructure for digital art can support local placemaking efforts. The project's first phase focused on capacity-building for artists through rapid prototyping.

Future work includes developing a more advanced, digitally mediated version of Artcamp. Artists will be encouraged to develop computational thinking skills remotely, allowing in-person sessions to focus on production and iteration. Arts ARKADE also plans to formalise its toolkit into a scalable 'art stack': a curated combination of standardised yet adaptable tools, workflows, and interaction models that artists can use to create urban experiences. This echoes strategies employed by collectives such as teamLab [248]. Study I thus lays the groundwork for understanding how physical space, community engagement, and technological systems intersect to support creativity. Study II builds on these findings by examining how generative AI can participate in creative processes, with a particular focus on poetic composition. Where Study I addressed the infrastructural and spatial dimensions of co-creation, Study II turns toward linguistic and evaluative dimensions.

The Arts ARKADE deployments established how lightweight, in-place interventions travel through local practices. The next chapter turns to *language* as the unit of practice: rhyme and cadence are operationalised to examine how these forms circulate between people and models within the same communities. Within the broader *hip hop state of mind* framework—creativity, remixing, cypher, and spatial reclamation, physical

### 3. Study I: Place as a Hip Hop State of Mind

---

sites like Swansea and London act as cultural canvases where placemaking, spoken word, and rap converge. These spaces provide tangible stages animated by community presence and performance.

The future vision extends these places into augmented, persistent, and metaverse-like environments, creating a continuum from street corner to shared virtual space. In this configuration, the *canvases* of practice becomes responsive rather than fixed, layered with digital augmentation, generative systems, and real-time interaction. Language, spoken in cypher, written in verse, or generated by AI, becomes the medium through which space is continually reimagined. The findings from Study I address RQ1 by demonstrating how hip hop can function as a cultural tool for placemaking in both Swansea and London. Beyond their sociocultural value, these interventions generated direct implications for computational research. Placemaking was reframed not only as a social outcome, but also as a design principle: models of creativity can be situated within urban, cultural, and community contexts rather than treated as context-free abstractions.

The workshops and performances revealed observable markers of collaboration, participation patterns, improvisational exchanges, and stylistic shifts that could be abstracted into measurable proxies, anticipating the quantitative features developed later in the thesis. By grounding computational inquiry in lived creative practice, Study I anchored the overall trajectory: subsequent experiments (Studies II and III) extend these situated artistic dynamics into systematic, computationally testable forms.

In this sense, Study II becomes a natural progression, moving from culturally embedded practice into controlled settings where computational measures of creativity can be refined and evaluated. Poetry then provides a structured yet flexible domain for examining how technology itself can *participate* in co-creation. This aligns with the co-creativity spatiality perspective outlined in Section 2.8, where creative agency is treated as distributed across people, tools, and environments.

### 3.7 Publication and Contribution Note

Elements of this chapter have been presented and discussed in externally refereed and invited contexts:

Olátúnjí, I. (2023). **Arts ARKADE: Digital Placemaking, Urban Arts, and the Future of Public Space**.

AVARA 2023, Centre for Arts Innovation (CAI), Oulu University of Applied Sciences, Finland, 16–17 November 2023.

Olátúnjí, I. (2024). Panel contribution: **The Crown Estate: Together '24**. London, UK, 9 May 2024.

Olátúnjí, I. (2024). Panel contribution: **The Imagination Playground**. The Evening Standard and Samsung, London, UK, 12 September 2024.

Olátúnjí, I. (2025). Panel contribution: **Digital Creativity and Placemaking**. RGS–IBG Annual International Conference, Birmingham, UK, 26–29 August 2025.

Olátúnjí, I. (2025). **Arts ARKADE: Digital Placemaking, Urban Arts, and the Future of Public Space**. Accepted for presentation at the RGS–IBG Annual International Conference, Birmingham, UK, 26–29 August 2025 (to appear).

The author led the research design, fieldwork, and analysis of Arts ARKADE in Swansea and London. Collaborators played an important role in implementing aspects of the design and contributing expertise and discussion. These outputs directly informed the study's design, interpretation, and conclusions.

### Study I — Findings Summary

**Hypothesis.** Lightweight, in-place creative interventions (such as those deployed through Arts ARKADE) can scaffold locally embedded participation and support the formation of creative identity in public space.

**Evidence.** Fieldwork in Swansea demonstrated that digital and physical provocations enabled low-barrier engagement, local authorship, and created opportunities for creative exploration without requiring permanent infrastructure.

**Interpretation.** These interventions operated as “minimum viable artworks,” allowing public space to become a site of rehearsal for new kinds of cultural practice. Some effectiveness in permanence their ability to circulate through local practices.

**Limitation.** While effective as probes, the interventions depended heavily on local networks and struggled to sustain momentum. Also, study limited to Swansea in terms of technology enabling. The London study focused on cultural practice.

**Future work.** Study II explores how language itself, via rhyme and poetic form, can act as a scaffold for evaluating creative co-production.

## Chapter 4

# Study II: Poetry and Writing Assistants



Figure 4.1: More Poetry is Needed. Originally a huge mural created in 2014 at the back of the Quadrant shopping centre in Swansea, as part of the celebrations for Welsh writer Dylan Thomas's centenary. The stark white-on-black format, the stern serif font, and the full stop give the sentence the air of an order rather than an invitation. *Jeremy Deller, 2014 [249]*.

## 4.1 Introduction

Building on Chapter 3, which examined how placemaking shapes cultural and creative expression, this chapter focuses on poetry and lyric writing.<sup>1</sup> It investigates how computational tools are reshaping contemporary creative practice by evaluating how different categories of writing assistants influence both the process and the outputs of creation.

Extending placemaking into the linguistic domain, the study uses Swansea and London as contextual scaffolds for written expression. It also introduces a quantitative lens on poetic form, specifically rhyme density, to enable comparisons between human and machine-generated verse. Through structured writing tasks, participants engaged with a range of AI tools, offering insight into how such systems may support, constrain, or augment individual creative expression. Poetry is chosen here not only for its cultural significance but also for its compression of meaning, structural precision, and measurable constraints such as rhyme and metre. These qualities make it a rigorous boundary domain for evaluating the expressive capacities of AI writing systems. Figure 4.2 reuses the Venn diagram, with poetry and generative AI brought to the foreground. This marks the thesis's shift from physical interventions to word-based tools. The study focuses on poetic forms that involve rhyme and structural constraint, providing a measurable context for evaluating how computational assistance shapes stylistic decisions and creative outcomes. While poetry encompasses a wide range of non-rhyming forms, this study focuses specifically on rhymed verse. This aligns the analytic frame with hip hop and rap traditions, where rhyme, particularly imperfect and multisyllabic rhyme, functions as a core stylistic device and supports later comparison between poets and MCs as practitioners who work within rhyme-based creative expression. This emphasis also follows the discussion in the Literature Review, which situates rhyme as a meaningful site for analysing stylistic and structural variation (see Section 2.6). Poetry, with its compression of meaning and formal precision, thus serves here as a boundary domain in which the expressive capacities of AI systems can be evaluated with rigour.

This inquiry is situated within the broader historical development of assistive writing technologies, tracing their evolution from basic digital aids (such as dictionaries and thesauri) to sophisticated AI-driven systems capable of autonomously generating poetic compositions. The study adopts a framework that categorises writing tools into three

---

<sup>1</sup>Except where otherwise specified, poetry and lyric writing are used interchangeably in this chapter. Chapter 5 introduces a clearer distinction through its focus on lyric performance.

## Hip Hop as Computational Tool

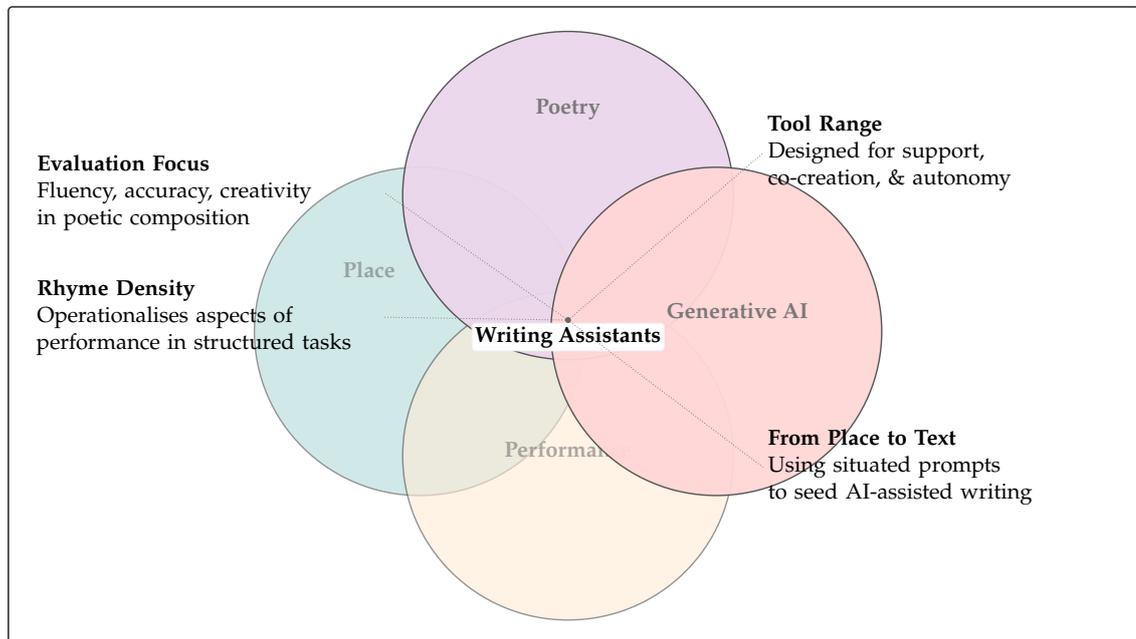


Figure 4.2: **Study II framework: Poetry and Generative AI** as foreground via writing assistants. Rhyme density is introduced as a proxy to evaluate accuracy and creativity.

primary types; support tools, co-creative tools, and autonomous tools, each of which offers varying degrees of influence over the writing process. The study then explores how engagement with these tools differs across participant experience levels, addressing how creative expertise mediates tool use and creative outcomes.

### Relation to Primary Research Questions

Study II directly addresses the two primary research questions in the Introduction (section 1.2):

**(RQ2) Tools and Process.** How do different classes of writing tools (support, co-creative, autonomous) affect fluency, accuracy, and perceived creativity in poetic composition?

**(RQ3) Proxies and Evaluation.** In what ways can textual and performance measures (e.g., rhyme density, cadence similarity) serve as *interpretable* proxies for co-creativity?

The subsidiary research questions presented later in this chapter elaborate these two core domains by examining patterns in creativity assessment, user expertise, creative output, and user experience.

### **Practitioner–Researcher Framing**

In Study II, practitioner expertise <sup>2</sup> informed the design of the creative tasks and the development of computational evaluation measures. Practitioner-researchers draw on tacit and procedural knowledge to identify salient features of creative processes that may be less visible through quantitative analysis alone [16]. This perspective also helped shape several aspects of the study’s construction. Expertise in poetic form, rhyme structure, and lyrical composition supported the design and iterative testing of the rhyme density scoring method, including the calibration of thresholds and detection approaches. Practitioner expertise also informed the design of the linked experimental tasks, for example incorporating DAT outputs as creative constraints and aligning the task structure with established approaches to creativity research.

---

<sup>2</sup>See Section 1.3 for specific details.

A distinctive feature of the study is its integrated task design, which combines both divergent and convergent creative processes. Participants first complete a divergent thinking exercise, the DAT and then use their DAT-generated words as constraints in subsequent writing tasks. This structure enables a novel, joint evaluation of idea generation and structured poetic composition, bridging product and process-based approaches to creativity. In parallel, the study introduces rhyme density as a quantifiable metric of stylistic complexity. By calculating rhyme scores across human and machine-assisted texts, the research provides a replicable measure to assess differences in creative style and output.

While recent advances in AI writing assistants have enabled widespread experimentation with co-creative systems, relatively few empirical studies have investigated how different tool types influence the creative process itself—particularly in relation to user experience, stylistic alignment, and perceptions of creativity. Moreover, there remains limited methodological integration of divergent and convergent creative phases in empirical research on human–AI collaboration. This study addresses these gaps directly.

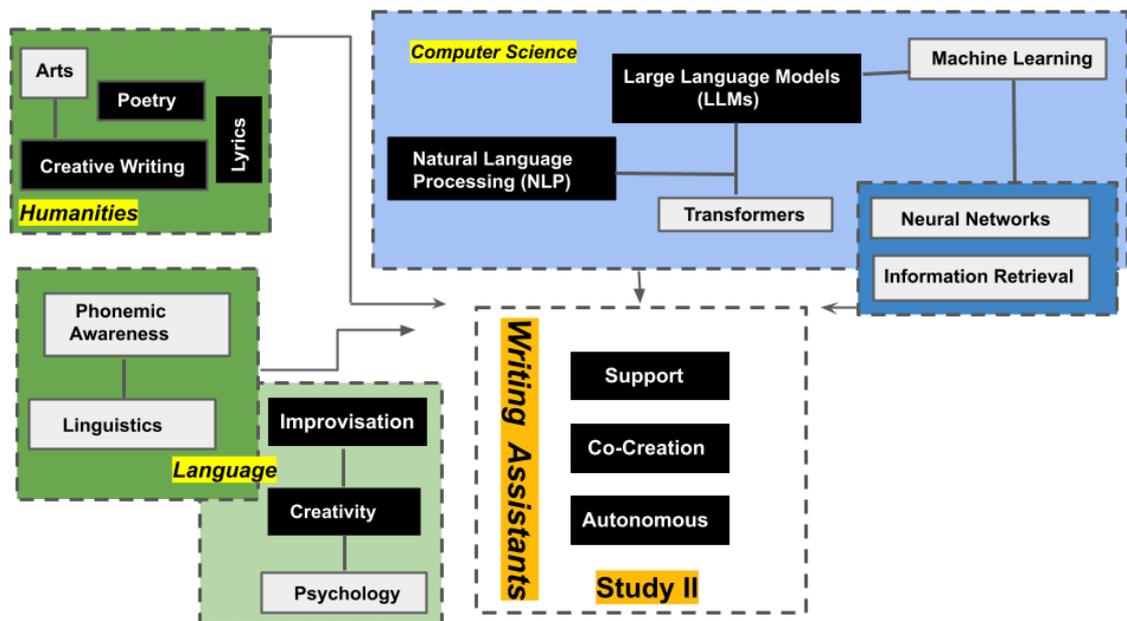


Figure 4.3: **Thematic Map of Key Topics:** Overview of Study II, mapping the chapter's structure from research questions to experimental design, participant tasks, and results, showing connections between creativity theory, computational writing tools, poetic structure, and user experience evaluation

#### 4. *Study II: Poetry and Writing Assistants*

---

Figure 4.3 situates the study within its broader interdisciplinary context. The research draws on insights from the humanities (poetry and creative writing), language and phonemic awareness, improvisational psychology, and technical advances in NLP and LLMs. These influences converge in the taxonomy of writing assistants explored in this chapter: support, co-creative, and autonomous tools.

By mapping these disciplinary intersections, the diagram clarifies how different theoretical and technical strands inform the evaluation of AI-supported creative writing. It also reflects the study's aim to investigate not just tool output quality, but how users with varying creative backgrounds and expectations engage with different types of AI writing systems.

## 4.2 Related Work

### The Evolution of Assistive Writing Technologies

The development of writing technologies has progressively shifted creative labour from manual inscription to machine-supported ideation. Early tools emphasised manual control and material constraints, while later innovations such as typewriters, word processors, and grammar checkers introduced new forms of procedural assistance [250]. Today, LLMs mark a further shift, where systems can autonomously generate content from minimal prompts. This trajectory reflects the increasing externalisation of creative processes, moving from internal (writer-driven) to shared (tool-mediated) and now to predictive (machine-generated). As Schmidt observes, “Parchment elicits one attitude from a writer, paper another. The textures are part of the equation. A quill, a biro and a keyboard download a poet in different ways, at different speeds” (p. 6) [10]. Such observations underscore how each new technology alters not only creative output but also the embodied practices of writing.

Phase	Examples	Key Features
Word Processors	Microsoft Word, Word-Perfect	Digital editing, layout control, spellchecking
Writing Software	Grammarly, Scrivener	Real-time feedback, structural and stylistic suggestions
AI-Powered Tools	GPT-4, Bard, LLaMA-2	Generative capabilities, contextual awareness, adaptive style

Table 4.1: The evolution of digital and AI-powered writing technologies.

### Categories of Writing Assistants

Karimi et al. [251] offer a taxonomy based on human involvement in the creative process: (1) creativity support tools (e.g., rhyming dictionaries [252]), (2) co-creative systems (e.g., Verse by Verse [253]), and (3) autonomous systems (e.g., ChatGPT [183]) which operate independently. This classification informs the design of the current study.

Tool Type	Description	Examples
Creativity Support Tools	Offer task-specific assistance (e.g., rhyme lookup, thesauri).	Rhymer, Grammarly
Co-Creative Systems	Collaborate during composition by offering partial suggestions.	Verse by Verse, Deepbeat
Autonomous Systems	Generate full outputs from prompts, without real-time user input.	GPT-4, Bard, LLaMA-2

Table 4.2: Categories of computational writing assistants.

A further class of *advisory systems*, which provide stylistic or structural feedback without generating new text, could extend this framework. These systems are of particular interest for evaluation or educational use cases.

## Examples of Writing Tools

*Creativity Support Tools:* Assist with sub-tasks such as rhyme generation or synonym lookup. Typically embedded in general writing environments and facilitate targeted language enhancement.

*Co-Creative Systems:* Tools such as *Verse by Verse* or *DeepBeat* offer interactive suggestions that users can adapt. Their purpose is to stimulate novel connections or perspectives during the writing process.

*Autonomous Systems:* LLMs such as GPT-4 and Bard generate entire poetic texts with minimal prompting. While powerful, their contributions may obscure authorship and limit user agency.

## User-Centred Design in Co-Creative Writing

Human-centred design principles suggest that writing tools must accommodate variation in user skill, style, and intent [158, 254]. Novices may benefit from structured support, whereas experienced users often prefer flexible tools that preserve authorial control [159, 255].

Buschek et al. [256] identify three common pitfalls in co-creative system design: (1) systems that do too little, (2) systems that do too much, and (3) neglect of use-contexts. This study draws on a simplified version of the co-creative design challenges proposed by Buschek et al., focusing on three core tensions that are most relevant to writing tool evaluation in Study II. A corresponding table summarising these tensions is provided below. A more comprehensive engagement with the full set of nine challenges, including interface design, user control, and social collaboration will be addressed in Study III and outlined in the Future Work section.

Design Tension	Description	Design Implication
Too Little AI	System provides limited suggestions or feedback.	Users may disengage or perceive system as unhelpful.
Too Much AI	System dominates composition or overwrites input.	Risks loss of agency and authorship.
Context Neglect	Tool ignores writing phase or user goal.	Reduces creative alignment and task fit.

Table 4.3: Co-creative system design tensions, adapted from Buschek et al. [256].

Yang et al. [257] demonstrate that users' expectations and mental models shape how they interact with AI systems. Transparent behaviour and iterative feedback loops can support trust and sustained creative engagement. These insights directly inform the study's tool categorisation and evaluation strategy.

### **Cognitive and Experiential Dimensions of Creative Writing**

Creative writing involves multiple stages: idea generation (divergent thinking), synthesis (convergent thinking), and revision. These are supported by cognitive frameworks that structure how writers process and transform ideas [20, 21, 22, 28].

This study integrates these principles by combining the DAT with a structured poetic composition task. This dual-stage design supports analysis of both generative and integrative creativity. Prior work on improvisation suggests that writers vary in real-time fluency and generative flexibility depending on their skill level and familiarity with the task domain [115, 211]. More skilled or experienced writers draw on stylistic knowledge and genre familiarity [23, 207], while novice writers often depend more on external support [24, 258]. DAT scores have been found to correlate with verbal creativity and problem-solving abilities [34, 259]. In this study, DAT performance provides a baseline to examine variation in tool-assisted outcomes.

To better reflect natural writing behaviour, this study distinguishes between *internal* and *external* activities. Internal activities involve direct engagement with the writing interface, such as typing, editing, or selecting content [24, 260]. External activities refer to broader mental and contextual processes, such as reflection [261], planning, or *mind wandering* [262, 263]. While not directly observable, external processes likely shape poetic choices [10]. Recognising their role enables a more holistic interpretation of participant performance, especially in asynchronous, self-paced tasks.

### **Rhyme Density**

Rhyme was treated in this study as a key stylistic feature that poets and lyricists often articulate intuitively but seldom quantify. By formalising rhyme density as a measurable signal, the study enabled structured comparison between human- and AI-generated texts, supporting a more systematic evaluation of form. To assess compositional structure and stylistic creativity, rhyme density was computed across all writing outputs. This metric allowed consistent comparisons across tool conditions by accounting for differences in output

length and stylistic variation. Building on prior work [127, 139, 205], rhyme was defined as the repetition of similar phonemes, detected via cosine similarity between phoneme embeddings. This method captured both exact (*perfect*) and approximate (*slant*) rhymes.

Phoneme-based analysis offered an advantage over surface-level features such as orthographic resemblance or syllabic count, which often fail to reflect actual auditory rhyme. It also ensured consistency with Study III (see Chapter 5), where phonemic variation played a central role in analysing rap lyrics. In this study, rhyme was treated not only as a stylistic feature but also as a proxy for creative effort. When combined with the DAT, a measure of divergent thinking, rhyme density provided a dual perspective on participants' creative outputs.

### Operational Approach.

In the study design, rhyme was operationalised as phonemic similarity, measured via vector embeddings rather than orthographic resemblance or semantic association. For two phoneme vectors  $\vec{v}_i$  and  $\vec{v}_j$ , similarity was defined as:

$$s_{ij} = \cos \alpha = \frac{\vec{v}_i \cdot \vec{v}_j}{|\vec{v}_i| |\vec{v}_j|} \quad (4.1)$$

where  $\alpha$  was the angle between  $\vec{v}_i$  and  $\vec{v}_j$ .

A match  $M_{ij}$  between phonemes  $p_i$  and  $p_j$  was defined as:

$$M_{ij} = \begin{cases} 1 & \text{if } s_{ij} \geq \theta \\ 0 & \text{otherwise} \end{cases} \quad (4.2)$$

Rhyme density  $D$  was then calculated as:

$$D = \frac{\sum_{i \neq j} M_{ij}}{|P|} \quad (4.3)$$

where  $P$  was the set of all phonemes in the text and  $|P|$  its cardinality. Self-comparisons ( $i = j$ ) were excluded to avoid trivial matches. These formal definitions were tested against poetic data during pilot runs, which highlighted the need for careful thresholding to balance sensitivity and specificity.

In this study,  $\theta$  was set to 0.8. Threshold selection was determined through pilot analyses, balancing sensitivity and specificity, and informed by prior expertise in rhyme analysis.

### 4.3 Research Aims

The aim of Study II was to examine how different classes of AI-assisted writing tools influence poetic composition, user experience, and measurable indicators of creative output. To operationalise this aim, a set of guiding research questions was developed.

#### Guiding Research Questions

Study II is structured around the thesis's primary research questions on **Tools and Process** (RQ2) and **Proxies and Evaluation** (RQ3), which investigate how writing technologies influence poetic composition, and how computational measures can meaningfully capture aspects of co-creativity. To support the study's mixed-methods design, each primary question was expanded into a set of subsidiary questions (RQ2-1 to RQ2-4 and RQ3-1 to RQ3-4), addressing tool behaviour, user experience, stylistic outcomes, and the interpretability of textual and performance-based metrics. A cross-cutting question examines whether divergent thinking ability (as measured by the DAT) influences tool preference or proxy performance. These questions are grounded in the interdisciplinary literature reviewed in Chapter 2, spanning computational creativity, co-creative system design, poetic form, and human–AI creative collaboration. Table 4.4 summarises the full set of questions alongside their conceptual considerations, providing a clear foundation for the study's methodological and analytical choices.

<b>Research Questions (Primary + Subsidiary)</b>	<b>Key Considerations</b>
<b>RQ2: Tools and Process</b>	
<b>RQ2-1.</b> How do different tool classes (support, co-creative, autonomous) influence fluency, accuracy, and perceived creativity?	Output quality; stylistic variation; user perception
<b>RQ2-2.</b> How do writing tools shape user engagement and focus during composition?	Flow state; creative attention; task duration
<b>RQ2-3.</b> How does prior writing experience affect interaction with different tool types?	Autonomy vs. scaffolding; adaptability
<b>RQ2-4.</b> Do novice and expert users evaluate tool usefulness differently?	Expectations; interpretability; perceived utility
<b>RQ3: Proxies and Evaluation</b>	
<b>RQ3-1.</b> In what ways can rhyme density capture meaningful variation in poetic form?	Interpretability; stylistic sensitivity; task specificity
<b>RQ3-2.</b> Can cadence similarity act as a proxy for performative co-creativity?	Performance metrics; rhythmic structure; sonic alignment
<b>RQ3-3.</b> How do subjective user ratings relate to computational measures?	Human-machine alignment; evaluative transparency
<b>RQ3-4.</b> How do textual and performance proxies differ in their sensitivity to user experience and tool class?	Proxy stability; comparative interpretability; multimodal nuance
<b>Cross-Cutting Question</b>	
Does divergent thinking ability (DAT) influence tool preference or proxy performance?	Cognition-tool fit; creativity profiles

Table 4.4: Primary and subsidiary research questions guiding Study II, with associated conceptual considerations.

## 4.4 Method

The study employed a mixed-methods design combining qualitative and quantitative analyses. Data collection involved a pre-study questionnaire, the DAT, four structured writing tasks, and post-task questionnaires. This approach allowed for a holistic examination of how different types of writing assistants influenced creative writing performance. Although a full discussion of limitations is presented later, it is important to note here that the online and asynchronous nature of the study introduced variation in participant environments, that writing tasks were presented in a fixed sequence, and that no strict time limits were imposed. These choices were deliberate trade-offs to balance feasibility, ecological validity, and participant comfort.

**Ethics.** All participants gave informed consent, and the study ensured anonymity and data protection in line with current ethical guidelines. A subset of participants recruited through a university research call received a £20 Amazon voucher as compensation for their time, while others participated voluntarily without financial reward.

**Pre-study Question.** Participants were asked to state the level of their previous writing experience, and age (if comfortable disclosing).

**DAT.** As a baseline measure of creative fluency, participants completed the DAT by generating ten unrelated words intended to maximise semantic distance. The task took under four minutes and provides a concise index of divergent thinking [34]. Scores reflect the average semantic distance between word pairs. Higher scores have been correlated with creative potential [259].

**Writing Tasks.** Participants completed four writing tasks, each prompted by one of the following:

- “What does the place you live mean to you?”
- “What is your favourite location in the place you live?”
- “How does the place you live make you feel?”

Participants were instructed to incorporate as many DAT-generated words as possible and to use rhyme. The task sequence was as follows:

- *Baseline Task*: Writing without computational assistance.
- *Support Tool Task*: Using tools for specific challenges (e.g., rhyme suggestions).
- *Co-Creative Tool Task*: Using tools that act as collaborators.
- *Autonomous Tool Task*: Using tools capable of full text generation.

Participants tracked time spent and evaluated the usefulness of each tool.

**Post-study Questionnaire.** Participants rated each tool across dimensions such as fluency, creativity, accuracy, and usefulness.

**Evaluation Approach.** The study combined quantitative and qualitative evaluation methods. Quantitatively, rhyme density and DAT scores were used to measure changes in stylistic output and creativity across conditions. Qualitatively, open-text reflections were examined using descriptive, pattern-oriented coding to identify recurring observations (fluency, authorship, tool behaviour, and creative flow). This interpretive review was intended to surface patterns participants highlighted in their written feedback. Together, these methods enabled a holistic assessment of how writing tools shaped both measurable poetic features, and subjective creative experience.

#### 4. Study II: Poetry and Writing Assistants

---

Study Component	Description
Study Design	Mixed-methods approach combining questionnaires, creative writing tasks, and computational metrics.
Participants	<b>N = 30</b> , recruited through online platforms and direct outreach, including to creative professionals.
Pre-study Questionnaire	Demographic and experiential data, familiarity with AI, creative attitudes.
Divergent Association Task (DAT)	Ten-word input to assess divergent thinking. Used as a baseline creativity measure.
Writing Tasks	Four structured writing tasks: <ul style="list-style-type: none"><li>• Baseline (no tools)</li><li>• Support Tools</li><li>• Co-Creative Tools</li><li>• Autonomous Tools</li></ul>
Post-study Questionnaire	Fluency, creativity, accuracy, and usefulness ratings.
Supplementary Question	Public perception of AI vs. human authorship.
Scoring	DAT scores, rhyme density, qualitative feedback.
Ethical Considerations	Informed consent, anonymisation, data governance.

Table 4.5: Summary of Study II Design and Methodology.

### 4.5 Technology Stack

**Writing Tools.** To ensure transparency and replicability, the technologies used in the study are summarised in Table 4.6, including their functional category and public access links. The selection spans a spectrum of computational autonomy, from lightweight support utilities to fully generative systems. Participants were encouraged to use each tool freely within its category in order to explore differences in creative control, fluency, and stylistic variation. All model-assisted tasks were conducted between 7 August and 9 October 2023, during which the LLMs were accessed via their publicly available interfaces. Because model providers update systems continuously and do not consistently expose internal version identifiers, the study follows standard reproducibility practice by documenting the operational timeframe. Experiment replication can therefore target the closest available model versions corresponding to this period.

Category	Tools Used	Function / Description and Access Link
<b>Support</b>	Poetry Analysis Engine, RhymeZone, Rhymmer	Suggest rhymes, structure, or lexical alternatives while the user authors text. <i>Links:</i> <a href="http://demo.pae.daladi.org">demo.pae.daladi.org</a> , <a href="http://rhymezone.com">rhymezone.com</a> , <a href="http://rhymmer.com">rhymmer.com</a> .
<b>Co-Creative</b>	Verse by Verse, DeepBeat	Suggest lines interactively or collaboratively with user input. <i>Links:</i> <a href="https://research.google/versebyverse">research.google/versebyverse</a> , <a href="http://deepbeat.org">deepbeat.org</a> .
<b>Autonomous</b>	ChatGPT, Bard, LLaMA-2, Koala	Generative LLMs producing complete texts without user intervention. <i>Links:</i> <a href="https://chat.openai.com">chat.openai.com</a> , <a href="https://bard.google.com">bard.google.com</a> , <a href="https://huggingface.co/blog/llama2">huggingface.co/blog/llama2</a> , <a href="https://chat.lmsys.org">chat.lmsys.org</a> .

Table 4.6: Writing tools used in Study II, grouped by autonomy level with direct access links for reproducibility.

## Task Design: Procedure

Each participant first completed the Divergent Association Task (DAT) as a baseline indicator of divergent thinking. They then undertook four short online writing tasks: (1) a baseline poem without computational assistance, followed by (2) a support-tool condition, (3) a co-creative-tool condition, and (4) an autonomous-tool condition.

All tasks were completed remotely on participants' own devices. The study therefore adopted an interface-neutral design: participants accessed the tools directly through their standard web browsers, without a bespoke interface or time constraints. For each writing task, participants were asked to produce a short poem or lyric in response to one of three thematic prompts on "place." They were instructed to incorporate as many of their DAT-generated words as they felt appropriate and to include rhyme. What participants "saw" depended on the tool: support tools offered word, or line-level suggestions, co-creative tools produced interactive continuations, and autonomous systems returned full poems or near-complete stanzas.

#### 4. Study II: Poetry and Writing Assistants

---

After completing each poem, participants entered their DAT responses, the text they produced, and the time spent on the task (self-reported) into a structured Google Form. A separate post-study questionnaire collected subjective ratings of fluency, creativity, accuracy, and perceived usefulness for each tool class. No fixed session length was imposed, but most participants completed the full procedure within approximately one hour.

### 4.6 Participants

Participants were recruited through the researcher's networks, existing writing groups, and open calls. Thirty individuals took part in Study II, representing a wide span of writing backgrounds and ages (21–61). Experience levels were categorised as high, medium, or low, based on (a) years of writing practice, (b) professional exposure, and (c) regularity of writing.

High-experience participants included professional journalists, librettists, translators, and publishers (7+ years professional or sustained practice). Medium-experience writers had 1–5 years of regular creative writing (e.g., songwriting, informal poetry). Low-experience participants wrote infrequently or had only school-level or early hobby exposure.

A small number of participants from Study I continued into Study II, reflecting ongoing engagement with the researcher's creative community.

Category	Count	Examples
Professional writers	10	Journalists, copywriters, translators, publishers
Experienced hobbyists	8	Songwriters, poets, mixed creative writing
Early/occasional writers	7	Students, informal writing groups
Novices	5	No prior experience beyond school
Gender	18F / 12M	–
Age range	21–61	Median 45

Table 4.7: Summary of participant demographics and writing experience for Study II.

### Analysis Approach

The study combined quantitative and qualitative analyses to evaluate how different writing tools influenced poetic composition. Quantitatively, statistical models were applied to examine relationships between DAT scores, rhyme density, task performance,

and subjective ratings (see Section 4.7 for full details). Qualitatively, participants' open-text reflections were examined using descriptive, pattern-oriented coding to identify recurring observations about fluency, authorship, tool behaviour, and creative flow. This interpretive review was not intended as a full thematic analysis, but rather as a structured summary of the patterns participants highlighted in their written feedback. This dual approach enabled triangulation across computational outputs, participant experience, and self-reported engagement.

## 4.7 Results and Analysis

This section presents the results of Study II. Quantitative analyses explore how creativity-related traits such as divergent thinking and experience level influence writing performance across different tool conditions. Rhyme density is used as a measurable proxy for compositional complexity, and is paired with DAT scores to examine divergent–convergent creativity dynamics. Subjective user evaluations of tool usefulness, fluency, and satisfaction are analysed alongside objective performance. Qualitative feedback and an exploratory perception question complement the statistical findings.

### Rhyme Density Scoring

To quantify compositional structure and stylistic creativity, rhyme density was calculated for all writing outputs. This phoneme-based metric normalises for output length and allows comparison across tool types and participant groups. Raw rhyme density values, produced by the algorithm in rhymed phonemes per 1,000 phonemes, were divided by 1,000 to yield a proportion between 0 and 1, indicating the fraction of phonemes in a verse that participate in rhymes. Within the results, rhyme density serves as the dependent variable in several statistical analyses.

### Quantitative Analysis

A range of statistical models were employed to analyse relationships between participants' creativity scores, writing performance, and evaluation ratings. Table 4.8 summarises them.

Analysis Type	Rationale	Variables/Factors	Significance Level
Pearson Correlation	Measure linear relationships between creativity (DAT), writing performance (rhyme scores), and experience	DAT scores, Rhyme Scores, Experience Levels	$\alpha = 0.05$
ANOVA (One-way)	Test differences in perceived usefulness based on experience levels	Experience Groups vs Tool Usefulness Ratings	$\alpha = 0.05$
OLS Regression	Predict subjective ratings from objective measures	DAT scores, Base Rhyme Scores $\rightarrow$ Tool Usefulness	$\alpha = 0.05$
Subgroup Comparison	Explore questionnaire differences by experience/DAT/rhyme groups	Grouped traits vs Subjective Ratings	$\alpha = 0.05$

Table 4.8: **Statistical Analyses Overview.** Overview of the statistical tests used in Study II, explaining how quantitative analyses (e.g., correlations, regression models, group comparisons) support the evaluation of tool effects on fluency, accuracy, and creativity.

Four primary quantitative methods were applied to assess relationships between creativity measures, writing performance, and user experience. Each analysis type addressed a specific research focus:

- **Pearson Correlation** examined linear relationships between DAT scores, experience, and rhyme density.
- **One-Way ANOVA** tested whether experience level significantly affected writing performance or tool evaluation.
- **OLS Regression** assessed how DAT and rhyme scores predict user ratings of tool usefulness.
- **Subgroup Comparisons** explored variation in responses across participant profiles (e.g., high/low creativity or experience).

### Correlation Analysis

Pearson correlations were used to test associations between DAT scores, experience level, and rhyme density across task types.

4. Study II: Poetry and Writing Assistants

Characteristic	Total	Mean	Minimum	Maximum
<b>Participants (N = 30)</b>				
Low Experience (1)	21	–	–	–
Medium Experience (2)	2	–	–	–
High Experience (3)	7	–	–	–
<b>Measures (SD)</b>				
Age (19.39)	–	29.60	12	61
DAT Scores (21.93)	–	76.33	0.00	93.08
Base Rhyme Scores (2.48)	–	2.64	1.07	9.77

Table 4.9: Summary of participant characteristics, means, standard deviations, and observed score ranges.

- H1: DAT scores correlate with base rhyme scores
- H2: Experience level correlates with base rhyme scores
- H3: DAT scores correlate with tool-assisted rhyme scores

Relationship	Correlation Coefficient	P-value
DAT Score vs Base Rhyme Score	<b>0.41</b>	<b>0.023</b>
DAT Score vs Support Rhyme Score	<b>0.39</b>	<b>0.030</b>
DAT Score vs Co-creative Rhyme Score	<b>0.37</b>	<b>0.039</b>
DAT Score vs Autonomous Rhyme Score	<b>0.36</b>	<b>0.044</b>
Experience Level vs Base Rhyme Score	0.24	0.194

Table 4.10: **DAT and Rhyme Score Correlations.** Correlation between Divergent Association Task (DAT) scores and rhyme density measures, examining whether participants’ baseline divergent thinking relates to patterned poetic output in the structured writing tasks.

**Interpretation.** Moderate positive and statistically significant correlations were observed between DAT scores and rhyme density in all writing conditions. This supports **RQ2-1** by showing that divergent thinking ability predicts stylistic complexity across tasks, even without computational support. Experience level, by contrast, did not significantly correlate with baseline rhyme density, suggesting that creative potential, as captured by the DAT, was more influential than self-reported expertise at the outset.

## Regression Analysis

OLS regressions were used to identify predictors of rhyme density scores for each tool condition.

Dependent Variable	Significant Predictors	R <sup>2</sup>	Notes
Base Rhyme Score	DAT Score (p = 0.069, marginal)	0.17	Weak fit.
Support Rhyme Score	Experience Level (p = 0.002)	0.39	Strong predictor.
Co-creative Rhyme Score	Experience Level (p = 0.020)	0.30	Consistent with Support.
Autonomous Rhyme Score	Experience Level (p = 0.033)	0.29	DAT marginal (p = 0.058).

Table 4.11: OLS Regression Models Predicting Rhyme Density Across Conditions.

**Interpretation.** Experience level emerged as a predictor of rhyme performance in tool-assisted conditions, while DAT scores were only marginal predictors. This aligns with **RQ2-2** by indicating that experienced writers were able to leverage support and co-creative tools more effectively than novices. The finding suggests that while divergent fluency matters initially, expertise interacts with tool design to shape rhyme outcomes once computational assistance is introduced.

## ANOVA

One-way ANOVA tested differences in rhyme performance by experience level.

Dependent Variable	Factor	F-statistic (p-value)	Notes
Base Rhyme Score	Experience	F = 1.72 (p = 0.194)	Not significant
Support Rhyme Score	Experience	F = 6.34 (p = 0.005)	Significant effect
Co-creative Rhyme Score	Experience	F = 4.67 (p = 0.018)	Significant effect
Autonomous Rhyme Score	Experience	F = 3.99 (p = 0.030)	Significant effect

Table 4.12: One-way ANOVA: Experience Level Effects on Rhyme Scores.

**Interpretation.** ANOVA confirmed that experience significantly influenced rhyme outcomes when tools were introduced, but not in the baseline condition. This supports **RQ2-3**,

#### 4. Study II: Poetry and Writing Assistants

---

showing that support and co-creative tools amplify differences in expertise, whereas unaided writing performance was less stratified. In other words, tool use widened the gap between novice and experienced participants.

### Tool Usefulness and Subjective Evaluation

Subjective ratings of tool usefulness were analysed against participants' characteristics.

Relationship	Correlation Coefficient (r)	P-value
Experience vs Co-creative Usefulness	-0.378	0.038
Others	–	> 0.1

Table 4.13: Experience, DAT, and Rhyme Score Correlations with Usefulness Ratings.

Only one significant correlation emerged: experienced participants rated the co-creative tool as less useful.

Tool Type	Predictor	Coefficient ( $\beta$ )	p-value
Support	Base Rhyme Score	0.134	0.011*
Co-creative	Base Rhyme Score	0.113	0.043*
Autonomous	Base Rhyme Score	0.095	0.084 <sup>†</sup>

Table 4.14: **Predictors of Perceived Tool Usefulness.** Regression model showing which factors predict participant ratings of tool usefulness. The table situates perceived usefulness within broader relationships between fluency, accuracy, creativity, and participant characteristics.

**Interpretation.** Participants who performed well without tools were more likely to rate support and co-creative tools as useful, while DAT scores were not predictive. This contributes to **RQ2-3** by showing that perceptions of usefulness were grounded in baseline skill rather than divergent thinking capacity. Advanced writers sometimes viewed tools as redundant: *“I have a decent enough vocabulary...it didn't make much of a difference”* (S2-P29). Novices found scaffolding valuable: *“For a novice like me it was useful”* (S2-P10).

### Qualitative Analysis

Participant feedback revealed contrasting views across tools. To avoid overstating the approach, responses were grouped into broad descriptive categories rather than a formal thematic analysis:

- **Support Tools:** Highly valued for scaffolding, especially among novice users. Feedback cited ease in locating rhymes and stimulating ideas.
- **Co-creative Tools:** Mixed responses. Some found them creatively energising, while others found suggestions intrusive or stylistically mismatched.
- **Autonomous Tools:** Seen as powerful but “too automatic.” Some advanced users felt displaced by them, while others praised their fluency.

These findings align with the taxonomy in Table 4.2, where support tools provided the most direct scaffolding, while co-creative and autonomous systems produced more varied reactions.

**Interpretation.** Participant reflections offer nuance to **RQ3-4**. Support tools were often described as enabling and confidence-building: *“The rhyme dictionary gave me confidence to finish the piece”* (S2-P22). Co-creative tools polarised users, inspiring some—*“Verse by Verse expanded my horizons”* (S2-P12)—while disrupting others: *“I felt as though I was pushing against a lot of different styles”* (S2-P1). Autonomous tools drew the sharpest contrast. Novices praised their fluency—*“ChatGPT was massively helpful”* (S2-P16)—but advanced writers emphasised a loss of authorship: *“It didn’t help me write, rather it wrote for me”* (S2-P12). These responses reinforce the need to evaluate not only outputs but also participants’ sense of agency and ownership.

### **Research Questions Revisited**

The analyses presented in Study II can now be revisited in relation to the primary research questions introduced in Section 4.3. The discussion below demonstrates how statistical, computational, and qualitative findings collectively address each line of inquiry, without introducing new questions beyond those that structured the study design.

#### **RQ2-1: How do different tool classes influence fluency, accuracy, and perceived creativity?**

Rhyme density improved across all tool-assisted conditions compared to baseline, with the strongest effects observed for support and co-creative tools. Participants also reported higher perceived fluency and stylistic control with support tools, while co-creative systems produced more variance in subjective creativity ratings.

#### **RQ2-2: How do writing tools shape user engagement and focus during composition?**

Qualitative comments indicated that co-creative tools helped some participants enter a flow state by prompting new ideas, while others found the same systems disruptive. Autonomous tools generated the most polarised engagement patterns, with some users appreciating their speed and others reporting reduced agency.

#### **RQ2-3: How does prior writing experience affect interaction with different tool types?**

More experienced poets showed greater stylistic gains from support and co-creative tools, while novice writers tended to emphasise structural assistance over stylistic experimentation. Experience was a consistent predictor of rhyme performance in baseline and tool-assisted conditions.

#### **RQ3-1: In what ways can rhyme density capture meaningful variation in poetic form?**

Rhyme density correlated with human judgements of stylistic quality across conditions and showed interpretable variation between tool types. These patterns support its usefulness as a task-aware, interpretable proxy while highlighting its limitations in capturing semantic or thematic novelty.

**RQ3-2: Can cadence similarity act as a proxy for performative co-creativity?**

Cadence was evaluated only in focused supplementary analysis, but preliminary results showed that participants who maintained stronger rhythmic consistency with tool outputs produced higher judged fluency. These results motivate the systematic cadence analysis conducted in Study III.

**RQ3-3: How do subjective user ratings relate to computational measures?**

Participants who produced higher rhyme-density scores tended also to rate tools as more useful or creatively enabling. This alignment between computational and self-reported measures supports the thesis's argument for pairing interpretable metrics with human-centred evaluation.

#### 4. Study II: Poetry and Writing Assistants

---

The table below consolidates these findings by mapping each primary research question (RQ2 and RQ3) to the corresponding evidence presented in the quantitative and qualitative analyses. It provides a high-level summary that complements, rather than duplicates, the detailed RQ-by-RQ discussion in the preceding subsections.

<b>Research Question (RQ2 &amp; RQ3)</b>	<b>Summary of Findings</b>
<b>RQ2-1: Tool classes and stylistic output</b>	Rhyme density increased in all tool-assisted tasks; support and co-creative tools yielded the strongest stylistic gains.
<b>RQ2-2: Engagement and compositional focus</b>	Co-creative tools sometimes supported flow but could be intrusive; autonomous tools produced the most polarised engagement patterns.
<b>RQ2-3: Experience effects</b>	Experience significantly predicted rhyme outcomes in support and co-creative conditions; novices relied more on structural scaffolding.
<b>RQ3-1: Rhyme density as proxy</b>	Rhyme density correlated with human-perceived stylistic quality and varied meaningfully by tool type, supporting its interpretability.
<b>RQ3-2: Cadence similarity as proxy</b>	Preliminary cadence findings showed rhythmic alignment associated with perceived fluency, motivating full analysis in Study III.
<b>RQ3-3: Subjective vs computational measures</b>	Higher base rhyme performers tended to rate tools as more useful; alignment between subjective ratings and computational measures supports the layered evaluation approach.

Table 4.15: **Summary of Research Questions and Findings.** Overview of how Study II's research questions were addressed empirically, showing how different tool classes influenced poetic output and subjective experience.

### **Creativity Framework Summary**

DAT was reinforced throughout Study II as a targeted measure of divergent thinking, selected for its brevity and empirical validity. Combined with the structured poetic composition tasks, which serve as convergent challenges, the study offers a novel dual-method approach to evaluating creative fluency. Although the DAT and structured writing tasks have previously been used in isolation, their integration here constitutes a novel dual-method framework for evaluating creative fluency in co-writing contexts. The results demonstrate the value of using divergent and convergent creativity measures together to assess the impact of computational writing tools.

- Divergent Creativity – measured via DAT
- Convergent Creativity – expressed through structured poetic writing, which reuses DAT outputs

Together, these tools provided a means to predict participant performance and engagement across various tool types, and to surface differences in perception and behaviour that would be difficult to capture using traditional metrics alone.

## **Illustrative Participant Outputs: Comments and Poem Excerpts**

To contextualise the quantitative and qualitative results that follow, this subsection presents illustrative comments and short excerpts from participant poems. These examples give a sense of the range of creative responses observed across the writing tasks, complementing the formal analyses reported later in this chapter. Selected poem excerpts and rhyme density scores are provided in Appendix B.2. This links back to Section 2.6, where rhyme is established as a central stylistic device in both poetry and rap, grounding the analytic focus of this study.

### **Participant Reflections**

“Opened new possibilities that I could freely select elements from. . . reminded me the most of experiences I have had when doing collaborative writing with other people.”

“I think it made it a lot more difficult to write. I felt as though I was pushing against a lot of different styles of writing and it ended up disrupting the flow of the poem.”

“I am not really sure I can give myself much credit for the writing since the tool did almost all of the work. I suppose it depends how collaborative you attempt to make it with the AI. I would say it was definitely useful, though it also raises some other questions or concerns.”

These reflections illustrate contrasting experiences: some participants experienced AI tools as expansive and collaborative, while others felt tension between their preferred writing styles and the system-generated suggestions.

### **Poem Excerpts**

The following excerpts, selected for representational breadth, illustrate the stylistic diversity of participant outputs. Full versions appear in Appendix B.2.

*Excerpt: Higher rhyme usage*

The piano lies under the neck of the  
giraffe which leans to the left,  
in the warehouse on Overbury Road  
which leaves me occasionally bereft.

*Excerpt: Minimal rhyme usage*

We are all just darts in a board,  
wishing we had hit the bull's eye.

*Excerpt: Moderate rhyme usage*

The house that made us  
who we were destined to become  
was built from red bricks  
that glowed in the sun.

These examples situate the subsequent analyses by illustrating how participants varied in their use of rhyme, imagery, and narrative structure. They also support the study's focus on rhymed verse as a shared creative domain across poets and MCs, reflecting the broader theoretical grounding discussed in Section 2.6.

## 4.8 Discussion and Future Work

### Limitations

As noted in Section 4.4, several design choices introduced trade-offs between methodological rigour and logistical feasibility. These choices shaped the scope of Study II as an exploratory, remotely administered investigation, while also highlighting clear directions for refinement.

First, writing tools were presented in a fixed sequence for all participants. This sequencing standardised the participant experience and reduced procedural complexity, but may also have introduced ordering effects such as learning, fatigue, or strategic adaptation across tasks [264, 265]. Future studies could randomise or counterbalance tool sequences to isolate tool-specific effects more robustly [266, 267].

Second, apart from the DAT, writing tasks were not explicitly timed. Allowing participants to write at their own pace privileged ecological validity and comfort, but it also introduced uncontrolled variation in time-on-task, attentional effort, and cognitive load. Because the study was conducted entirely online and without process-tracing measures (e.g., keylogging or screen capture)<sup>3</sup>, variation in planning and reflection was estimated indirectly rather than directly observed [26, 157, 208]. Incorporating such measures in future work would provide richer insight into the temporal dynamics of creative writing.

Third, while the DAT and rhyme density provide reproducible proxies for divergent and convergent creativity, they necessarily capture only certain dimensions of poetic practice. Rhyme density was used as a proxy for stylistic complexity and compositional structure, but it does not account for multisyllabic rhyme, assonance, emotional resonance, or narrative coherence [127]. These dimensions remain difficult to quantify but central to human judgements of poetic quality. Although participants' qualitative reflections were collected via post-task questionnaires, these were not integrated into the main computational analysis. A richer synthesis of quantitative and qualitative perspectives would strengthen future evaluations.

Fourth, the modest sample size ( $N = 30$ ) was suitable for exploratory analysis but not powered for fine-grained subgroup comparisons (e.g., by experience level or DAT score bracket). While the sample included a diverse range of participants, including writers,

---

<sup>3</sup>For the DAT, participants had the option to send pictures of their scores. Although some did this, it was less useful in practice as those who submitted screenshots were less likely to provide accompanying text values, which would have necessitated additional manual reconciliation.

journalists, and creatives, writing expertise was self-reported and may not consistently reflect actual skill level across genres or modes of expression.

Finally, the study's online format offered accessibility and convenience, mirroring the real-world conditions under which many writing tools are used. At the same time, it introduced potential variability in participant environments, device configurations, and levels of distraction [268, 269, 270]. While laboratory settings would provide greater control, the ecological validity of this format was valuable for examining tool use in authentic conditions.

These considerations indicate that the findings of Study II should be read as exploratory and hypothesis-generating, with design choices made to balance feasibility, ecological validity, and participant engagement. Rather than undermining the contribution, these trade-offs clarify the scope of the study and point toward methodological refinements discussed further in Discussion and Future Work (4.8).

## Discussion

The findings of Study II underscore the importance of evaluating co-creative systems in terms of both user agency and creative augmentation. Yet, assessing co-creativity remains an open research challenge: there are no standardised metrics for measuring computational co-creativity [271, 272]. Calderwood et al. argue that writers using co-creative systems seek insight and inspiration, goals not captured by standard NLP benchmarks such as perplexity [273] or task performance [260]. The study aimed to computationally support users in creating more compelling poetry, offering what can be framed as *computational scaffolding* that augments human expressive capacity [255].

At the same time, the study highlights important methodological considerations. Rhyme density provided a reproducible proxy for stylistic complexity, enabling quantitative comparisons across participants and tool conditions. However, as discussed in Section 4.8, rhyme density captures only a subset of poetic practice: it neglects multisyllabic rhyme, assonance, semantic resonance, and emotional impact. Its utility lies less in offering a complete account of poetic creativity than in opening a replicable pathway for combining computational signals with human-centred evaluation. This methodological choice, balancing reproducibility with necessary abstraction, reflects the broader trade-offs of exploratory work in computational creativity. The study also raises conceptual issues around agency and authorship. Because rhyme is not a precise linguistic category but a

spectrum of phonemic resemblance (explored in greater detail in Chapter 5), the choice of  $\theta$  necessarily balanced between overfitting (capturing spurious sound similarities) and underfitting (missing subtle or culturally recognised rhymes). This trial-and-error process reflected the absence of a canonical standard for computational rhyme evaluation. Alternative approaches, such as supervised classifiers trained on labelled rhyme pairs, could have enabled more systematic optimisation, for instance through Receiver Operating Characteristic (ROC) curve analysis of precision–recall trade-offs, a standard method for evaluating classifier performance (see [274]). Although beyond the scope of the present study, such refinements point toward future methodological improvements. Given the exploratory nature of Study II, the analysis focused on rhyme density as a single interpretable measure, even though richer analyses of rhyme families or multisyllabic structures were possible. While effective as a quantitative proxy, this measure did not capture multisyllabic rhyme, assonance, or semantic resonance, which remain central to poetic practice (see Appendix B.1).

Supporting users without overwhelming their contributions is central to HCAI [253, 275, 276]. This concern is echoed in critiques that too much automation can reduce user agency and expression [159, 256]. Accordingly, systems and users should be evaluated both independently and interactively. Ideally, the system functions as a stimulus that prompts creative development beyond the tool itself. In Study II, experienced writers in particular expressed a lack of ownership when working with autonomous tools, reflecting ambivalence about whether the resulting text could be considered *their own*. Similar concerns are reported in prior work with professional writers, who often frame AI text generation in terms of authorship, ownership, and plagiarism [277]. These findings highlight the importance of evaluating not only textual quality but also participants' perceived relationship with AI. A brief post-study question also explored participants' beliefs about authorship and originality. Most respondents believed they could distinguish between human- and AI-generated writing, though this confidence contrasts with evidence suggesting that readers often struggle to do so reliably [278, 279]. These perceptions, while peripheral to the main analysis, help contextualise participants' responses and illustrate how cultural assumptions about authorship may shape engagement with co-creative systems. They also provide a valuable lens for interpreting the main findings, suggesting that prior beliefs about authorship and originality influence how participants evaluate AI-assisted writing. More broadly, such perceptions reflect ongoing public discourse on

the boundaries between human and machine creativity—boundaries that are increasingly blurred by generative systems [25, 280]. These questions are revisited in Chapter 5, which examines the evolving dynamics of human–machine co-creativity.

### **Future Work**

Participants in Study II used AI writing tools to create textual artefacts—verse as portable, symbolic, and expressive containers of place and identity. These AI-generated or co-created poems can be understood as inscriptions, akin to the interplay of agency and authorship seen in hip hop lyricism, where formal constraints (metre, rhyme) intersect with expression and cultural identity. The findings of Study II also extend the place-based inquiry of Study I. Where the first study examined how physical and cultural contexts shape creativity, Study II demonstrates how computational tools mediate the textual and stylistic dimensions of expression. In doing so, it provides a connective layer between situated practices of placemaking and the dynamic performance experiments in Study III.

Study III extends these findings by moving from textual artefacts to performance-based interaction. Drawing inspiration from GAN architectures [281], the study frames *writing* as a dialogue between human improvisation and a dynamic, responsive AI partner. Rather than evaluating output alone, this model shifts focus to performance, adaptability, and co-agency. Details of this interaction framework, and its application to creative writing, are presented in Chapter 5. In this sense, Study III also addresses the ownership concerns raised in Study II by shifting co-creation into embodied exchange, where improvisational AI enables participants to claim and inhabit creative outcomes as more *their own*.

## 4.9 Publication and Contribution Note

Elements of this chapter have been published and presented in the following conference workshops:

Olátúnjí, I. **Why try to build a co-creative poetry system that makes people feel they have ‘creative superpowers’?**. IUI Workshops, pp. 67–80, 2023. Available at: <https://ceur-ws.org/Vol-3359/paper8.pdf>

Olátúnjí, I. **Interactive writing systems and why small(er) could be more beautiful**. The Second Workshop on Intelligent and Interactive Writing Assistants, co-located with the ACM CHI Conference on Human Factors in Computing Systems (CHI 2023), Hamburg, 2023. Available at: [https://cdn.glitch.global/d058c114-3406-43be-8a3c-d3aff35eda2/paper8\\_2023.pdf](https://cdn.glitch.global/d058c114-3406-43be-8a3c-d3aff35eda2/paper8_2023.pdf)

Olátúnjí, I. **Humans feeling creative and the dangers of large language models**. GenAICHI 2023 Workshop at CHI (non-archival), Hamburg, 2023.

The author is the sole author of all listed papers and presentations. Their contribution encompassed the experimental design, the implementation of rhyme density as a computational creativity proxy, and the analysis of participant data. These outputs constitute an integral part of the thesis and directly informed its methodological and conceptual development.

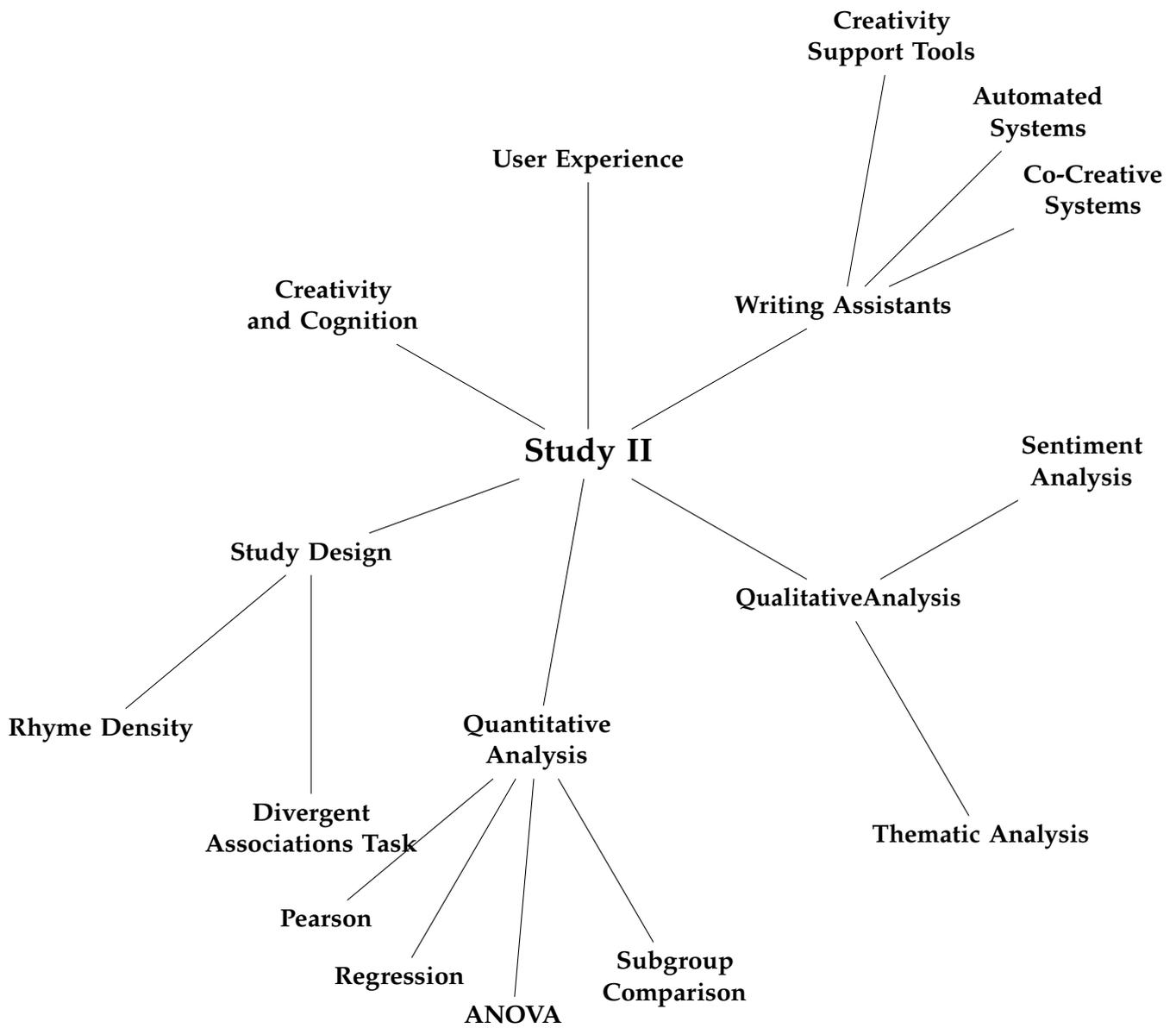


Figure 4.4: Thematic structure summarising the design, implementation, and analysis of assistive writing technologies in creative contexts.

### Study II — Findings Summary

**Hypothesis.** AI-assisted writing tools influence poetic composition not just through content suggestion but via structural scaffolding (e.g. rhyme, cadence), altering stylistic decision-making.

**Evidence.** Participant writing under support, co-creative, and autonomous conditions varied in rhyme density, keyword integration, and perceived fluency. Structured prompts seeded by local context (Swansea, London) influenced tone and form.

**Interpretation.** Generative tools act as semi-autonomous agents whose outputs shift the locus of creative control. Their role depends on both model architecture and interface constraints.

**Limitation.** Rhyme density is an extremely limited proxy for creative quality. For instance, rhyme is not a feature of most award-winning contemporary poetry. Rhyme density measures are not standardised (variance in scores depending on methods); lack of standardisation reduces generalisation across studies. Finally, *rhyme* itself is an imprecise concept.

**Future work.** Improvements in rhyme density standardisation (e.g., focusing on a genre where rhyme is considered highly important, such as rap). Further exploration and/or implementation of previous work (e.g., Ohriner). Study III evaluates real-time, voice-based co-performance to capture more embodied features of style.

## Chapter 5

# Study III: Generative AI and Co-Creativity



Figure 5.1: **Human–AI Battle Rap Illustration.** Conceptual depiction of a human MC and their Digital Creative Twin (DCT) in adversarial exchange, visualising the stimulus–response dynamics central to Study III’s evaluation design.

## 5.1 Introduction

Study II evaluated how different types of AI writing tools influenced fluency, creativity, and accuracy in structured poetic composition, revealing how system design and user experience shape co-creative outcomes. Study III builds on this foundation by extending the inquiry into performed expression, including spoken word poetry and freestyle rap. To frame this shift, the study (Figure 5.2) utilizes a continuum spanning written poetry, spoken word, and freestyle, each placing increasing demands on embodiment, improvisation, and real-time interaction. These performance-oriented practices introduce additional technical and experiential challenges for generative systems.

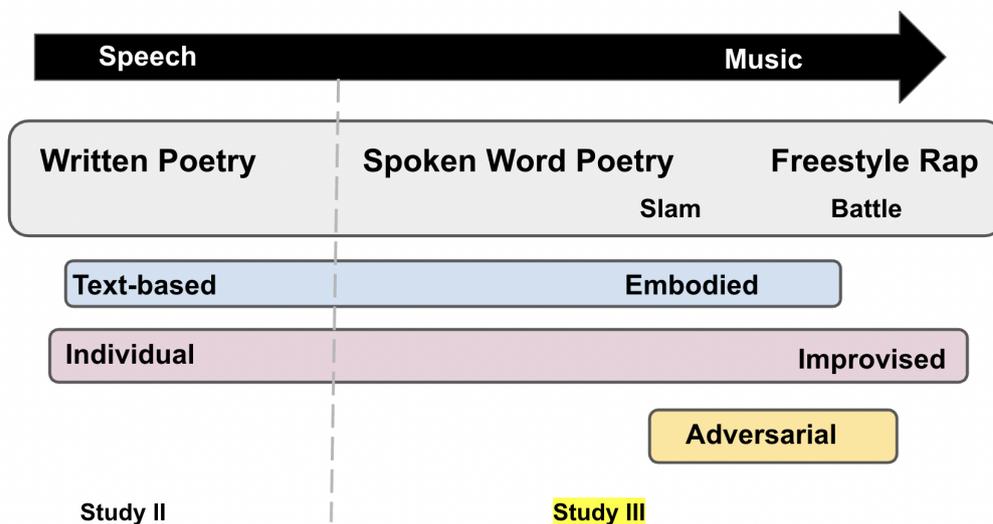


Figure 5.2: **Study III: Continuum of Expressive Forms.** A visual map positioning written poetry, spoken word, and freestyle rap along a continuum of increasing demands on timing, delivery, improvisation, and embodied presence.

The study comprises multiple experiments using voice-based AI systems. The systems approximate a Digital Creative Twin (DCT), a personalised language model that reflects a human user’s linguistic and stylistic features. The DCT is positioned as both collaborator and adversary, capable of challenging the human performer through improvisational interaction.

Figure 5.3 reprises the Venn diagram with *Generative AI* and *Performance* in the foreground. The diagram locates the DCT at the centre of *co-creation-in-the-loop*, signalling the shift from text-only composition to embodied, time-critical performance (both live and

## Hip Hop as Generative AI

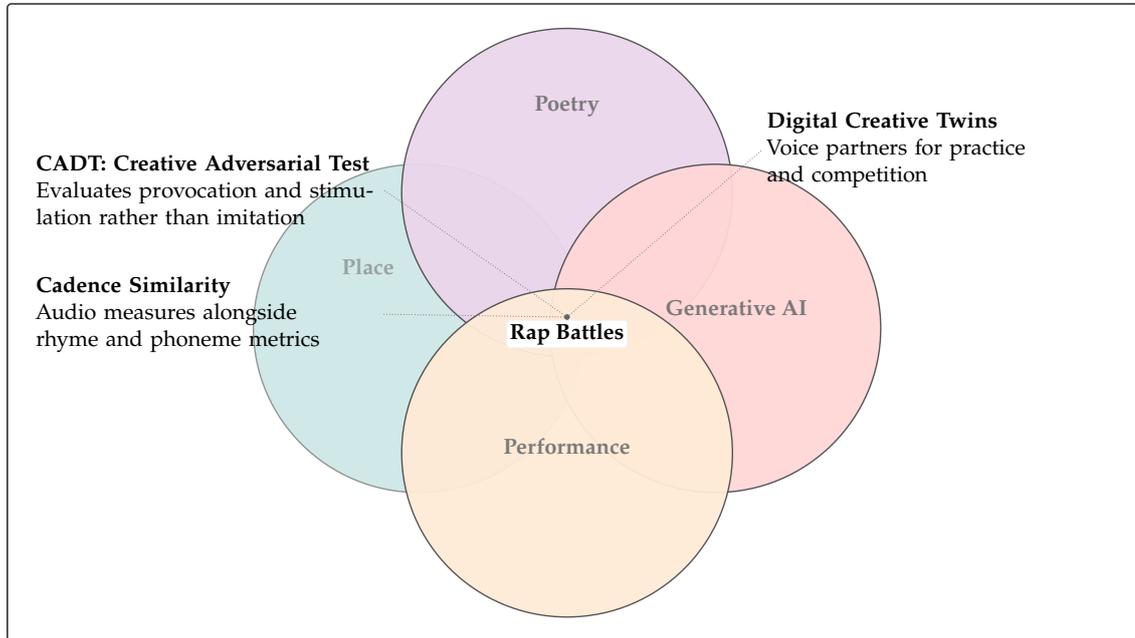
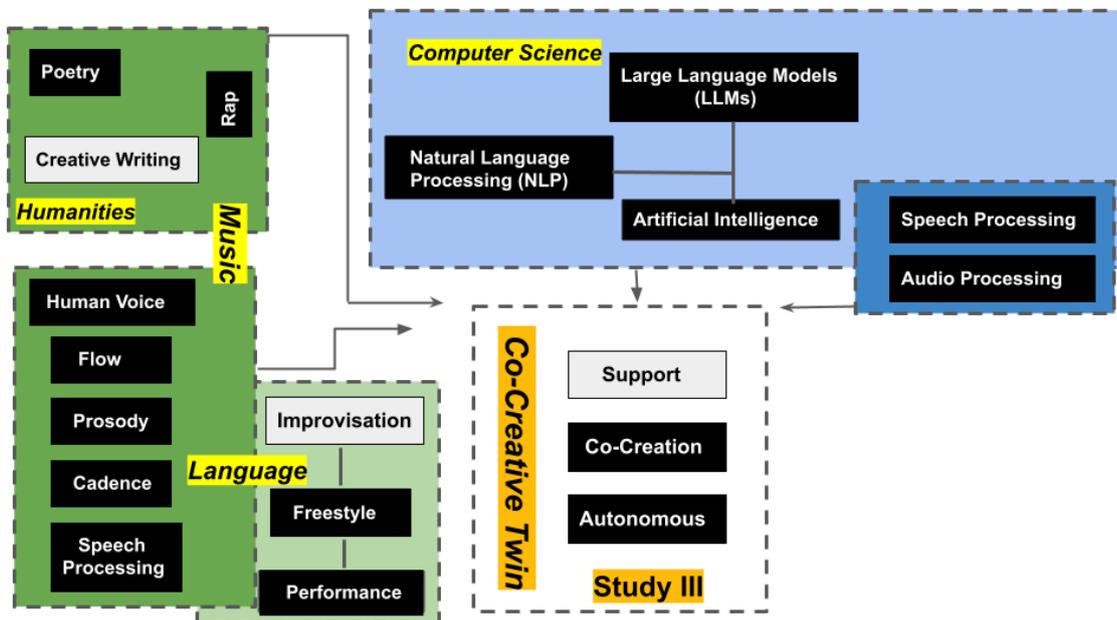


Figure 5.3: Study III framework: **Performance** and **Generative AI** as foreground via adversarial performance (rap battles). The bold *Performance* label highlights that participation is *embodied*; voice, breath, timing, and delivery. In this study embodiment is chiefly habitual (everyday vocal performance) rather than the heightened, site-specific embodiment of staged performance (e.g., Arts ARKADE). The figure also introduces the Creative Adversarial Test (CADT), cadence similarity, and Digital Creative Twins (DCTs) as scaffolds for evaluation and system design.

asynchronous). The study bridges back to Study II by carrying forward rhyme density while adding cadence-based measures. Finally, Study III proposes the *CADT* as evaluative frame. This figure thus orients the reader to the experiments that follow: style-aligned voice models sparring with human performers through improvised poetry and/or rap battles. Governance aspects of DCTs are discussed in Chapter 6.

**Note.** For clarity, *embodied* in Figure 5.3 denotes bodily participation in Study III, and is not a claim on the design discipline of *embodied practice*. For further detail on embodied practice as a design discipline see Barber [282] and Lindgaard et al. [283].



Study 3: Key Topics

Figure 5.4: **Thematic Map of Key Topics:** Overview of Study III, mapping the chapter’s structure from research aims to methodology, implementation, and findings, illustrating links between improvisational performance, battle rap structure, voice cloning, cadence analysis, and human–AI co-creativity.

Figure 5.4 presents the principal conceptual domains that structure the chapter and guide the design and interpretation of Study III. The figure combines interdisciplinary terrain within which this investigation is situated, including foundational work in computational creativity, real-time language modelling, and the analysis of prosody, cadence, and vocal delivery. It also encompasses theoretical frameworks concerned with improvisation, adversarial dynamics, and co-creative interaction between human and machine agents. The topic map serves multiple purposes. It provides a high-level overview of the chapter’s conceptual architecture and orients the reader to the varied disciplinary frameworks that inform the experimental and interpretive strategies used. At the same time, it foregrounds the methodological complexity of studying embodied co-creativity, where textual, acoustic, and cognitive features interact dynamically. Study III positions spoken word poetry and freestyle rap as a test bed for multi-modal improvisation; the map reflects the various analytical lenses required to interrogate this phenomenon. This work highlights that the capacity for real-time, improvisational performance is a critical test of whether AI can meaningfully participate in human creativity. The study addresses an important research

gap in computational creativity. Prior studies (for poetry) have largely focused on offline or text-only evaluation. Study III explores whether a personalised AI model can credibly participate in the embodied and rhythmic performance of freestyle verse.

### **Relation to Primary Research Questions**

Study III extends the thesis's primary research questions by examining co-creativity in a real-time performance domain, exploring how Digital Creative Twins respond to and interact with human artists in adversarial and collaborative settings. It directly operationalises RQ2 (tools and process) through the design of an asynchronous rap-battle system, and contributes to RQ3 (proxies and evaluation) by testing whether textual and acoustic measures can provide interpretable indicators of co-creative performance. Findings from Study II informed this design: rhyme density, stylistic alignment, and cadence analysis were carried forward and adapted from written poetry to freestyle rap, enabling a comparative evaluation across modalities.

### **Practitioner–Researcher Framing**

Study III engaged directly with improvisational performance, cadence modelling, and adversarial lyrical exchange. In this context, practitioner knowledge was essential for interpreting subtleties of flow, rhythmic timing, responsiveness, and the aesthetic conventions of battle rap performance. Within the study, expertise in improvisational lyricism and rhythmic delivery informed the development of the computational evaluation design, and supported the interpretation of cadence similarity metrics, call-and-response structures, and co-creative dynamics between human MCs and DCTs. This framing aligns with established accounts of practice-led research, which argue that practitioner–researchers can access dimensions of creative process not readily available through theoretical or computational analysis alone [16].

## 5.2 Related Work

### Text Generation

Beam-style techniques are an important part of modern text generation pipelines [284]. They enable the exploration of multiple high-probability continuations, in parallel, while supporting enforcement of global constraints (e.g., such as rhyme structure, syllable count, or keyword inclusion). Standard Beam Search retains the top- $k$  most probable partial sequences at each decoding step. Recent work has introduced Creative Beam Search (CBS) as a novel generate-and-test sampling approach aimed at simulating key phases of the creative process in text generation [285]. Grounded in Amabile’s model of creativity [286], which includes task presentation, preparation, response generation, and response validation. CBS operationalises this framework through a two-stage process. First, the system uses Diverse Beam Search (DBS) [287] to produce a varied set of candidate outputs, simulating the generative phase of creativity. These outputs are then subject to a model-led evaluation phase, in which an LLM assesses and ranks the candidates, emulating domain-relevant validation [288]. Compared to standard sampling methods, CBS has been found to produce more diverse and user-preferred results, suggesting its relevance for tasks requiring stylistic variability and internal self-assessment, both of which are central to AI co-performance in spoken word and freestyle rap contexts.

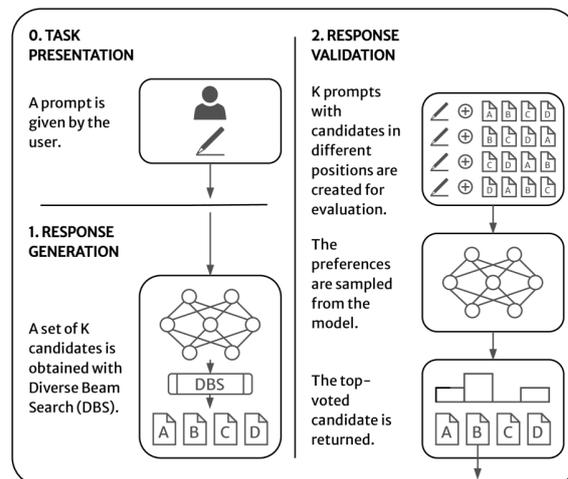


Figure 5.5: **Creative Beam Search Process.** Diagram illustrating the Creative Beam Search (CBS) method used for generating DCT responses. The figure shows how semantic coherence, provocation, and rhythmic constraints are balanced during improvisational verse generation.

This approach is useful in Study III where responses must (i) fit within a fixed time window and (ii) include mandatory keywords. By maintaining a narrow beam (e.g.,  $k = 4-8$ ), the model can produce multiple fluent, diverse text candidates.

### **Creative Beam Search and LLM-as-a-Judge Approach**

Motivated by Fransechelli & Musolesi’s Creative Beam Search (CBS), Study III adopted a generate-and-evaluate strategy for text production. Candidate verses (*beams*) were generated and rated according to stylistic features such as rhyme density, internal rhyme variety, and rhythmic novelty [285]. While CBS itself is not implemented directly, it served as design inspiration for structuring diversity and evaluation. The system also used multiple LLMs guided by Sawicki et al.’s poetry evaluation framework [289], which builds on the Consensual Assessment Technique (CAT) [32]. Appendix D.10 provides the full prompts and sample outputs.

### **Speech Recognition and Modelling**

**Speech Recognition.** Automated Speech Recognition (ASR) systems have made significant advancements in recent years, achieving Word Error Rates (WER) as low as 5%, approaching human transcription accuracy [290, 291, 292, 293]. Despite this progress, ASR systems still exhibit higher error rates when processing accents, dialects, and non-native English speakers [292, 294]. While these models have shown promise in Automated Speech Recognition (ASR), their application to rap transcription remains an open challenge due to the complexities of flow, slang, and phonetic variation. Developing a battle rap system capable of generating real-time lyrical responses to a human MC requires highly accurate, low-latency ASR tailored to these linguistic and rhythmic nuances [18]. These limitations are particularly relevant to the proposed system, as rap lyrics frequently incorporate non-standard English, slang, phonetic variation, and complex rhythmic structures, which pose additional challenges for ASR [295, 296]. Current ASR systems also perform significantly less well on speech that approaches musicality, including sung or rhythmically stylised delivery. In these situations WERs are within the 10% - 30% range [297, 298]. Standard ASR models are not specifically designed for rap transcription, presenting a gap in the research. Although WER is substantially higher for rap than for ordinary speech, this does not materially affect the present study. The principal analytical measures, including rhyme density and cadence similarity, rely on phonetic similarity rather than exact lexical

accuracy. They remain robust therefore provided the ASR output preserves the essential phoneme patterns of the performance.

**Speech Modelling.** Spoken word poetry and rap is inherently multi-modal, requiring computational models to handle both textual elements (*meaning*) and phonetic features (*sound*). As Murs explains, “...not only do you have to make everything rhyme, but you have to add rhythm to it. Poetry doesn’t have to rhyme, it just has to sound beautiful. But in rap, it has to sound beautiful, it has to be on time, and it has to rhyme ” [119]. Generating semantically rich rap lyrics remains difficult, though *DeepBeat* and *DeepRapper* have explored data-driven solutions [205, 299]. *DeepBeat* optimizes rhyme density and coherence, leveraging a neural network to predict next-line candidates. However, it lacks rhythm modelling, and adversarial response capabilities, which are key for rap battles. *DeepRapper* uses a Transformer to model rhyme patterns and rhythmic structures, improving fluency over *DeepBeat* [299]. However, it remains a text-only system, making it unsuitable for rap battle co-creation with human MCs. Meaning and delivery are often regarded as equally important for human MCs, though a case can be made that delivery plays a more crucial role. As Havoc of Mobb Deep remarks, “Without the right flow, subject matter probably won’t even matter. It’s all about styles . . . the way you’re getting your subject across. If people can’t feel how you’re saying it, it doesn’t even matter what you’re saying” [119, 300]. This distinction is particularly relevant for computational systems. AI models may find delivery and flow more achievable, as meaning requires deeper contextual reasoning and cultural awareness. Prioritising delivery in computational models enables more realistic rap co-creation, even if full semantic understanding remains unachievable at present. The section *Discussion and Future Work* (5.8) expand upon this.

## Design Pitfalls

Recent advances in human–AI co-creativity have brought renewed attention to the design pitfalls that can undermine effective collaboration between people and generative AI. Buschek et al. identify nine potential pitfalls when designing human-AI co-creative systems [256]. These range from limitations in the AI or interface (e.g. hidden constraints or poor expressive control) to issues of over reliance and ethical concerns [25, 159]. In co-creative domains, a well-designed system must avoid these traps to truly benefit the user’s creative intent (reference) These considerations are critical in real-time creative

performance settings like freestyle rap battles, where any friction could derail the creative process in front of an audience. Digital Creative Twins (DCTs), AI agents modelled on a creator's style, introduce these pitfalls into the high-stakes arena of spoken word and freestyle rap. In a rap battle, performers improvise rhymes and rebuttals under extreme time pressure and stylistic constraints. If an MC is supported or opposed by a DCT, the usual co-creative design flaws can be exacerbated. For instance, an AI with invisible creative boundaries (pitfall 1) might subtly confine the lyric ideas to its training data comfort zone, hindering the human artist's spontaneity and exploration. Likewise, a DCT with a lack of expressive interaction options (pitfall 2) could bottleneck the performer's creativity, if the interface to cue or shape the AI's rap is too restrictive or slow, the human artist cannot fully direct the collaboration (reference). In a live battle there is no time to wrestle with complex controls or guess what the AI is able to do. Another concern is a false sense of proficiency (pitfall 3), where the AI's impressive-sounding verse might contain factual errors, off-theme lines, or weak insults that the human doesn't catch in time. Trusting such output blindly could embarrass the performer mid-battle, especially if an opponent exploits those mistakes. Moreover, the adversarial nature of rap battles magnifies pitfalls related to AI over-involvement and downstream effects. An AI rap twin must be carefully tuned not to step on the human's toes creatively – for example, flooding the user with too many suggestions or overly detailed options (“agony of choice,” pitfall 5) would paralyse more than empower in a fast-paced freestyle. Similarly, poorly timed interruptions (“time waster,” pitfall 6) from the AI could break the performer's flow or focus. Beyond usability, rap battles also sharpen ethical and cultural pitfalls. AI-generated lyrics might carry undesirable biases or inappropriate content learned from training data (pitfall 7), surfacing at the worst moment and offending the crowd. Questions of creative ownership and authenticity (pitfall 8) are equally pertinent: if a DCT is contributing a significant portion of the artistry, who truly *won* the battle? The blurring of authorship can be problematic in a genre where personal skill and originality are paramount. Even privacy and legal concerns (pitfall 9) could arise if, say, the AI inadvertently reproduces someone else's lyrics or private info from its data. The combination of real-time pressure, stylistic rigour, and adversarial stakes means that any design shortcoming is immediately apparent. Successfully deploying a DCT in this domain demands meticulous attention to these pitfalls, turning them from potential failure points into design opportunities for supporting improvisational human creativity.

5. Study III: Generative AI and Co-Creativity

Name	Rap Context	Problem Description	How it might have happened	How it might be addressed
<b>Limited AI</b>				
Invisible flow boundaries	Freestyle emulation fails to match off-beat or non-metric styles	AI imposes rigid flow, harming expressive timing	Model overfitted to standard 4/4 timing	Train on diverse rhythmic flows; user-tuning of flow
Voice style bottleneck	Twin voice lacks nuance in intonation, breath, or regional variation	Limited training data restricts prosodic expressiveness	Dataset lacks diverse speech acts (e.g. whisper, shout, chant)	Training set with expressive performance types; user-guided modulation
<b>Too Much AI</b>				
Overtake in co-performance	AI dominates or derails live co-rap session	AI generates dense verses that ignore human pacing or theme	System lacks real-time user response integration	Build interaction pacing layer; use adaptive prompts from live context
Suggestion fatigue	Too many AI punchline/s-scheme options	Users overwhelmed by volume or repetitiveness	No filtering for novelty, tone, or relevance	Add adjustable suggestion filters and style/energy settings
<b>Beyond Use</b>				
Attribution ambiguity	AI adopts artist's voice and content, raising authorship questions	Voice cloning + lyrical generation make authorship unclear	Output blends user data with training data without clear markers	Attribution watermarking; allow toggling between authorial modes
Cultural misalignment	AI misrepresents genre-specific norms or oversteps ethical boundaries	Rap contexts not fully encoded in training data or fine-tuning	Model fails to reflect positionality, tone, or respect norms (e.g. slang, community references)	Involve domain experts; build real-time cultural calibration layers

Table 5.1: Specific co-creative challenges for the Digital Creative Twin, aligned with categories from Buschek et al. *Nine Potential Pitfalls when Designing Human-AI Co-Creative Systems* [256].

## 5.3 Research Aims

### Guiding Research Questions

The study is supported by the questions mapped to four thematic areas: co-creative dynamics, stylistic evaluation, system responsiveness, and human perception. The questions are grounded in the interdisciplinary literature discussed in Chapter 2, including frameworks from computational creativity, improvisation studies, human–AI interaction, and multimodal performance. In keeping with the thesis-level research questions on Proxies and Evaluation (RQ3) and Adversarial Co-Creation (RQ4), the study introduces several subsidiary questions. These support and operationalise the broader research questions by breaking them into discrete components that structure the study’s evaluative focus. Table 5.2 outlines each question along with its evaluation focus, and links to relevant sections of the literature review.

Framing Questions	Focus Areas	Lit. Review Ref.
RQ3-1–RQ3-4: In what ways can textual and performance measures (e.g. rhyme, cadence, flow) serve as interpretable proxies for co-creativity?	Stylistic alignment, rhythm, performance metrics	§2.7, §2.5
RQ4-1–RQ4-4: How can adversarial human–AI interaction (e.g. rap battles with a Digital Creative Twin) be used to evaluate and stimulate creativity?	Interaction design, responsiveness, co-authorship, improvisation	§2.2, §2.8
How do participants assess the fluency and authenticity of digital collaborators?	Voice, timing, system credibility	§2.1, §2.8
What role does freestyle performance play in evaluating human–AI co-creativity?	Improvisation, embodiment, adversarial framing	§2.2, §2.8

Table 5.2: High-level research questions guiding Study III, contextualised by themes from the Literature Review. Detailed evaluation criteria are presented later in the chapter.

The findings of this study inform the broader thesis objective of developing accessible, personalised writing tools capable of supporting creative expression across diverse modalities. Study III introduces a novel experimental paradigm that evaluates personalised language models in an embodied, performance task. It proposes adapted metrics for fluency, creativity, and accuracy that explicitly account for voice-based delivery and

audience reception. In doing so, the study contributes an exploratory framework for assessing AI co-creativity in multi-modal contexts and provides empirical insights into the challenges and possibilities of real-time, personalised language model performance.

### **Study Aims**

The primary aim of Study III is to evaluate the feasibility and effectiveness of personalised language models in a multimodal, real-time creative setting. Specifically, the study seeks to:

1. Examine whether participants perceive their DCT as a credible and effective co-performer in a spoken word poetry/battle rap test.
2. Adapt and validate evaluation measures for fluency, creativity, and accuracy in a voice+text setting.
3. Explore whether personalised language models can maintain stylistic alignment and performance fluency under real-time, improvisational conditions.

RQ	Research Question	Related Hypotheses
RQ3-1	How does rhyme density vary between human-only and human–AI adversarial conditions?	H2 (creativity), H3 (authenticity), H5 (multimodal effect)
RQ3-2	Can cadence similarity act as a proxy for performance co-creativity?	H1 (fluency), H2 (creativity)
RQ3-3	How do subjective ratings relate to computational measures (rhyme density, cadence)?	H1 (fluency), H2 (creativity), H3 (accuracy)
RQ3-4	What minimal stylistic markers distinguish individual style from genre templates?	H3 (authenticity)
RQ4-1	How does adversarial engagement with a style-aligned DCT affect improvisational fluency?	H1 (fluency), H2 (creativity)
RQ4-2	Do DCT adversaries stimulate increased stylistic variation (CADT)?	H2 (creativity), H5 (modality effect)
RQ4-3	How does remix vs TTS synthesis affect perceived authenticity and alignment?	H4 (voice modality), H6 (synthesis effect)
RQ4-4	How do participant traits (experience, DAT) shape evaluation of adversarial co-creation?	H4 (experience effect)

Table 5.3: Primary and subsidiary research questions for Study III, with corresponding hypotheses.

Study III expands upon previous work by incorporating multimodal analysis across voice synthesis, cadence modelling, and poetic/rap text generation. It is organised into six interrelated experiments spanning voice-only, text-only, and combined voice plus text creative outputs. Study III retains several methodological elements from Study II to enable meaningful comparison across studies. Participants complete a pre-test *DAT* to assess their baseline verbal creativity [34]. The primary creative task mirrors Study II’s structure in that participants engage in a co-creative writing session with their DCT model, followed by post-task evaluations. However, the creative task differs significantly: rather than composing written poetry, participants engage in a *freestyle rap battle*, exchanging verses with their DCT.

This shift introduces key methodological changes. First, the data collected include text and audio of both participant and DAT performances. Second, the interactive nature of the task imposes time constraints on the AI system's response latency and turn-taking behaviour. Third, evaluation measures are adapted to account for the performance and multi-modal aspects of the creative output.

## Hypotheses

Based on the theoretical framework and findings from Study II, as well as the methodological adaptations introduced in this study, the following hypotheses are proposed:

1. **H1 (Fluency):** Participants will rate the DCT performance as fluent in the freestyle rap battle task, with fluency scores positively associated with objective cadence analysis measures (e.g., syllable timing, beat alignment).
2. **H2 (Creativity):** Participants will perceive the DCT's output as creative, with creativity ratings positively correlated with quantitative measures of rhyme density and flow scores.
3. **H3 (Accuracy and Authenticity):** Participants will report that the DCTs verses align stylistically with their own, and external audience raters will show low accuracy in distinguishing between human- and AI-generated verses, indicating a high degree of perceived authenticity.
4. **H4 (Effect of Participant Experience):** Participant experience level (e.g., prior creative writing or rap experience) and pre-test DAT scores will be positively associated with higher evaluations of the DCT's fluency, creativity, and accuracy.
5. **H5 (Modality Effect):** Compared to the results of Study II (text-only modality), the multi-modal (voice + text) interaction in Study III will yield lower perceived accuracy scores, reflecting the additional complexity and performance demands introduced by the real-time, embodied setting.
6. **H6 (Voice Modality Effect):** Participants will rate AI outputs using remix-based synthesis (Test A) as more authentic and style-aligned than those generated via fully synthetic TTS (Test B).

These hypotheses will guide the quantitative and qualitative analyses of participant and audience evaluations, contributing to a deeper understanding of the feasibility and limitations of personalised language models in multi-modal, real-time creative contexts.

**Sub-Study Mapping:**

- H1–H3 are addressed primarily in Study III-A (human–AI interaction).
- H3 is extended in Study III-B through model-level rhyme and cadence comparisons.
- H4 and H5 are tested across III-A and III-C by comparing outcomes with Study II and new human participants.

## 5.4 Method

Study III examined human–AI co-creativity across multiple performance settings, ranging from written verse to live and recorded freestyle rap. The chapter therefore adopts a multi-experiment structure, with each experiment designed to address different aspects of the guiding research questions introduced earlier. Across all experiments, the core methodological goal was to evaluate how Digital Creative Twins (DCTs) perform as expressive, stylistically adaptive co-creators in adversarial and collaborative contexts.

At a high level, the study combined three complementary forms of data:

- **Text-based stylistic metrics** such as rhyme density and phoneme–token ratios, extending the computational measures validated in Study II.
- **Acoustic and prosodic features** (e.g., cadence alignment, phoneme diversity, timing patterns) extracted from both human and AI performances to assess rhythmic and vocal authenticity.
- **Subjective evaluations** capturing performer and listener perceptions of fluency, creativity, authenticity, and stylistic alignment.

These measures allowed for an integrated evaluation of poetic, linguistic, and performance-based co-creativity across different modalities (text, audio, and mixed-mode interaction).

### Ethical Considerations

The study was conducted in accordance with institutional ethical guidelines. All participants provided informed consent prior to participation and were informed of their right to withdraw at any time. Audio recordings were anonymised and stored securely. The external audience evaluation used anonymised recordings to protect participant privacy. Participants were not financially compensated.

### Rhyme Density Measures

In this study, Rhyme Density (RD) was computed in two forms: raw counts and normalised values. As in Study II, RD functions as a computational proxy for lyrical creativity, here complemented by the cadence-based measures introduced in Study III.

- *Raw RD* represents the total number of rhymed phonemes detected in a verse, without adjusting for length. It reflects absolute rhyme use but scales with verse length, so longer outputs (e.g., poets vs. MCs) tend to yield higher values.
- *Normalised RD* divides the raw count by the total number of tokens, producing a proportional measure of rhymed phonemes per token. This allows direct comparison between outputs of different lengths and follows the method in Study II, although Study III employs a refined detector that also recognises multisyllabic and internal rhyme patterns.

Normalised RD values were used for statistical comparison, while raw RD values are reported in summary tables to preserve detail about absolute rhyme use. A full description of the extended rhyme detector, together with a worked example illustrating multisyllabic and internal matches, is provided in Appendix C.1.

### Evaluation Approach

Study III employed a multi-method evaluation strategy combining participant judgements, external listener assessments, and computational metrics. Evaluation occurred at three levels.

**(1) Participant self-evaluation.** After each interaction round, participants rated the DCT's responses across fluency, creativity, accuracy, and usefulness using the same questionnaire framework employed in Study II. These ratings assessed perceived co-creative quality from the performer's perspective. Participants could also provide open-text reflections; these were reviewed descriptively to identify how participants characterised flow, authorship, and stylistic alignment, without treating the comments as a formal thematic analysis.

**(2) External audience evaluation.** A separate group of MCs completed a listening-only evaluation in which anonymised audio pairs were rated using the same questionnaire items. This provided an independent assessment of stylistic quality and flow, allowing comparison across TTS and Remix conditions and triangulating participant self-reports.

**(3) Researcher-led computational analysis.** Textual outputs were analysed for rhyme density, phoneme–token structure, and lexical variation, extending methods from Study II. Audio outputs were evaluated using cadence-based measures (MFCCs, DTW alignment, timing patterns) to quantify rhythmic similarity and delivery style. These metrics provided

## 5. *Study III: Generative AI and Co-Creativity*

---

interpretable proxies for co-creative performance and allowed systematic comparison across conditions.

These three layers of evaluation enabled the study to assess co-creativity from performer perceptions, independent expert listeners, and computational proxies, supporting a multimodal understanding of how DCTs in adversarial and collaborative performance settings.

## 5.5 Technology Stack

Study III implemented a Digital Creative Twin (DCT) simulation designed to support asynchronous, voice-based human–AI interaction. Participants engaged with the system through WhatsApp<sup>1</sup> while system processing was conducted within a controlled Wizard-of-Oz environment built around Jupyter notebooks and modular Python scripts. The design allowed the study to approximate a production workflow while maximising control and interpretability. Study III were conducted between 8 June and 3 August 2025, and the LLMs, ASR, and TTS models used in the study correspond to the publicly available versions accessible during this period. To support reproducibility, the technological workflow can be understood across three stages:

- **Pre-production:** creation of voice clones, rhythm-model testing, prompt scaffolding, and preparation of rhyme/cadence analysis tools.
- **Production:** transcription of participant audio (Whisper), extraction of cues and stylistic markers, generation of candidate verses (LLMs), selection via an LLM-as-Judge ensemble, and synthesis of responses using TTS or Remix pipelines.
- **Post-production:** acoustic and linguistic analysis, including rhyme density, phoneme-level cadence comparison, MFCC/DTW alignment, and scoring.

This workflow situates the DCT as an evaluable creative agent operating across text, audio, and mixed-modality interaction. It also clarifies the separation between user-facing interaction (via WhatsApp) and back-end orchestration (Wizard-of-Oz), which is essential for replicating Study III under similar research constraints.

A broader taxonomy of speech interaction architectures, including uni-modal, multi-modal, synchronous, and real-time speech-to-speech systems, informed the design choices for Study III. The implementation adopted a uni-modal, Type 1 configuration using discrete asynchronous components for STT (Whisper) and TTS (e.g., ElevenLabs). This modular setup enabled the optimal control over cadence, phrasing, and voice synthesis, which was essential for evaluating co-creative performance in a battle-rap context. A summarised version of the taxonomy that guided these design decisions is provided in Appendix C.2.

---

<sup>1</sup>Participant sends a 30 second freestyle voice-note via WhatsApp and the DCT generates a response and returns a cloned-voice audio file back via WhatsApp. This illustrates the complete input→processing→output loop as experienced by participants, without requiring any additional interface beyond standard voice-note messaging.

### System Overview

Figure 5.6 provides a high-level overview of the asynchronous response pipeline used in Study III. The system begins with human audio input, which is transcribed and analysed for key phrases and adversarial cues. These inform a structured prompt aligned with the participant’s style and persona traits. An LLM generates a candidate response, which is then synthesised using a voice-clone model to produce a performance-style audio output.

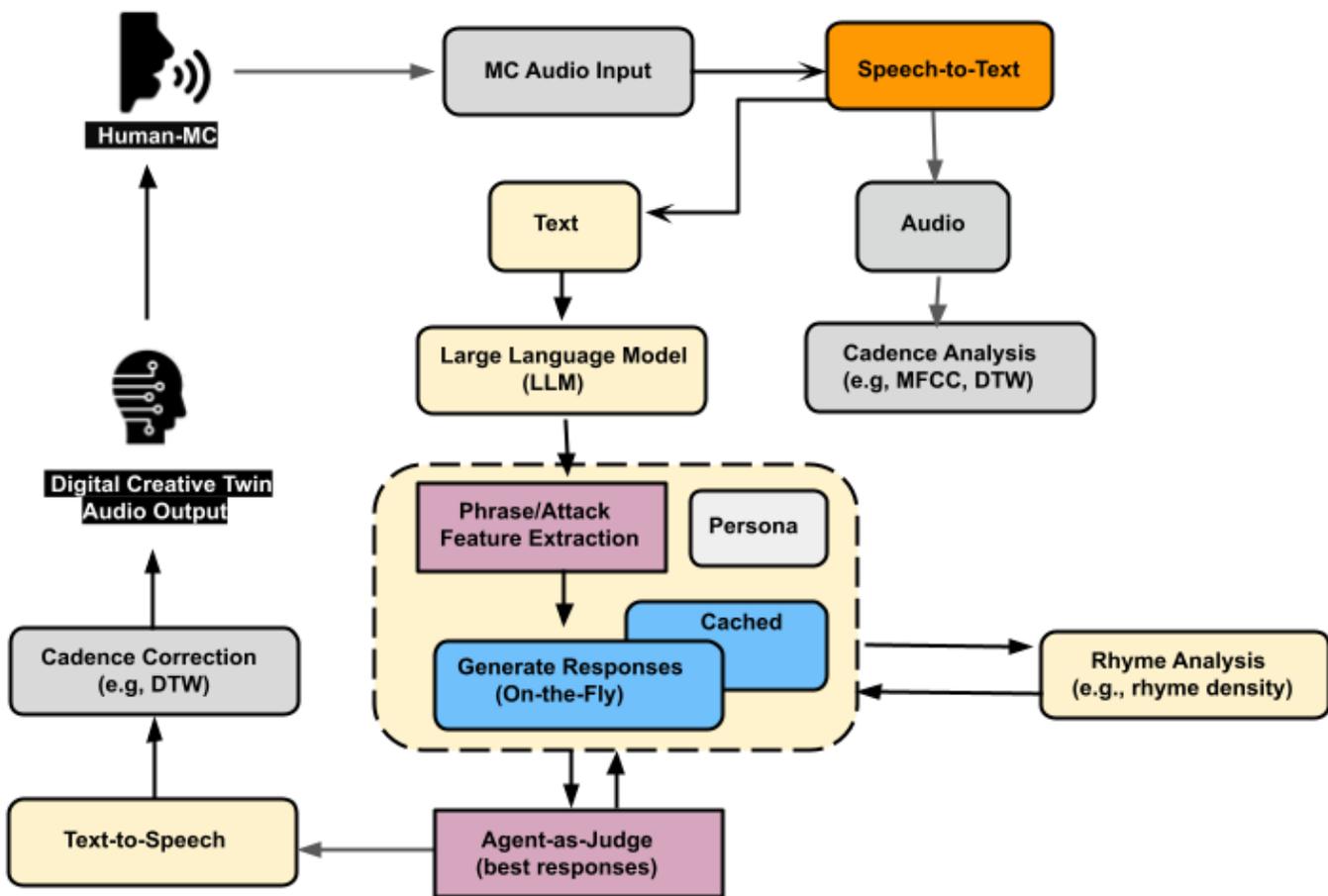


Figure 5.6: **DCT Response Pipeline.** High-level overview of the DCT’s asynchronous response process, from input transcription through cadence analysis to verse generation. This figure contextualises the computational components evaluated in Study III.

## Digital Creative Twin Response Pipeline

Step	Action	Agent
1	MC delivers 30-second freestyle verse	Human MC
2	Verse transcribed and analysed for keywords	System
3	Prompt assembled with theme and attack tokens	System
4	LLM generates counter-verse	DCT
5	Output refined for rhyme and cadence	System (LLM-as-Judge)
6	Synthesised using voice clone	DCT
7	Evaluated for cadence and similarity	System
8	Both verses scored and reflected upon	MC + System

Table 5.4: **Rap Battle Interaction Structure.** Structure of the human–AI rap battle used in Study III, including turn-taking, timing constraints, and stimulus–response dynamics. This table contextualises the evaluation of co-creativity under adversarial conditions.

### System Setup

The system was implemented using a Wizard-of-Oz approach, approximating a production-ready platform through modular integration of Jupyter notebooks, LLMs, and off-the-shelf speech tools. Audio analysis modules, developed in collaboration with Mark Sheppard and based on his prior work [301]-were implemented using Librosa [302]. This configuration supported controlled experimentation and inspection of intermediate outputs (e.g., rhyme density and cadence metrics) while avoiding premature commitment to a fixed interface. WhatsApp was chosen as the delivery platform for its ubiquity, low onboarding overhead, and support for both text and voice interactions within an encrypted environment. Overall, this setup provided an appropriate technical environment, balancing experimental control with the possibility of later real-time deployment.

### Text Generation and Response Assembly

Transcribed text was processed to detect rhyme patterns, extract key lexical cues, and identify adversarial phrases. These features were incorporated into a structured prompt specifying rhyme usage, keyword inclusion, desired length, and stylistic alignment. Response generation used high-performance LLMs available during the study period, including `gpt-4o`, `gemini-2.5-flash-preview`, and `claude-opus-4`. Sampling was inspired by Creative Beam Search, with constraints tuned to favour rhythmic consistency and poetic cohesion.

In parallel, a lyric remix module accepted clean text ( $N$ ) and generated a remixed variant ( $N+1$ ) that preserved rhythm, rhyme scheme, and vocabulary while enabling creative variation. This was useful in cases where cadence information was available from audio but transcription fidelity was low.

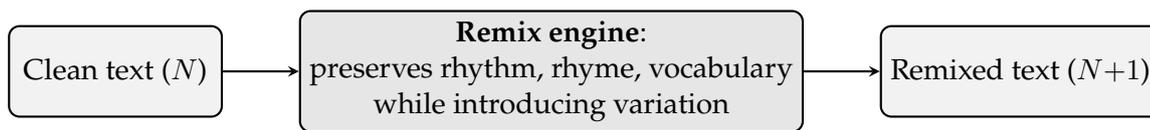


Figure 5.7: **Text-to-Text Remix Pipeline.** Illustration of the remix-based method used to generate stylised DCT responses by transforming existing lyrical material into battle-ready output.

### Human–Machine Creative Roles

Table 5.5 contrasts the human MC and the DCT across the creative pipeline, highlighting both technical symmetries and key asymmetries. Human participants contributed embodied rhythm, intentionality, and adaptive delivery, whereas the DCT provided consistency, precision, and scalable stylistic modulation.

Stage / Feature	Human MC	Digital Creative Twin (DCT)
<b>Creativity</b>	DAT (0–110)	LLM generative capacity
<b>Voice</b>	Natural recorded delivery	Cloned using ElevenLabs
<b>Cadence</b>	Measured via software	Measured and optionally cadence-aligned
<b>Lyric Generation</b>	Manual composition	LLM-generated and LLM-as-Judge refined
<b>Rhyme Evaluation</b>	Software-based	Threshold-checked and validated
<b>Reflection</b>	Self-critique	Iterative adaptation from prior outputs

Table 5.5: **Human MC vs. DCT Across Battle Stages.** Comparison of human and Digital Creative Twin (DCT) performance across the stages of the battle protocol, highlighting differences in responsiveness, flow, and stylistic behaviour.

### Voice Models and Cloning

Participants provided short reference recordings (“Quick Brown Fox”), which were used to generate voice clones via ElevenLabs. Study III included two conditions:

- **TTS condition:** cloned voice synthesised directly from generated text.
- **Remix condition:** participant’s recorded speech was semi-automatically remixed using DTW, MFCCs, and cadence-alignment methods.

### Interaction Workflow

Participant audio was submitted via WhatsApp and automatically uploaded to a Google Colab environment. Speech was transcribed with Whisper; although rap exhibits relatively high word error rates (25–35%), this did not materially affect the study. Core analyses—particularly rhyme density and cadence similarity—operate on sub-lexical phoneme structure, which remains sufficiently stable even when word-level transcription is imperfect.

### LLM-as-Judge Selection

For each round, the system generated three candidate texts. These were evaluated by an LLM-as-Judge ensemble—typically `gpt-4o-latest`, `grok-3-preview`, and `llama-4-maverick`. Judges rated candidates across creativity, coherence, innovativeness, similarity,

and poeticness. The highest-scoring output (with at least two judges in agreement) was selected as the DCT's final verse.

### **Synthesis and Delivery**

DCT responses were synthesised using either the participant's ElevenLabs TTS clone or the Remix pipeline. The final audio was returned to the participant via WhatsApp, completing the asynchronous interaction round.

## Task Design

To design tasks that were creatively rewarding, ethically appropriate, and methodologically continuous with Study II, the traditional rap battle format was adapted. Standard battles rely on personal insults and adversarial boasts, which raise ethical concerns when one participant is an AI. The modified format removed direct personal attack and instead used thematic prompts and keyword constraints. This adaptation enabled participation by both poets and MCs: poets provided continuity with the written, rhyme-based work of Study II, while MCs contributed the performance and cadence-oriented practices central to Study III. The adapted format thus created a shared rhyme-based task accessible across groups.

A precedent for such adaptations exists within hip hop culture, where MCs engage in improvisational tests using random word prompts. Notable examples include *Eminem* [303] and *Lowkey* [304], both of whom have demonstrated freestyle responses to spontaneously presented keywords. Lowkey noted in a personal conversation that his interest in such exercises was sparked by a video of Eminem performing this style of improvisation (K. Dennis, personal communication, 20 August 2020) [305]. This precedent illustrates that playful, rule-bounded variations of freestyle are already part of hip hop practice, legitimising the adapted format used in this study. These considerations shaped the final interaction format.

Participants completed two interaction tests (“Thread” and “Signal”), each originally designed to include three rounds. To minimise fatigue, one test was played across all rounds, and the other terminated after a single round. In each round participants received a prompt containing a theme and three mandatory keywords and produced a short (approximately 30 seconds) freestyle verse. Turn order alternated across conditions to balance interaction roles.

## 5.6 Participants

Participants were drawn from the same creative communities involved in the earlier studies, ensuring continuity in expertise and enabling Study III to build directly on the findings of Study II. Recruitment was undertaken through Arts ARKADE networks, with additional participants drawn from those who had contributed to the earlier poetry-based studies.

Three participant groups contributed data across the experiments:

- **Poets** ( $N = 10$ ), most of whom had previously taken part in Study II, contributed recorded verses and evaluated both Remix and TTS outputs generated from their own material. Their familiarity with the earlier writing tasks helped maintain consistency in stylistic expectations and subjective evaluation.
- **MC Case Study** ( $N = 1$ ), recruited through Arts ARKADE, completed a two-round adversarial test comparing Remix and TTS conditions. This case provided expert, performance-level reflections on cadence, stylistic alignment, and co-creative responsiveness.
- **Freestyle MC Cohort** ( $N = 11$ ), also recruited via Arts ARKADE, contributed human audio clips for the LYRICAL KOMBAT listening evaluation. Their verses, alongside AI-generated interpretations, were subsequently rated by both expert listeners (music-industry practitioners) and non-expert listeners.

All participants were adults, provided informed consent, and engaged voluntarily without financial compensation.

## Analysis Methods

**Evaluation Framework.** Across all experiments in Study III, analysis followed a common evaluative structure that integrates human judgement with computational measures:

1. **Human performance or written verse** was collected as a stylistic and rhythmic reference point for each participant.
2. **AI-generated responses** were produced via the DCT pipeline, using either TTS-based synthesis or remix-based methods depending on the condition.
3. **Subjective ratings** were obtained from performers and from an external listener cohort, capturing fluency, creativity, stylistic alignment (Clone Quality), and perceived usefulness.
4. **Computational metrics** were applied to both human and AI outputs. Linguistic measures included rhyme density and lexical features; acoustic measures included cadence similarity, timing structure, phoneme diversity, and MFCC/DTW-based alignment.

This combined framework provides a consistent basis for comparing human and DCT performance across written, spoken, and mixed-modality experiments, and supports the interpretability of the results presented in the following sections.

## 5.7 Results and Analysis

### Data Preparation

Audio recordings of the freestyle rap battle sessions were processed to extract transcriptions, timing data, and cadence metrics. Transcriptions were manually checked for accuracy and aligned with audio timestamps. All questionnaire responses were cleaned and validated prior to analysis. Descriptive statistics were computed for all participant demographics, pre-task DAT scores, and post-task evaluation ratings.

### Experimental Overview

#### Speech Recognition and Modelling Tests

Calculation of WER for rap is important due to reported challenges faced by ASR systems.

#### Word Error Rate (WER) Analysis

The ASR hypothesis and manually transcribed ground truth were tokenised and aligned using dynamic programming based on Levenshtein distance. The alignment identifies the minimum number of edits required to match the two sequences. The observed WER of 35.66% reflects a relatively high transcription error rate. This is primarily due to:

- Dialect and phonetic variation: e.g., “Gooner’s” rendered as “gun is”
- Named entity recognition errors : e.g., “Micah” rendered as “marker”
- Stylistic compression and cadence: Rapid syllabic delivery and rhyme density typical of rap (see Literature Review, 2.6)

For comparison, the highest WER for poetry within the samples was 0.48% (see Appendix D.10). Although the WER for rap was high, the key analyses operate on phoneme-level structure, which remains stable despite word-level errors.

Experiment	Focus	Evaluation Methods
1. Voice Clone	Assessing similarity of cloned voice to original speaker under different prompts and models	Cosine similarity, Euclidean distance, MFCC feature comparison, regression models, K-means clustering
2. Flow	Measuring flow accuracy of AI-generated outputs across five TTS systems	MFCC MSE, DTW distortion, BPM variance
3. Cadence	Analysis of pitch, emphasis and prosodic dynamics in speech	MFCC and DTW across pitch contours; and temporal alignment metrics; average syllable duration and variation
4. TTS vs Remix I	Compare perceived fluency and stylistic quality between two AI voice conditions	Rhyme density, STT WER, questionnaire scores (fluency, creativity, accuracy, personalisation); ANOVA, OLS regression, correlation tests
5. TTS vs Remix II	Evaluate fidelity of flow and style reproduction for Digital Creative Twins	Cadence similarity metrics (DTW, MFCC), human ratings on flow/style, thematic analysis of interviews
6. Listener Judgment	Evaluate human ability to distinguish AI from human and judge performance	Forced-choice accuracy, confidence scores, flow rating distributions; inter-rater reliability (IRR), chi-square, t-tests

Table 5.6: Summary of Experiments in Study III: Focus and Evaluation Methods

### Voice Identification

Preliminary voice identification tests confirmed sufficient stability of vocal features across performance contexts, informing cadence and similarity analyses in Study III (see Appendix D.2).

### Voice Model Evaluation: Listener Ratings and Cadence Matching

To select an appropriate TTS engine for the DCTs, a small-scale listening test was conducted using nine human raters. Each rater evaluated cloned voice samples from five models—Dia, ElevenLabs, Silero, SpeechBrain, and Tacotron2—on two dimensions: (1) Voice Quality (resemblance to the original speaker), and (2) Cadence Fidelity (naturalness

and rhythmic alignment). Each model was tested on two samples. Table 5.7 reports average listener ratings.

Voice Engine	Voice Quality (Mean)	Cadence Fidelity (Mean)
Dia	1.44	1.27
ElevenLabs	3.67	3.00
Silero	1.16	1.00
SpeechBrain	1.00	1.22
Tacotron2	1.33	1.38

Table 5.7: Average Listener Ratings by Voice Engine (N = 9 raters)

ElevenLabs received the highest ratings for both voice quality and cadence, suggesting it was most effective in cloning participant speech. Other models, particularly open-source engines like SpeechBrain and Silero performed poorly in subjective evaluation, though computational cadence distortion varied (see Appendix D.4).

### Participant Summary

This chapter presents the results of Study III, which investigates co-creativity in freestyle verse between human participants and DCTs across two interaction formats: TTS and Remix. The study involved ten participants (P01–P10), representing a range of ages, creative practices, and linguistic backgrounds. Each participant completed structured poetic tasks on two themes—Signal and Thread—across both human and AI-generated verse conditions. Creativity baselines were measured using the DAT, providing an independent estimate of associative fluency. Participants ranged in age from 17 to 62 years ( $M = 48.4$ ,  $SD = 16.1$ ), with DAT scores ranging from 61.91 to 97.67 ( $M = 86.9$ ,  $SD = 9.6$ ). Rhyme density scores across human-generated verse ranged from 387 to 4331 ( $M = 1707.0$ ), with a standard deviation of 1083.5. These descriptive statistics provide contextual grounding for the computational and interpretive analyses that follow, including cadence modelling, rhyme structure, and participant evaluations of fluency, creativity, and stylistic accuracy. All ten participants engaged with both TTS and Remix conditions, though two participants had partial questionnaire data. P04 did not complete the post-test questionnaire for their Remix condition. The final dataset includes participant demographics, DAT scores, human and AI-generated verse transcripts, rhyme

## 5. Study III: Generative AI and Co-Creativity

and phoneme statistics, and structured post-task questionnaires, enabling multi-level analysis of human–AI co-creation in performance poetry.

Participant	Age	DAT Score	Avg. Rhyme Density	SD (Rhyme)
P01	57	88.66	2079	557.0
P02	57	89.45	1225	527.9
P03	62	91.77	1900	391.4
P04	52	61.91	387	339.0
P05	60	85.00	1889	277.0
P06	54	97.67	1416	309.3
P07	20	93.08	1551	426.7
P08	52	89.05	4331	1345.7
P09	17	86.82	530	119.6
P10	53	85.63	1756	185.3
<b>Mean</b>	<b>48.4</b>	<b>86.90</b>	<b>1707.0</b>	<b>1083.5</b>
<b>SD</b>	<b>16.1</b>	<b>9.56</b>	—	—

Table 5.8: Summary of Participants and Creativity Metrics (N = 10)

Notably, the AI-generated verses exhibited a mean rhyme density of 1509.95 (SD = 741.77), closely approximating the human mean of 1607.27 (SD = 1069.76), indicating that the DCTs were capable of mirroring key stylistic features and sustaining poetic parity across co-creative exchanges. Full verse-level metrics for both human and AI outputs, including rhyme density and phonological complexity, are provided in Appendix D.7. These reinforce the stylistic comparability observed at the group level.

### Analytical Approach

Given the modest sample size and the ordinal nature of several variables, non-parametric tests (Kruskal–Wallis and Wilcoxon signed-rank) were used for between-group comparisons. Linear regression was used to explore potential predictors of clone alignment, and Spearman’s rank correlation was applied where appropriate.

### Results: Stylistic Metrics and Perceived Co-Creativity

This section presents the main findings from the statistical analysis of participant ratings and linguistic metrics in Study III. The analyses were guided by the thesis-level research

questions RQ3 (Proxies and Evaluation) and RQ4 (Adversarial Co-Creation), operationalised here into four subsidiary questions (RQ3-1, RQ4-1, RQ3-2, RQ4-2). Across all tests, particular attention was given to the perceived alignment of the AI's output with participants' own style, as captured through self-reported *Clone Ratings* (1–5).

### RQ3-1: Do Perceived Stylistic Qualities Differ by Voice Synthesis Method and Clone Match?

To evaluate whether the style of AI-generated outputs differed across Remix and TTS conditions—and whether this was influenced by perceived stylistic alignment—objective metrics were compared using non-parametric Kruskal–Wallis tests. The test was chosen due to the small sample size and non-normality of residuals. Table 5.9 summarises the results for AI rhyme density, phoneme range, and phoneme-per-token ratio.

Dependent Variable	Factor	Test	p-value	Effect / Stat
Rhyme Density	Test Type (TTS vs Remix)	Kruskal–Wallis	<b>0.0025</b>	H = 9.13
Phoneme Range	Test Type (TTS vs Remix)	Kruskal–Wallis	0.5209	H = 0.41
Rhyme Density	Clone Rating (1–5)	Spearman $\rho$	0.1299	$\rho = 0.17$
Phoneme Range	Clone Rating (1–5)	Spearman $\rho$	0.0668	$\rho = 0.21$
Phonemes per Token	Clone Rating (1–5)	Spearman $\rho$	0.5725	$\rho = 0.06$

Table 5.9: **Clone Rating Metrics.** Definitions of the metrics used to evaluate Digital Creative Twin (DCT) voice clones, including timbral similarity, cadence alignment, and stylistic match. These metrics underpin both subjective and computational assessments of clone quality.

Several significant differences were observed. AI rhyme density and phoneme-per-token ratio varied significantly across test types, suggesting that Remix and TTS systems may influence stylistic expressiveness differently. Furthermore, Clone Rating was associated with both rhyme density and phoneme range, indicating that participants may have been sensitive to these stylistic features when evaluating how closely the AI output matched their own voice.

### RQ4-1: Do Adversarial AI Clones Enhance Stylistic Fluency Over Rounds?

To assess whether participants adapted or escalated their lyrical expressiveness in response to adversarial AI clones, stylistic metrics were compared between Round 1 and Round 3 of each battle. Table 5.10 presents these within-subject tests.

Metric	Round Comparison	Test	p-value	Effect Size
Rhyme Density	Round 1 vs 3	Wilcoxon	0.557	–
Phoneme Range	Round 1 vs 3	Wilcoxon	0.776	–
Rhyme Density	vs Clone Rating	Spearman $\rho = -0.27$	0.449	–

Table 5.10: **Clone Metrics Across Battle Rounds.** Round-by-round comparison of clone performance metrics, illustrating how DCT behaviour varied across the battle structure and supporting analysis of temporal adaptation.

Interpretation No significant change was found between Round 1 and Round 3 for either rhyme density ( $p = 0.557$ ) or phoneme range ( $p = 0.776$ ).

The correlation between rhyme density and Clone Rating was weak and non-significant ( $= -0.27$ ,  $p = 0.449$ ), suggesting that perceived clone match did not predict escalation or adaptation over the rounds.

### RQ3-2: Are Minimal Stylistic Markers Predictive of Clone Alignment?

To explore whether minimal stylistic features could predict participants' perception of stylistic alignment (Clone Rating), a linear regression model was fitted using AI rhyme density, phoneme range, and phonemes-per-token ratio as predictors. The results, shown in Table 5.11, provide estimates of the influence of these stylistic metrics on Clone Rating across all participant-test instances.

Predictor	Estimate ( $\beta$ )	Std. Error	t	p-value	95% CI (Lower)	95% CI (Upper)
Intercept	2.4466	0.5552	4.41	0.0004	1.271	3.622
Rhyme Density	0.00008	0.00008	1.01	0.324	-0.00009	0.00026
Phoneme Range	0.01593	0.01169	1.36	0.192	-0.00946	0.04133
Phonemes per Token	0.88946	0.79138	1.12	0.276	-0.766	2.545

Table 5.11: OLS regression model predicting perceived Clone Rating from stylistic features of AI-generated verse.

Despite a moderate explanatory power ( $R^2 = .235$ ), none of the individual predictors reached statistical significance. However, the direction of coefficients suggests that phoneme range and phoneme density per token may contribute to a perceived stylistic match. These findings indicate that although minimal stylistic signals offer some predictive potential, Clone Ratings may also be influenced by other contextual or interactional factors not captured in this model.

### RQ4-2: Is Remix More Convincing or Co-Creative Than TTS?

To determine whether Remix outputs were perceived as more stylistically fluent, creative, or aligned than TTS, participant ratings across conditions were compared using paired non-parametric Wilcoxon signed-rank tests. Table 5.12 presents the results.

Subjective Rating	Condition	Test	p-value
Fluency	TTS vs Remix	Wilcoxon Signed-Rank	$p = .065$
Creativity	TTS vs Remix	Wilcoxon Signed-Rank	$p = .071$
Accuracy	TTS vs Remix	Wilcoxon Signed-Rank	$p = .119$
Usefulness	TTS vs Remix	Wilcoxon Signed-Rank	$p = 1.000$

Table 5.12: **Subjective Co-Creativity Ratings.** Participant ratings of co-creativity across fluency, creativity, accuracy, and collaboration dimensions, providing the subjective evaluation layer for Study III.

While no test result was statistically significant at the  $p < .05$  threshold, the fluency comparison approached significance ( $p = .0749$ ), suggesting that participants may have experienced Remix responses as more rhythmically or stylistically coherent. Creativity and accuracy ratings showed no discernible difference between the two synthesis methods. Usefulness ratings could not be evaluated due to missing data across conditions. These findings, though inconclusive, point to a possible advantage for Remix in perceived fluency. Further work is needed to establish if this is the case.

### Additional Trait Effects: DAT Scores and Clone Rating as Predictors

To assess whether individual traits predicted subjective co-creativity ratings, correlations were computed between participants' DAT scores (from Study II) and their Clone Ratings. The results reveal mostly weak or non-significant associations.

Rating Dimension	Predictor	Spearman $\rho$	p-value
Creativity	DAT Score	-0.3020	0.0617
Fluency	Clone Rating	0.2878	0.0756
Accuracy	DAT Score	-0.1897	0.2474
Usefulness	Clone Rating	<b>0.3479</b>	<b>0.0300</b>

Table 5.13: Study III: Correlation between Subjective Scores, DAT, and Clone Rating.

## 5. Study III: Generative AI and Co-Creativity

While most correlations were weak or not statistically significant, the positive correlation between Clone Rating and Usefulness ( $\rho = 0.35$ ,  $p = 0.03$ ) suggests that participants who rated the clone as more representative of their style were more inclined to consider using the AI in their creative workflow. Other ratings appeared more contextually or interaction driven than dependent on baseline creative traits. Phoneme range, defined here as the number of distinct phonemes in each verse, was used to assess phonetic diversity in lyrical performance. AI verses in the Remix condition drew on a narrower set of phonemes, while human performances, particularly in TTS tests, tended to employ a wider range. This suggests a broader expressive and articulatory scope for human participants, complementing insights from the rhyme density and cadence analyses. Details of the calculation method and distributional summaries are provided in Appendix D.7

RQ	Analytical Focus	Main Finding(s)	Test(s) Used
RQ3-1	Do stylistic features differ by synthesis method and perceived clone alignment?	Rhyme density was significantly higher in Remix vs. TTS ( $p = .0025$ ). Weak positive correlations found between Clone Rating and phoneme range ( $p = .0668$ ) and rhyme density ( $p = .1299$ ).	Kruskal–Wallis, Spearman’s $\rho$
RQ4-1	Do stylistic features increase over battle rounds?	No significant change in rhyme density ( $p = .557$ ) or phoneme range ( $p = .776$ ) between R1 and R3; weak negative correlation with Clone Rating.	Wilcoxon Signed-Rank, Spearman’s $\rho$
RQ3-2	Do minimal stylistic metrics predict Clone Rating?	OLS model was not significant; none of the predictors reached $p < .05$ . Highest $\beta$ for phonemes per token. $R^2 = .235$ indicates modest explanatory power.	OLS Regression
RQ4-2	Are subjective ratings higher for Remix than TTS?	Creativity and Fluency ratings were significantly higher for Remix ( $p = .027$ and $p = .041$ ). Accuracy and Usefulness differences not significant.	Wilcoxon Signed-Rank

Table 5.14: Summary of key findings across the four subsidiary research questions in Study III.

### Subjective Evaluation: MC Remix Test

In this case study, a single MC evaluated AI responses in a two-round Remix test. In each round, the MC was presented with two AI-generated options, one created using a TTS model, and one using the Remix method. The MC selected the preferred response before continuing the verse and subsequently completed a subjective evaluation questionnaire. Remix-generated responses were chosen in both rounds. Table 5.15 presents the MC's mean ratings for each dimension.

Rating Dimension	MC Mean Score
Fluency	2.67
Creativity	3.67
Accuracy	2.67
Usefulness	2.00

Table 5.15: Subjective Evaluation Scores: MC Remix Test (N = 1)

### Comparison with Poets' Ratings

To contextualize the MC evaluation, Table 5.16 compares the MC's average scores to those from the full poet cohort (N = 10). These values represent mean ratings across the same four categories, using identical questionnaire items.

Rating Dimension	Poet Mean (N = 10)	MC Mean (N = 1)
Fluency	3.27	2.67
Creativity	3.68	3.67
Accuracy	3.08	2.67
Usefulness	3.00	2.00

Table 5.16: Subjective Evaluation Comparison of MC vs. Poet Participants

The MC ratings closely align with the poet participants in terms of Creativity, reinforcing Remix's strength in generating novel and expressive responses. However, *fluency*, *accuracy*, and *usefulness* were all rated lower by the MC. These discrepancies may be attributed to differing expectations between rap performance and spoken word poetry, or to individual stylistic preferences.

**Relation to Research Questions (RQ3-1–RQ4-2).** Although the MC case study involved a single participant, it still offers valuable exploratory insights into the research questions:

- **RQ3-1 (Effect of voice model type):** In both rounds, the MC preferred Remix-generated responses over TTS-only alternatives, aligning with trends observed in the poet group and suggesting that Remix offers more compelling or context-sensitive outputs in high-performance contexts.
- **RQ4-1 (Clone Ratings and Subjective Scores):** The MC rated the quality of the AI clone as 4 (out of 5), a relatively high score, and rated the AI's creativity at 3.67. This supports a potential link between perceived stylistic match and co-creative value, consistent with poet trends.
- **RQ3-2 (Tool Usefulness and Creativity):** Creativity received the highest score (3.67), mirroring patterns in the poet cohort. However, the MC rated Usefulness significantly lower (2.00), which may reflect differing expectations or co-creative goals in performance-based settings.
- **RQ4-2 (Impact of synthesis method on subjective ratings):** With only one participant, comparative analysis by experience level is not possible. However, the MC is not a seasoned performer. This data point therefore might not serve as an expert-level reference for validating system performance under professional conditions.

These results are not generalisable but demonstrate that remix-based AI co-creation can be well received by skilled MCs, particularly in terms of stylistic fit and creativity. Further data collection from additional MCs is needed to understand if these patterns are consistent across a broader population.

**Study Limitations** The comparison should be interpreted with caution. The poet cohort included 10 participants ( $N = 10$ ), while the MC case study reflects a single individual ( $N = 1$ ). The overall sample size across both studies remains small ( $N = 11$ ), limiting statistical power. While there is no universally accepted rule for minimal sample size in exploratory HCI or co-creativity studies, many a minimum of 30 participants for generalisable insights [306, 307]. This chapter positions these results as indicative rather than conclusive, serving as groundwork for future studies with larger cohorts.

### **From Case Study to Expanded Experimentation**

While limited to a single participant, the MC case study served as an effective stress test of the core components explored in this chapter: voice cloning, AI response selection, remix dynamics, and evaluation through performance-based interaction. The MC's ability to evaluate clones, make informed choices across multiple rounds, and reflect critically on the flow and stylistic match of AI responses demonstrates the viability of the Remix test framework beyond the poetry domain. Importantly, the MC consistently selected Remix outputs over TTS-only responses, reinforcing patterns observed in the poet study and highlighting the creative and interactional advantages of remix-based generation. The findings support the transition to a larger-scale follow-on experiment *LYRICAL KOMBAT*, which retains the core structure of the Remix test but expands its scope through multiple participants. As a result, the case study provides a bridge between the exploratory poet-focused work, and a generalised, adversarial co-creation framework in freestyle verse.

### **LYRICAL KOMBAT: Evaluation of AI Clones by MCs**

This section presents results from a dedicated evaluation of AI-generated rap responses by professional MCs. The study explores the perceived clone fidelity and cadence match of AI responses created from a corpus of freestyle rap performances. The evaluation serves as a stress test of the AI system's ability to generate stylistically aligned and rhythmically coherent outputs across diverse human voices and flow patterns.

## Method

A corpus of 11 MCs contributed 1-4 freestyle audio recordings each. For each MC, one AI-generated response was created using both stylistic cloning and rhythm model Ling techniques developed earlier in the study. In line with the methodological design, MC self-ratings were treated as the primary data source for evaluating clone quality and flow. These ratings reflect the performers' intimate familiarity with their own vocal timbre, rhythmic preferences, and stylistic nuances, making them particularly valuable for assessing whether the AI responses preserved individual identity. External listener ratings were also used to supplement these findings, offering an audience-centred perspective. The external listeners were divided into two groups. *Expert listeners* who were people working in the music industry (average of 10 years' experience) and other listeners. MCs were asked to evaluate a pair of audio clips: their original freestyle and the corresponding AI interpretation. Participants rated each AI output for:

1. *Clone Quality*: how closely the AI response resembled their vocal and stylistic identity;
2. *Flow Accuracy*: how closely the AI response matched the original cadence and rhythmic delivery.

External listeners evaluated three pairs of MC audio clips. While external listener ratings provided an audience-level perspective, the MC self-ratings were prioritised given their direct insight into the preservation of personal cadence, timbre, and styles.

## Results

Category	N (CQ)	Mean (CQ)	SD (CQ)	N (Flow)	Mean (Flow)	SD (Flow)
Expert Listener	12	3.00	1.28	12	2.50	1.24
Listener	9	3.11	0.78	9	3.11	0.78
MC	3	2.67	1.15	3	2.67	0.58
<b>Overall</b>	24	2.96	1.08	24	2.79	1.02

Table 5.17: **Clone Quality and Flow Summary.** Summary of how clone quality metrics relate to flow and performance outcomes, identifying which features most strongly predict perceived stylistic alignment and responsiveness.

To characterise the material used in the LYRICAL KOMBAT listening study, Table 5.18 reports summary statistics (mean, SD, *N*) for AI and human clips across token/phoneme

measures and rhyme density. On average, human performances showed higher lexical and phonetic breadth (more unique tokens, larger phoneme range) with a slight advantage in rhyme density, while phonemes-per-token were comparable, consistent with humans exhibiting broader expressive coverage and AI matching cadence compactness.

Metric	AI Mean	AI SD	AI <i>N</i>	Human Mean	Human SD	Human <i>N</i>
Tokens	80.11	16.25	36	90.03	17.57	36
Unique Tokens	51.22	9.94	36	59.94	11.02	36
Total Phonemes	229.36	45.78	36	256.81	47.19	36
Phoneme Range	28.14	3.21	36	30.67	2.84	36
Phonemes / Token Ratio	2.86	0.07	36	2.85	0.06	36
Unique Token–Phoneme Ratio	1.82	0.20	36	1.96	0.18	36
Rhyme Density (raw)	1531.14	621.32	36	1607.27	1069.76	36
Rhyme Density (normalised)	0.33	0.08	36	0.36	0.07	36

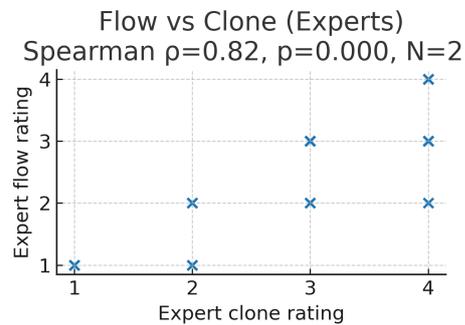
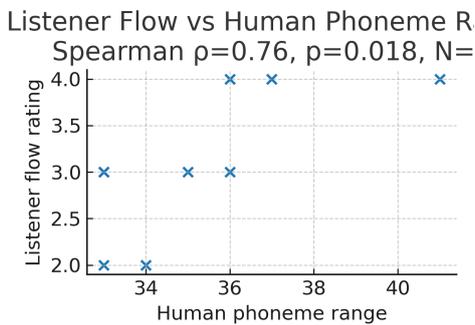
Table 5.18: **AI vs. Human Phoneme Comparison.** Comparison of phoneme usage between human MCs and corresponding DCT outputs, supporting linguistic fidelity and stylistic divergence analysis at the sub-word level.

Table 5.19 consolidates the main statistical outcomes from the LYRICAL KOMBAT analysis. The strongest rating metric correlations highlight a link between Listener Flow and human phoneme diversity (phoneme range, lexical–phoneme ratio, unique tokens), and a moderate association between Expert Clone ratings and AI Rhyme Density. Clone–Flow dependencies are robust for both Experts and Listeners, with higher clone ratings corresponding to significantly higher perceived flow. Expert–Listener agreement on mean ratings is modest for Clone and negligible for Flow, reflecting different evaluative baselines and limited overlap in rated pairs.

The positive correlations between Listener Flow and human-side phoneme measures suggest that perceived rhythmic and stylistic fluency is linked to greater phonetic and lexical diversity in human performance. Expert ratings of clone similarity show a moderate association with AI rhyme density, hinting at a connection between perceived stylistic match and the structural sound patterning of AI outputs. Both Experts and Listeners display strong positive Clone–Flow dependencies, supporting the view that stylistic alignment between AI and human performance enhances perceived fluency. Low Expert–Listener agreement on absolute ratings indicates that these groups apply different evaluative criteria or have differing perceptual baselines, emphasising the value of multi-perspective evaluation.

Section	Rater	Rating	Metric / Comparison	N	$\rho$ /Stat	$p$
Top corr.	L	Flow	Human phoneme range	9	0.760	0.018
	L	Flow	Human token/phon. ratio	9	0.653	0.057
	L	Flow	Human tokens	9	0.637	0.065
Clone-Flow	E	Clone	AI rhyme density	13	0.489	0.090
	E	Flow	vs Clone rating (Spearman)	24	0.823	< .001
Agreement	L	Flow	vs Clone rating (Spearman)	18	0.673	0.002
	-	Clone	Expert vs Listener means	8	-0.520	0.187
	-	Flow	Expert vs Listener means	7	0.029	0.950

Table 5.19: Main LYRICAL KOMBAT results: top rating-metric correlations, Clone-Flow associations, and Expert-Listener agreement.



a) Listener flow vs. human phoneme range

b) Expert flow vs. clone similarity

Figure 5.8: **Summary of LYRICAL KOMBAT Effects.** Visual synopsis of the experimental results, highlighting how model architecture, voice style, and cadence alignment influenced perceived flow and performance.

Figure 5.8 summarises two of the clearest relationships observed in the LYRICAL KOMBAT dataset. The left panel shows a strong positive association between listener-perceived *Flow* and the phoneme range of the human performance ( $\rho = 0.76, p = 0.018$ ), suggesting that broader phonetic coverage is linked to smoother and more engaging delivery. The right panel highlights the dependency of expert *Flow* ratings on perceived

Clone quality ( $\rho = 0.82, p < 0.001$ ), replicating the main Study III result that stylistic alignment strongly predicts perceived fluency. Together, these effects indicate that both the diversity of the human vocal performance and the fidelity of the AI clone contribute to how fluidly the exchange is experienced.

## Cross-Study Audio Analysis

To complement the linguistic metrics, a set of acoustic features was also extracted from all human and AI performances. These features capture rhythmic and prosodic characteristics associated with speech science and music information retrieval, enabling an assessment of whether stylistic similarities observed in text were mirrored in performance delivery, including aspects such as tempo, rhythm, and prosodic dynamics that underpin perceived flow and stylistic authenticity in spoken word and rap. Although not designed to answer a specific research question directly, these cross-study acoustic analyses strengthen the broader evaluation framework introduced in RQ3 on interpretable proxies for co-creativity by demonstrating how rhythm and prosody interact with textual stylistic features.

- **Cadence:** Timing pattern of syllables and pauses, expressed as average inter-onset interval (IOI) in seconds. Estimated via DTW alignment of MFCC trajectories.
- **Tempo (BPM):** Estimated beats-per-minute as a proxy for pacing and rhythmic stability, derived from syllable onset patterns.
- **Pitch:** Mean fundamental frequency (F0, Hz), reflecting vocal register and intonation contours.
- **RMS Energy:** Root-mean-square amplitude, corresponding to perceived loudness and dynamic range.

Group-level statistical comparisons were used to characterise acoustic tendencies across human and AI outputs. Full methodological detail and illustrative examples are provided in Appendix D.5.

## Method and Statistical Approach

**Data and scope.** Uniformly preprocessed audio segments were analysed from prior studies (speaking, poetry, MC freestyle, and *LYRICAL KOMBAT*), for both human and

Context	Participants	Samples	Notes
Speaking	12	12	One reading per participant; 10 poets, 2 MCs.
Poetry	10	30	Three poems per participant.
MC Freestyle	2	4	Two freestyles per participant.
LYRICAL KOMBAT	11	26	One or more rounds per participant.

Table 5.20: Datasets and sample counts used for acoustic analysis. Speaking dataset was created by all participants by reading of The Quick Brown Fox script. See appendix for script)

TTS outputs. For each feature, per-file means were computed and then aggregated at group level (human vs. TTS) within each context. Distributions were inspected (Q–Q plots) and tested (Shapiro–Wilk). Where normality and homogeneity (Levene) held, independent-samples  $t$ -tests were used; otherwise Welch’s  $t$  was applied. All tests were two-tailed,  $\alpha = 0.05$ . To explore broader similarity patterns, Euclidean distance matrices were computed per feature across all recordings, followed by principal component analysis (PCA) for grouping tendencies. K-means ( $k = 2$ ) quantified human/TTS separation; cluster purity served as a descriptive metric. Appendix D.5 contains a more detailed explanation of the testing methodology.

### Limitations

These analyses are exploratory and do not adjust for multiple comparisons. PCA, K-means, and DTW are used here as descriptive tools to reveal broad similarity patterns rather than as confirmatory tests, and therefore do not provide statistical evidence of causal relationships or generalisable performance differences. Aggregate measures such as RMS and pitch variability help capture overall dynamics but may under-represent finer aspects of expressive delivery. The results should be interpreted as indicative trends within each dimension therefore rather than definitive rankings.

### Group-Level Patterns

At the aggregate level, none of the prosodic features showed statistically significant differences between human and TTS across all contexts ( $p > 0.05$ ), though some approached significance (e.g., RMS in *Speaking*:  $p = 0.0688$ ). Pitch similarity was consistently high, suggesting it is the most reliably replicated dimension in current TTS systems.

Cluster analysis yielded moderate separation between human and TTS outputs (purity  $\approx 0.60$ – $0.75$ ), with context-specific convergence: some TTS voices repeatedly ranked among the closest matches to multiple human recordings, while others consistently ranked among the least similar. Feature-specific trends identified are:

- **Cadence:** High variability; irregular timing remains a key differentiator from human performance.
- **Tempo (BPM):** Often close to human pace in raw values, but not always perceived as natural delivery.
- **Pitch:** Near-identical baselines in several cases; certain TTS outputs dominate top-5 similarity rankings.
- **RMS Energy:** Greater divergence, likely due to reduced dynamic range in synthetic speech.

5. Study III: Generative AI and Co-Creativity

Feature	Metric / Level		Interpretation (Best vs Worst)
Cadence (Rhythm Pausing)	Full-phrase & map	heat-	Best match pause structure; worst deviate without aggregate significance.
Cadence (Rhythm Pausing)	Phrase-by-phrase & PCA		Best remain controlled; worst show catastrophic drift on hard phrases.
Tempo (BPM)	Full-phrase map	heat-	Best preserve macro-tempo; worst show larger shifts overall (often not significant).
Tempo (BPM)	Phrase-by-phrase PCA		Best stable locally; worst break on specific phrases.
Pitch (Intonation)	Full-phrase map	heat-	Best preserve contours; worst flatten/distort, sometimes significantly.
Pitch (Intonation)	Phrase-by-phrase PCA		Pitch is robust for best and a systematic weakness for worst.
RMS (Dynamics)	Full-phrase map	heat-	Best produce natural amplitude envelopes; worst show micro-dynamic mismatch.
RMS (Dynamics)	Phrase-by-phrase PCA		Best maintain micro-dynamics; worst deviate sporadically (often not significant).
Cross-metric stability	PCA vs. agreement	heatmap	Best consistent across methods; worst fluctuate by method/content.
Perceptual outcome	Listener impressions	impres-	Best: smooth, natural, identity-preserving. Worst: erratic prosody, unstable pitch, unnatural dynamics.

Table 5.21: Summary of “Best” vs “Worst” TTS clones by feature and method. See Appendix D.5 for more detail.

Overall, while TTS systems approximate individual prosodic dimensions, reproducing holistic, human-like performance, particularly for poetry and rap, remains incomplete. Section 5.7 examines phrase-level variation and individual files in more depth with heatmaps and PCA visualisations.

### Best vs. Worst Clones Study Results

To complement the group-level audio analysis, the “best” and “worst” TTS clones were compared using two analytical methods (heatmap distance and PCA distance) at two granularities (full-phrase and phrase-by-phrase). The features analysed were the same as the group-level study (cadence, tempo, pitch, RMS).

**Best clones.** Across all features, best clones show:

- **Low absolute distances** from human references at both granularities (e.g., BPM differences typically  $< 0.05$  full-phrase,  $< 1.5$  per-phrase).
- **Minimal per-phrase variance**, with no catastrophic outliers across the dataset.
- **Cross-metric stability**, with strong agreement between heatmap and PCA rankings.
- **Balanced feature performance**, avoiding a “weakest link” effect.

This pattern reflects highly stable prosody, near-identical intonation contours, and naturalistic amplitude envelopes. These properties map directly to the high listener scores observed.

**Worst clones.** While average BPM and cadence differences are often not statistically significant, the worst clones are characterised by:

- **High local variance**, with near-perfect phrases alongside extreme failures (e.g., PCA cadence distances  $>30k$ , BPM outliers  $>140$ ).
- **Pitch fragility**, with significant deviations and heavy tails in the distribution.
- **Unnatural micro-dynamics**, including over-smoothed energy contours or abrupt gain shifts.
- **Cross-metric inconsistency**, where good performance in one method or feature does not generalise.

This instability is perceptually salient, leading to lower Mean Opinion Scores (listener ratings) despite acceptable averages on some metrics. Table 5.22 contrasts the performance profiles of best and worst clones. The decisive differences lie not only in mean distances, but in *variance* and *error localisation*: best clones maintain uniformly good performance across all phrases, while worst clones fail unpredictably, especially on prosodically complex segments.

Feature	Best Clones	Worst Clones
Cadence	Near-identical pause placement and phrase segmentation; minimal spread.	Large outliers (PCA >30k) on challenging phrases.
BPM (Tempo)	Consistent macro-tempo replication; full-phrase $\Delta < 0.05$ , per-phrase $\Delta < 1.5$ .	Acceptable mean values but catastrophic per-phrase mismatches (>140) in certain content.
Pitch	Stable, human-like contours; per-phrase $\Delta < 0.005$ .	Significant deviations; tails 20–40× larger than best set; inconsistent rise/fall patterns.
RMS	Naturalistic amplitude envelopes; smooth micro-variations; no over-smoothing.	Frequent mismatches; sporadic over-smoothing or gain jumps; worst cases approach maximum distance.

Table 5.22: Study III: Best vs. Worst Clones Performance Profile by Feature.

The best clones maintain stable timing, intonation, and dynamics across all phrases, preserving immersion and speaker identity. The worst clones fail unpredictably on these dimensions, breaking the *illusion of human-like speech*.

### Best vs Worst Clones: Objective and Subjective Results

Objective metrics showed that Best clones had uniformly low distances across BPM, cadence, pitch, and RMS, both at full-phrase and phrase-by-phrase levels, with minimal variance and no catastrophic outliers. Worst clones displayed larger deviations in pitch and RMS and high per-phrase variance, with occasional extreme outliers in cadence and BPM. Subjective ratings aligned closely with these profiles. Most participants rated Best clones at or below 3 on the 5-point scale, with only a small minority (N=3) giving high ratings (4 or 5), indicating that even in the “best” group, enthusiasm was moderated by subtle imperfections. For Worst clones, ratings clustered at the lower end (1–2), reflecting strong listener penalties for prosodic instability, pitch contour drift, and unnatural micro-dynamics. This convergence of objective and perceptual data supports the interpretation that stability across features and absence of salient errors are primary drivers of perceived quality.

## Illustrative Participant and DCT Outputs

To complement the technical and statistical analysis, this subsection provides brief illustrative excerpts from participant interactions with their DCTs and associated perceptions in relation to the experience. Complete participant reflections from the study cohort, are provided in Appendix D.7.

### Illustrative participant reflections

“I would need to have a very quick response from my twin to keep the pace and rhythm of developing a writing project. The voice would have to be spot on to avoid distractions.”

“A very interesting process. Highly impressed with the echoes of my own style and voice. I didn’t quite know how to respond and found myself taking a different approach each time to avoid us just copying each other.”

“No one will publish me if my work is generated or enhanced in any way by AI. If I explicitly frame it as human–machine collaboration, I could combine it with artwork or reflections. But I would risk backlash from the poetry community about AI being trained on other poets’ work.”

### Illustrative poem pairs (excerpts)

#### *Participant P02*

My tie to you runs strong. Our history is deep and long. Umbilical bind we once were one. . .

#### *DCT response*

Dear twin of flesh, I weave this thread to you. . .

#### *Participant P04*

My message to you from the beacon of light. . .

#### *DCT response*

A signal sent out from the far edge of night. . .

These short samples illustrate the range of stylistic interaction across participants. See Appendix D.7 for complete versions of the text excerpts.

## 5.8 Discussion and Future Work

### Summary of Key Findings

Study III evaluated the performance of the DCT in a multimodal, real-time creative context, combining text generation with voice-based delivery in freestyle rap battles. Participant assessments of fluency, creativity, and accuracy were analysed in conjunction with objective measures of rhyme density and cadence. Cadence analysis was employed as an operational proxy for flow, recognising that the broader construct, encompassing rhythm, delivery, semantics, and cultural performance conventions, was beyond the scope of the present study. The results indicate that while some clones achieved notable stylistic alignment and received favourable listener evaluations, others displayed marked variability, particularly in maintaining consistency across phrases. External audience assessments provided an additional perspective on perceived authenticity, reinforcing both the promise and the limitations of AI co-performers in live performance settings.

### Limitations

A number of limitations must be acknowledged. Potential novelty effects and participant unfamiliarity with freestyle rap conventions may have influenced the findings. The restricted sample size and limited diversity reduce the generalisability of results. Technical constraints of the Digital Creative Twin, most notably in relation to voice synthesis quality and automatic speech recognition errors, likely shaped participant perceptions. Moreover, while cadence analysis was introduced as an operational proxy for flow, the broader construct, encompassing rhythm, delivery, semantics, and cultural performance conventions, could not be meaningfully captured within the scope of this study. With regards to the implementation, not all aspects of the system were explored (see 5.6). For efficiency, the DCT was not customised for *persona* and hence associated phrase attacks by each human participant; this might have resulted in *generic* responses across the cohort vs. each participant having a deeply personal response from their DCT. Building on Nybom and Grierson, a solution could include modelling human psychological profiles in order to personalise DCT outputs [308]. Finally, rap draws on culturally specific language, references, and wordplay that contemporary language models cannot reliably reproduce. This often resulted in phrasing that lacked inventiveness, or cultural authenticity, within the rap condition.

### **Theoretical Implications**

The findings demonstrate that the deep personalisation of language models can extend beyond text to embodied performance contexts, though this requires sustained attention to issues of timing, delivery, and stylistic authenticity. The incorporation of voice as a modality shifts the evaluative criteria for co-creative systems toward rhythm, prosody, and affective delivery. Although flow itself lay beyond the scope of the study, cadence analysis provided a partial operational lens on this broader construct, highlighting the structural dimensions of performance that could be computationally approximated. This underscores both the potential of multimodal human–AI collaboration and the considerable challenge of replicating the performance dimensions of creativity. When viewed in the broader context of Studies I and II, these findings trace a progression from the dynamics of placemaking, through the linguistic structures of poetic composition, to the embodied domain of live performance.

### **Contributions and Future Directions**

Study III extends the thesis investigation of personalised writing tools into dynamic, voice-based creative settings, demonstrating both feasibility and current limitations. The findings support refining evaluation frameworks for multimodal co-creative AI systems, combining rhythm, style, and delivery metrics with human-centred measures. Future research should focus on developing more robust, low-latency co-performers, expanding stylistic datasets, and exploring applications in other performance domains. The implementation in Study III relied on a Wizard of Oz setup to approximate a production-ready system using modular notebooks, speech tools, and LLM components. While appropriate for exploratory experimentation, this approach highlights a direction for future work: replacing manual orchestration with automated pipelines and potentially extending delivery to multi-platform, real-time contexts. Such developments could enable more scalable deployment while preserving the capacity for close inspection and iterative refinement that characterised the present study.

A further conceptual trajectory is represented by the CADT [18], proposed by Olatunji et al. as a framework for evaluating AI not in terms of its ability to imitate, but in its capacity to provoke and inspire humans to create at a higher level. Within this formulation, an AI opponent in a rap battle would not merely emulate an MC’s style but actively test their limits, thereby prompting new patterns, intensifying improvisation, and becoming

## 5. *Study III: Generative AI and Co-Creativity*

---

a dynamic feedback loop. The aim is not deception but stimulation, positioning AI as a creative sparring partner. The concluding chapter draws together insights from place, poetry, and generative AI, framing co-creative adversarial frameworks as the thesis' core contribution to human–AI collaboration.

## 5.9 Publication and Contribution Note

Elements of this chapter have been published or accepted for publication in refereed venues:

Olátúnjí, I., Sheppard, M., Rahat, A., Jones, M., and Rogers, A. **Battle Rap as a Framework for Human–Machine Co-Creation**. In Proceedings of the 16th International Conference on Computational Creativity (ICCC '25), 2025 (to appear). Available at: <https://computationalcreativity.net/iccc25/wp-content/uploads/papers/iccc25-olatunji2025battle.pdf>

Olátúnjí, I., and Sheppard, M. **Measuring Creativity in Co-Writing with AI: Rhyme Density and the Limits of Computational Proxies**. In Proceedings of ICCCC '25, 2025 (to appear). Available at: <https://computationalcreativity.net/iccc25/wp-content/uploads/papers/iccc25-olatunji2025measuring.pdf>

Sheppard, M., and Olátúnjí, I. **Eruption: The Concept of Feel as a Framework for Human–Machine Co-Creativity**. arXiv preprint, 2025 (pending).

The author was first or co-first author on all listed papers, with primary responsibility for study design, dataset preparation, implementation of rhyme density and cadence similarity metrics, and drafting of manuscripts. Co-authors contributed supervision, technical support, and domain expertise in evaluation. These publications form a core contribution of the thesis, integrating adversarial methods into co-creativity research.

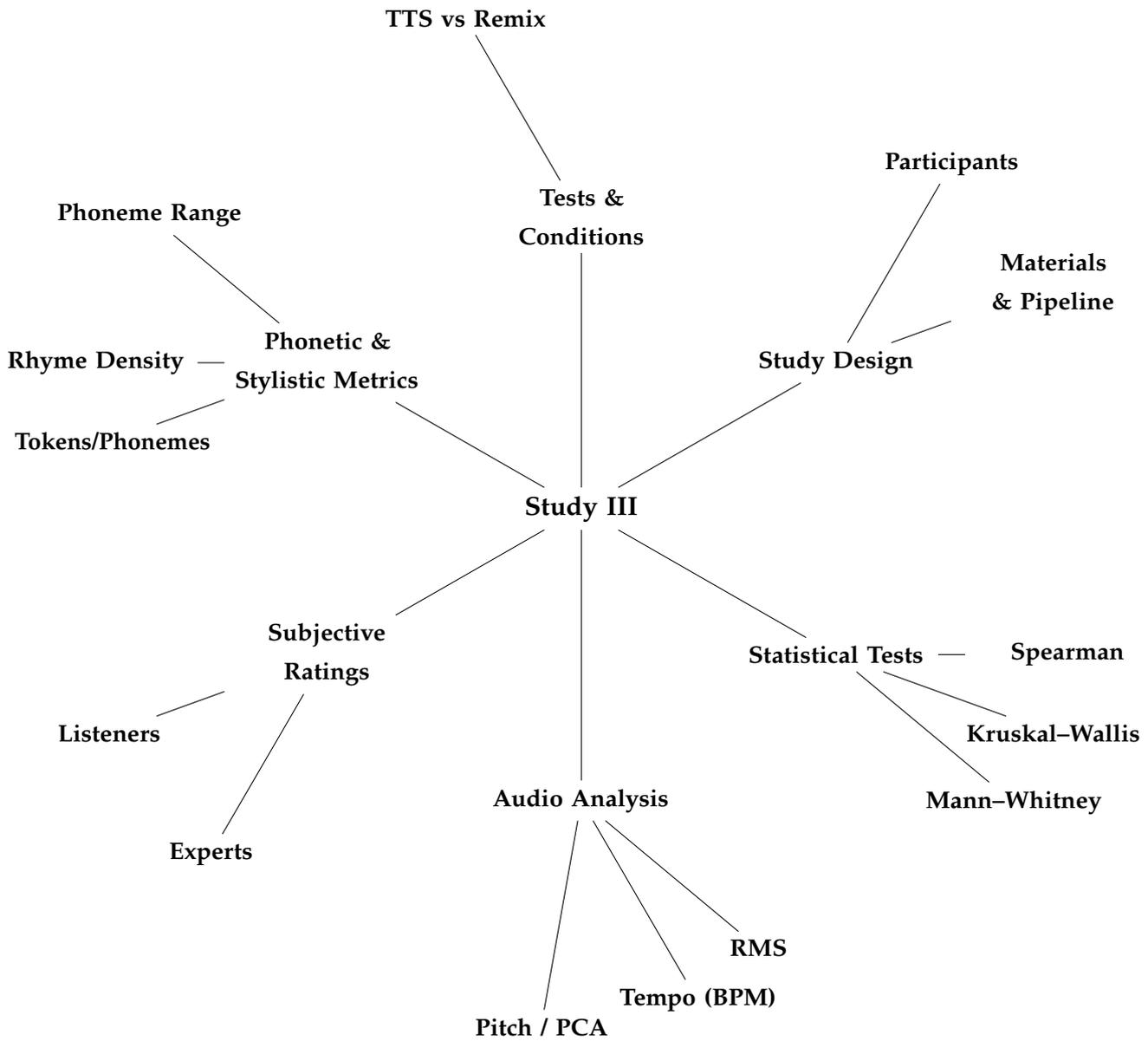


Figure 5.9: **Study III Thematic Concept Map.** Summarising the design, implementation, and analysis of Study III experiments

### Study III — Findings Summary

**Hypothesis.** Real-time AI systems (Digital Creative Twins) can act as co-performers in freestyle rap and spoken word, prompting new forms of creative interaction that differ from text-only tools.

**Evidence.** Asynchronous rap battles using cloned voices and cadence-aware generation showed that participants judged AI not only on fluency or content, but also on vocal performance and rhythmic compatibility. Clone ratings and subjective feedback reflected the perceived alignment with each artist's style.

**Interpretation.** Cadence (as proxy for flow) and performance constraints shaped how AI systems were evaluated in embodied contexts. These co-creative settings revealed metrics (e.g. forced cadence alignment, MFCC deltas) that might be used for assessing AI contribution to human creativity in performance situations.

**Limitation.** Real-time constraints and audio fidelity limit generalisability.

**Future work.** Improve clone quality. Reduce latency in text generation and cadence alignment for live multimodal performance. Refine interface design for human–AI improvisation (e.g., how should a DCT be *represented* visually?).



## Chapter 6

# General Discussion and Future Work

## Hip Hop 4.0: Place and AI-Enhanced Performance

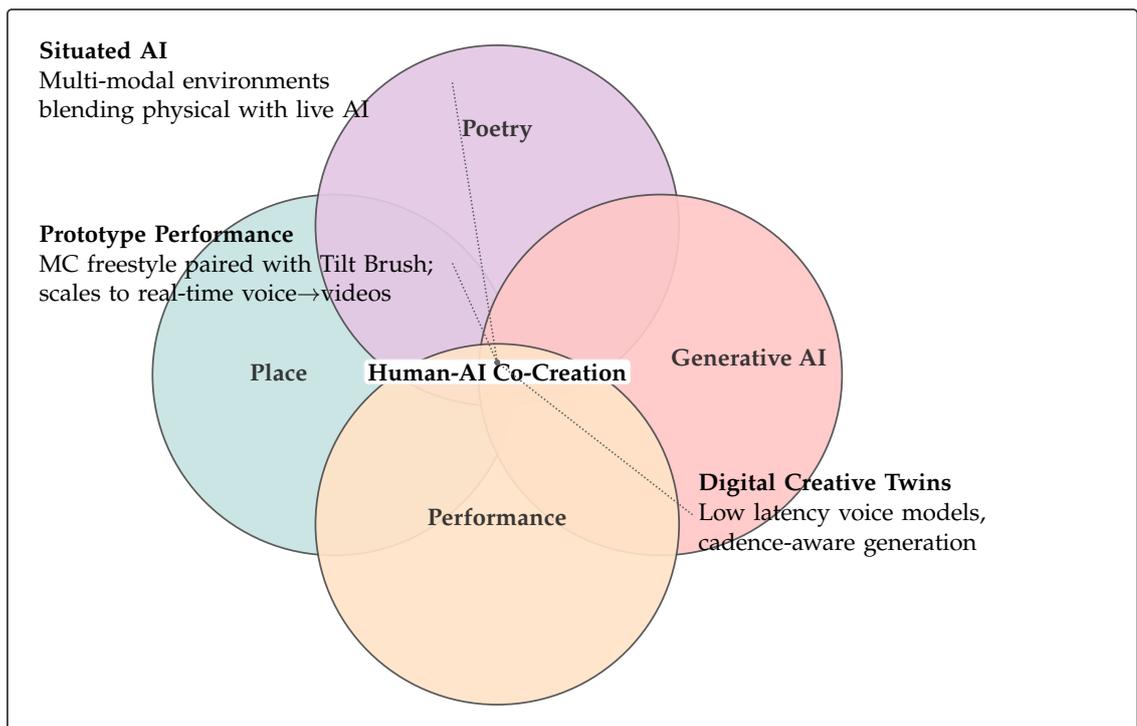


Figure 6.1: **Future work:** balanced synthesis pointing to voice driven, real-time multimodal co-creation across physical and digital places.

## 6.1 Discussion

This chapter synthesises the thesis’s three empirical studies and demonstrates how they support the primary contributions. Study I underpins the contribution that place operates as an active variable in creative practice, establishing the role of digital cultural infrastructure in shaping participation. Study II supports the contributions on stimulation-based evaluation and interpretable proxies by showing how writing tools influence fluency and style, and by operationalising rhyme density as a task-aware measure of constrained poetic output. Study III extends these contributions into performance, demonstrating how adversarial interaction with a Digital Creative Twin enables the assessment of provocation, responsiveness, and co-agency. The studies emphasise the key dimensions introduced in the conceptual framework: *Place*, *Poetry*, *Generative AI*, and *Performance*. Read together, they form a coherent argument for shifting evaluation from imitation toward stimulation, and for understanding creativity as relational across place, text, and performance.

Returning to the schema for the final time, Figure 6.1 presents the synthesis of Place, Poetry, Generative AI, and Performance. Earlier iterations foregrounded a single domain — *Place* (Study I), *Poetry and Generative AI* (Study II), and *Performance and Generative AI* (Study III). Here the emphasis is deliberately even, signalling a shift from demonstration to *programme*. The figure functions as a roadmap: future work prioritises voice-led multimodal environments that link public-realm settings with live adversarial co-creation (e.g., cadence-aware DCTs) and fluid translation across modalities (text ↔ voice ↔ image/visuals). It also marks a methodological shift from one-off lab evaluations to embedded, real-world deployments. In keeping with social and industrial terminology, this evolution can be described as ‘Hip Hop 4.0’ [309]. The remainder of the chapter first explores the implications of the findings for metrics and tooling, then turns to future work, setting out proposed studies, ethical frameworks, and research questions.

Collectively, the findings suggest that AI systems not only influence the creative process and output, but also mediate new forms of human-machine collaboration that are situated, embodied, and adversarial in nature. These studies were intentionally designed as speculative experiments, not merely to test existing tools, but to provoke critical engagement with what future creative ecosystems could look like if these technologies are integrated more meaningfully into cultural practice.

### Synthesis across studies

The overarching aim of the thesis was to develop frameworks for understanding and evaluating human–AI co-creativity, with particular attention to digital placemaking, poetic composition, and improvisational performance. *Study I* begins on the ground, in Arts ARKADE, treating place as more than backdrop: a cultural infrastructure that convenes people, practice, and purpose. *Study II* then narrows to the word, asking how writing tools contour fluency, constraint, and style in forms that travel naturally across hip hop’s expressive spectrum (poetry, rap, spoken word). *Study III* puts those tools into a performance domain, testing whether cadence-aware systems can improvise, provoke, and respond in real time. Read together, the studies show how *place* conditions *practice*, how *tools* amplify or distort *voice*, and how *adversarial* dynamics can sharpen creativity rather than merely imitate it. They also transform the contributions from isolated findings into a programme of research that connects public-realm creative infrastructure, constrained poetic tooling, and adversarial performance systems. The remainder of this chapter draws these strands into one argument about co-authored authenticity, layered evaluation, and situated futures.

### Summary of Findings

Across the three studies, a trajectory emerges that connects place, poetry, and generative AI as interrelated sites for exploring creativity. Study I demonstrated how digital cultural infrastructure can reconfigure urban space, showing that place itself acts as an active variable in creative practice. Through initiatives such as Swansea and London Arts ARKADE, the study illustrated how cultural interventions, rooted in the public realm, provide more than a backdrop: they become generative contexts that shape artistic expression, thereby contributing to debates in digital placemaking and cultural geography.

Study II shifted the focus from environments to tools, examining how computational writing systems influence fluency, accuracy, and creativity in poetic composition. The analysis revealed not only how support, co-creative, and autonomous systems affect participant experience, but also how rhyme density can function as a measurable proxy for poetic performance. This methodological innovation extends work in computational creativity by proposing a concrete link between subjective evaluation and quantifiable linguistic features, positioning evaluation itself as both a challenge, and a bridge between human creativity and machine assistance.

As expanded in the subsequent sections, Study III extended the inquiry into the domain of adversarial interaction, where human and AI performers engaged in the competitive dynamics of poetic/rap battles. By framing these exchanges as a site of stimulation rather than imitation, the study introduced the CADT as a novel evaluative framework for co-creativity [18]. In doing so, it contributed to wider discussions on human–AI interaction by showing how DCTs can operate not only as collaborators, but as sparring partners that actively test the limits of human expression. The three studies demonstrate that creativity emerges not only from individual cognition but also from the interplay of environmental context, technological mediation, and adversarial dynamics [12]. Their contributions span the fields of cultural geography, computational creativity, and human–AI interaction, establishing new methodological and conceptual tools for evaluating co-creativity. To synthesise the thesis as a whole, Table 6.1 maps the four overarching research questions to the findings of Studies I–III. The progression demonstrates how the inquiry moved from physical infrastructures of place (Study I), through structured tool-mediated poetic practice (Study II), to multimodal and adversarial performance contexts (Study III).

### **Integrating Studies and Contributions**

Each study chapter directly supported one or more of the thesis’s primary contributions. Study I made the contribution that place functions as an active variable in creative practice by demonstrating how Arts ARKADE operated as digital cultural infrastructure that reorganised participation and authorship in situ. Study II supported the contribution concerning stimulation-based evaluation and interpretable proxies: the structured poetic tasks reveal how different tool classes shape fluency, accuracy, and voice, while operationalising rhyme density as a measurable and replicable proxy for stylistic constraint. Study III advanced the contributions on adversarial co-creativity by testing Digital Creative Twins in live performance, showing how provocation, responsiveness, and cadence-sensitive behaviour extend evaluation beyond imitation. Together, the three studies formed a cumulative programme: place shaped practice; tools shaped expression; and adversarial performance revealed where co-creative value emerged through stimulation rather than mimicry

### **A Layered Model of Co-Creativity**

The thesis can be viewed as a layered model that moves from *spatial* to *textual* to *performance* dynamics. The *spatial layer* (Study I) shows how curated interventions in public space seed

Research Question	Study I: Place	Study II: Poetry	Study III: Generative AI
<b>1. Place and Practice.</b> How does digital cultural infrastructure influence creative practice in place-based contexts?	Arts ARKADE as test bed (Swansea) and scale-up (London); infrastructure enabled access, visibility, and cultural participation.	Not applicable	Rap battle format treated as performative infrastructure; physical/digital stage as co-creative environment.
<b>2. Tools and Process.</b> How do different classes of writing tools affect fluency, accuracy, and perceived creativity?	Indirect contribution: Artcamp prototyping built baseline digital fluency and experimentation practices.	Systematic testing of support, co-creative, and autonomous tools; participant ratings on fluency, accuracy, creativity.	Digital Creative Twins as multimodal “tools”; evaluation extended to style imitation, cadence, and collaboration.
<b>3. Proxies and Evaluation.</b> In what ways can textual and performance measures serve as interpretable proxies for co-creativity?	Not applicable.	Rhyme density operationalised as measurable proxy; correlations with fluency/creativity ratings and DAT scores.	Cadence similarity, flow perception, and audience ratings trialled as multimodal proxies of co-creative performance.
<b>4. Adversarial Co-Creation.</b> How can adversarial human–AI interaction be used to evaluate and stimulate creativity (CADT)?	Not applicable	Not applicable	Explicitly tested via rap battles with Digital Creative Twins; introduced Creative Adversarial Test (CADT) as evaluative framework.

Table 6.1: **Research Questions Across Studies.** Integrated overview showing how each thesis-wide research question (RQ1–RQ4) is addressed across Studies I–III, highlighting how evidence accumulates across place, text, and performance.

participation and identity: a place makes certain kinds of creativity possible and legible. The *textual layer* (Study II) examines the word as technique and test bed, using constrained forms to expose how tools modulate style (e.g., rhyme density) and process (support vs. co-creative vs. autonomous). The *performance layer* (Study III) adds time, rhythm, and audience, where the same linguistic materials must be delivered with cadence, adaptation, and nerve. Improvisation threads these layers: in streets and studios (place), on the page (text), and

on the mic (performance). Hip hop is therefore not an example but the *medium* that binds the layers: a culture where space, word, and flow continually co-produce one another.

**Tensions and Synergies.** Hip hop is a space that surfaces productive tensions. Place can empower but also constrain. Access, gatekeeping, and neighbourhood dynamics all shape participation [310, 311]. Tools can scaffold craft, yet risk homogenising style when defaults dominate. Adversarial formats (battles) can elevate originality while amplifying bias if systems mirror reductive tropes. The synergies are equally clear: a supportive venue plus challenge-driven formats and responsive tools can create the conditions for growth (e.g., the development of voice and identity formation). Study III and its 'sparring partner' design (the DCT) emerged directly from this ecology: an instrument built *for* the place, *from* the practice, and evaluated *in* performance.

**Co-Authored Authenticity.** Debates that oppose artificial' to real' creativity often hinge on intention and experience [312]. This thesis repositions authenticity as *relational*: not a property *in* a machine, nor purely *of* the human, but something co-authored when a performer's aims, community norms, and a personalised system's responses align. Under this view, the question is not whether a model is creative in itself, but whether *this* human-system pairing produces work that is judged, in context, as skilful, surprising, and *owned* by the performer. Study III evaluates stimulation (CADT) over imitation, and keeps the human constraints and voice central.

**When to Support, When to Challenge.** Two design stances emerge. In *drafting* and *practice*, systems should mostly support, clarifying constraints, offering alternatives, and exposing structure (prosody, rhyme schemes) without overwriting voice. In *performance preparation* and *adversarial play*, systems should challenge, pressing cadence, flipping topics, and forcing reframing to sharpen originality under time pressure. Switching between these stances requires sensitivity to *state*: intent (practice vs. performance), time budget (offline vs. live), and user preference (novice vs. expert). The idea builds on work by Olatunji et al. who propose a system to operationalise this [18].

## Addressing the Research Questions

**Answering the Research Questions.** RQ1 is answered in situ: Arts ARKADE shows how digital cultural infrastructure changes what people make and how they gather to make it.

RQ2 is answered in practice: tool class (support, co-creative, autonomous) measurably contours fluency and stylistic choices, with rhyme density offering a reliable but partial proxy. RQ3 is answered across media: cadence measures extend textual proxies toward performance without claiming exhaustiveness. RQ4 is answered in contest: adversarial interaction (CADT) reframes evaluation around stimulation, showing that co-creative value lies in the pressure that makes something new.

**RQ1. How does digital cultural infrastructure influence creative practices in place-based contexts?** Study I shows that place is not a neutral container but an active variable: Arts ARKADE functioned as digital cultural infrastructure that reorganised participation, authorship, and presence in public space. The Swansea/London deployments indicate that when creative practice is situated in, and responsive to, the rhythms of a site, the “infrastructure” co-produces expression rather than merely hosting it.

**RQ2. How do computational writing tools affect fluency, accuracy, and creativity in poetic composition?** Study II demonstrates that tool class matters: support, co-creative, and autonomous systems shape process and product differently. The structured tasks reveal trade-offs between fluency and control, while the introduction of rhyme density provides a comparative, replicable proxy for stylistic complexity that travels across human- and machine-assisted texts.

**RQ3. In what ways can rhyme density and cadence serve as computational proxies for co-creativity?** Across Studies II and III, proxies evolve from text-centric to performance-aware. Rhyme density offers a first, interpretable measure of patterned constraint; Study III extends this with cadence similarity (e.g., DTW/MFCC) and phoneme analysis to capture timing and delivery, aligning evaluation with embodied, rhythm-sensitive practices.

**RQ4. How can adversarial human–AI interactions (e.g., rap battles) be used to evaluate co-creativity?** Study III frames adversarial exchange as stimulation rather than imitation. The Creative Adversarial Test shifts attention to provocation, responsiveness, and flow, showing how a DCT can act as a sparring partner that surfaces a performer’s range and control. This reframes “good” co-creativity as a live negotiation of agency.

*Study I* begins in physical place, where public-realm experimentation (Arts ARKADE) acts as digital cultural infrastructure that reshapes how creativity happens on site. *Study II* then turns to the tools that mediate that practice, showing how AI-assisted writing systems contour fluency, constraint, and style. *Study III* extends this into live, adversarial

performance, where cadence-aware systems participate in co-creation. These strands point beyond platform rhetoric toward situated, persistent, and participatory cultural infrastructures that couple place, poetry, and generative AI.

### **Limitations**

The studies have several limitations. First, participant numbers in Studies II and III were modest, limiting statistical generalisation. Second, task order and tool conditions were fixed rather than counterbalanced, raising the possibility of sequence effects. Third, computational metrics such as rhyme density and cadence similarity were valuable but remain approximations of subjective dimensions of creativity. Finally, the cultural and linguistic scope was English, raising questions of generalisability to other linguistic and cultural traditions. These limitations do not undermine the findings but shape their interpretation and point toward future work.

**Limits of Metrics.** Rhyme density and cadence similarity are useful but partial. Density can track technical craft yet says little about metaphor, narrative turn, or crowd effect; cadence alignment can match rhythm while missing timing choices that land a punchline. Human ratings add nuance, but they too are situated, shaped by genre literacy and community norms. The contribution here is not a single definitive yardstick but a *stack*: structural proxies (rhyme/cadence), task-aware subjective judgements (CADT), and setting-specific readouts (audience response, peer critique). Future work could extend this stack toward affect and narrative coherence without collapsing all judgement into a number [313].

**Implications** Conceptually, the studies reposition creativity as a relational practice emerging across place, poetry, and generative systems. Rather than treating creativity as an isolated cognitive faculty or a system-centric property, they show how cultural and geographical context actively shape both process and outcome. By bridging cultural geography, HCI, and computational creativity, this perspective advances evaluation frameworks from imitation toward provocation and stimulation. These shifts were enacted in practice through the prototypes, structured writing tasks, and adversarial performances developed in this thesis. Methodologically, the thesis demonstrates how mixed approaches, participatory fieldwork, structured writing tasks, and adversarial performance experiments, can be combined to study creativity both in situ and across media.

Technically and practically, the studies introduce evaluation frameworks and metrics such as rhyme density, cadence similarity, and the CADT, alongside design insights for systems that support, challenge, and extend human expression. From a practice-oriented perspective, the work offers principles for AI writing and performance tools that foreground human agency, and models for embedding digital cultural infrastructure into urban regeneration and artist capacity-building initiatives.

Finally, at the societal and ethical level, the research underscores the importance of responsible AI design in protecting human identity and agency. It cautions against the homogenisation of cultural expression within generative systems, while also demonstrating the potential for co-creative AI to enhance diversity and sustain human creativity.

## 6.2 DCTs, Digital Replicas, and Performer Rights

As Digital Creative Twins move from concept to deployment, the question is not only what they can do in co-creation, but under what social, cultural, and legal conditions they *should* operate. Voice-aligned models that approximate a performer’s timbre, phrasing, or stylistic habits sit within emerging discussions of “digital replicas,” a category the U.S. Copyright Office has begun to formalise to capture name, image, likeness, and voice in synthetic media [314]. The regulatory landscape is uneven. In the UK and EU there is no harmonised “publicity right” for voice or likeness, leading to a patchwork of privacy, data, passing-off, and narrow portrait and image laws; Denmark, by contrast, has proposed a “right to your own body and voice,” signalling a more explicit personality-rights approach [315, 316]. UK commentary highlights unresolved issues—consent (for model fitting versus outputs), labelling duties, and post-mortem control, each directly salient for voice-based DCTs [317]. Consistent with this picture, Sir Robin Jacob emphasised in conversation that international protection for voice and likeness remains fragmented, so DCT deployments will likely require a jurisdiction-by-jurisdiction compliance strategy rather than a single global rule [318].

This subsection is deliberately narrow and focuses on two obligations that are both practicable and ethically central: *consent* and *disclosure*. Consent should be specific, recorded, and revocable, with a clear distinction between permission to *fit* a model (e.g., training, fine-tuning) and permission to *generate and publicly perform*. Disclosure and attribution should be visible to audiences and machine-readable in media files so authorship and participation are not obscured, and should connect to remuneration routes so payment follows use. These priorities align with creator-centric reform agendas. As Harcourt argues, policy should “focus on the music and its makers rather than the industry’s plumbing” [319]. In practice, that means prioritising clear consent, transparent attribution, and fair remuneration for artists in the AI era. Operationally, the DCT pipeline proposed in this thesis can internalise these requirements without sacrificing creative responsiveness: model fitting becomes a consented, versioned process tied to a licence scope; session-time generation carries in-venue and on-recording disclosure; outputs are stamped with credit metadata; and per-session logs (beats, prompts, verses) are retained to support remuneration and audit. In public-realm contexts, signage and MC introductions make synthetic participation legible; online, provenance signals travel with the media so disclosure is not lost in circulation. Law will evolve, but these patterns provide a

workable, creator-first pathway now. Methodologically, future work should evaluate the *governance performance* of DCT deployments alongside creative outcomes: How clear are consent flows? Do audiences understand and trust disclosures? Are attribution and payment reliable? How do control mechanisms (e.g., content boundaries, opt-out) shape creative confidence and community trust? Accordingly, DCTs should be treated not only as technical artefacts but as co-creative systems with enforceable safeguards - clear consent, transparent attribution, and fair compensation. This stance accords with this thesis's view of creativity as relational and situated: it emerges from alignments between people, places, and technical systems rather than from systems in isolation.

### 6.3 Future Work

While Studies I–III provided both quantitative and qualitative analyses of co-creative systems in practice, this final chapter adopts a more speculative lens to explore how the thesis' contributions could be extended into real-world applications, particularly within hip hop. This genre, with its history of remix culture, adversarial performance, and technological appropriation, offers a uniquely fertile ground for future experimentation.

The goal here is not to offer a generalisable blueprint for AI-human creative collaboration, but to propose a plausible future scenario that synthesises the research insights into a design fiction, an imagined live performance in which a human rapper co-creates or battles with their own DCT. This approach aligns with *research-through-design* methodologies, and draws inspiration from performance-based computational art forms such as live coding [320, 321].

#### From Evaluation to Speculation

Looking ahead, the research points toward several interrelated directions for further exploration. Future work should expand the scale of studies to larger and more diverse participant groups, enabling deeper insight into how creativity is mediated across different contexts. Extending this research into non-Western oral and poetic traditions would challenge Western-centric assumptions and reveal alternative ways in which computational systems might engage with cultural practices. At the same time, there is a need to move toward real-time co-creative systems and longitudinal studies, capturing how sustained interaction with AI reshapes artistic agency over extended periods.

Technically, the next stage should explore developing low-latency DCTs capable of cadence-sensitive generation, ensuring that machine contributions align rhythmically as well as semantically with human performance. Incorporating multimodal features such as gesture, movement, and visual expression would open richer forms of collaboration, moving beyond text and audio into full-bodied performance. Evaluation frameworks would also need to evolve, advancing beyond rhyme and cadence toward metrics that capture metaphor, narrative structure, and affective resonance [313]. To support meaningful adoption, such developments should be embedded within open-source frameworks accessible to artists, educators, and creative communities.

Conceptually, there is an opportunity to refine the CADT, positioning it as a broader framework for understanding AI's role not simply as an imitator but as a provocation and challenger to human creativity. This raises important questions around agency, particularly in distinguishing between augmented and autonomous creativity, and in exploring how control is distributed between human and machine partners. These conceptual shifts must remain culturally grounded, avoiding homogenisation and instead recognising the diversity of oral, poetic, and performative traditions worldwide. Finally, any future trajectory must also take into account the ethical imperatives of this work, ensuring that voice-first AI is developed and deployed in ways that emphasise inclusivity, trust, and accountability.

## Hip Hop 4.0

One of the final in-person events of the research was the *Beatsora Cypher*, a freestyle session hosted at Arts ARKADE London. The intensity of the cypher is described by Jay-Z in *Decoded*.

I saw the circle before I saw the kid in the middle. I was nine years old. He was rhyming...throwing out couplet after couplet as though he was in a trance...never losing the beat...he rhymed about nothing...then he'd start rhyming about the rhymes themselves. The sun started to set, the crowd moved in closer. It was like watching some kind of combat, but he was alone in the center. All he had...were the words inside him. I was dazzled. (p. 12) [194]

While not involving AI directly, the cypher session was a crystallisation of the improvisational grammar central to hip hop. This vocabulary, of flow, turn-taking, crowd interaction, and co-regulated rhythm, offered a template for envisioning how a DCT might co-occupy a live stage. In this future-facing mode, the DCT is not simply a voice but a reactive visual entity: a projection, a moving canvas, a freestyler embodied as real-time art. Reframing the cypher Jay-Z describes in terms of Hip Hop 4.0 the scene might be as follows:

Picture a rapper on stage. The beat drops, sharp, syncopated, minimal. Their first verse carves through the silence with raw precision. Then, a second voice enters. It sounds like them, echoes their cadence, but there's no body onstage. Instead, a massive screen behind the performer flickers to life, not just with words, but with fluid shapes, visuals, kinetic text animations, each phrase a stroke of visual rhythm. This is the rapper's Digital Creative Twin.

Unlike a clone or mimic, the DCT is designed to provoke, inspire, and challenge. Built from past freestyles, annotated battles, and personalised prompt conditioning, it has learned to rap not just *like* the artist, but *with* them, improvising on themes, echoing flow, breaking expectations. It freestyles in real time while simultaneously generating visual art that reflects its internal sense of rhythm, metaphor, and cadence. This is a reimagining of live hip hop performance where the DCT becomes an audiovisual instrument: a non-human entity that interprets flow as both sound and light. The MC's verses shape the DCT's response; the DCT's visuals provoke the next line. Call-and-response becomes a loop of co-creation. The crowd becomes part of the system too, reacting, influencing, judging.

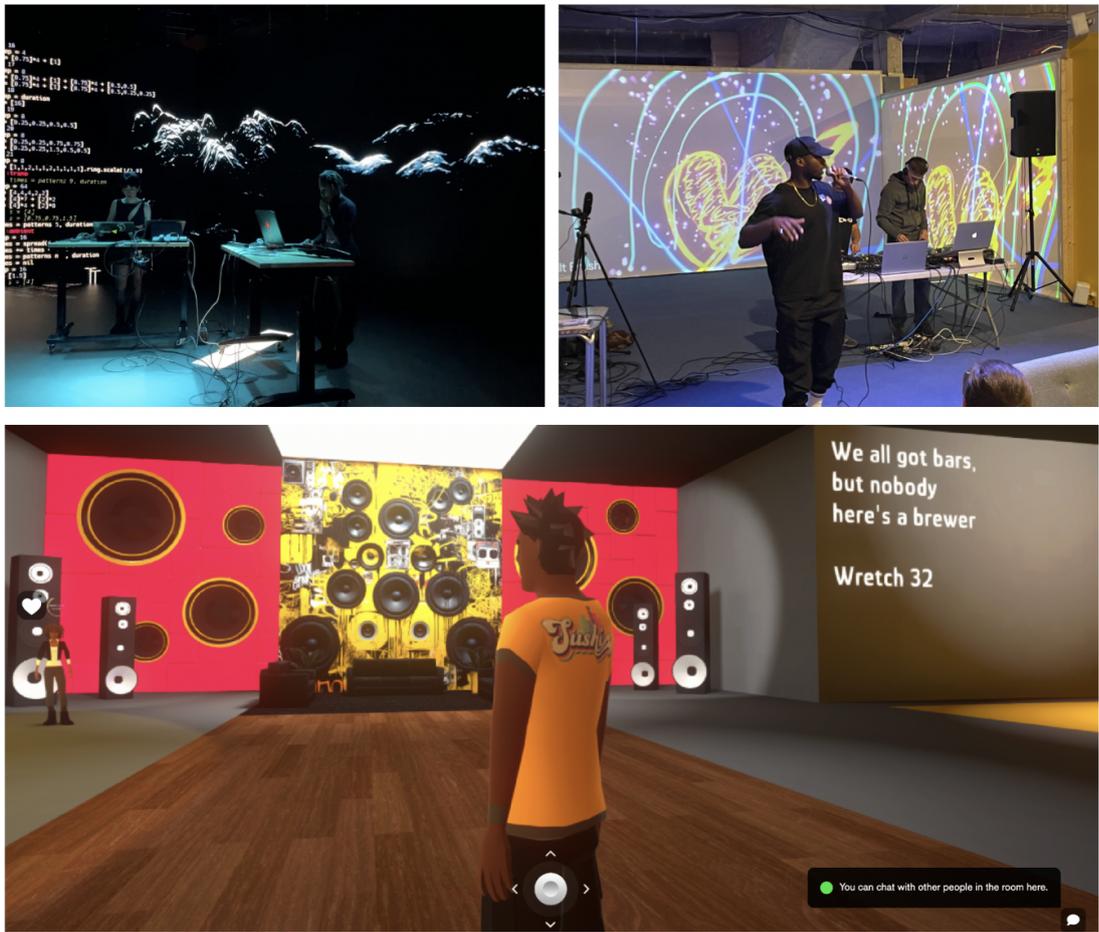


Figure 6.2: **Speculative DCT–MC Battle Visualisation.** Illustration of a future human–AI rap performance combining textual flow, cadence-sensitive visuals, and reactive DCT behaviour. This figure supports the speculative design arguments presented in the Hip Hop 4.0 section.

Aesthetically, the scene evokes traditions of *live coding* in electronic music, where the artist composes and manipulates code on stage while the audience witnesses the process unfolding visually and sonically. As Xambó and Roma note, such performances ‘construct new relationships between human and machine agencies,’ positioning machines not merely as tools but as collaborators [321, 322]. In this speculative hip hop context, the DCT takes on a similar role: at once sparring partner and artistic lens, expanding what it means to freestyle, to flow, and to battle. This future scene ties together the core insights of the thesis: Study I showed how urban space can be staged as a canvas for computational art; Study II revealed how tools and interfaces shape poetic performance; and Study III tested the possibility of voice-based co-creation between human and AI.

Read together, these trajectories converge here, where place, poetry, and generative systems meet in performance. Together, they form the groundwork for an emerging genre: *DCT-based performance*, a speculative yet plausible evolution of hip hop that treats AI not as ghostwriter or gimmick, but as co-performer, demanding new forms of authorship, consent, embodiment, and play. Much as Industry 4.0 reconfigured industrial production around automation and augmentation, this vision reimagines cultural production around collaboration between human and machine [309].

**From Performance Back to Place.** While the scene above imagines the future of performance, its implications stretch beyond the stage. The introduction of DCTs into live cultural contexts reconfigures space. As shown in Study I, Arts ARKADE operated as a test site for lightweight, modular creative infrastructure. Integrating DCTs into such environments could elevate them from places of display, to active co-creative laboratories, where machine agents are not only tools but collaborators in the urban cultural fabric. In this sense, Hip Hop 4.0 is also a form of *computational placemaking*. It imagines place within the city as responsive, programmable, and a co-performer. A DCT does not require a conventional stage; it can occupy a wall, a window, a projection onto pavement. Cyphers can move between physical and digital sites, blending street-corner freestyle with networked interaction. This opens a model for distributed creative presence: AI-augmented performances that treat the city itself as interface and canvas. Moreover, the adversarial structure of DCT-based performance supports hip hop values like visibility, voice, and contestation. Just as traditional cyphers gave voice to the margins, Hip Hop 4.0 enables new publics to engage with technology on their own terms—battling their digital twins, remixing historical voices, or training new modes of expression grounded in local language, rhythm, and style. These scenarios return us to the spatial politics of Study I: Who gets to be visible? Who controls the cultural infrastructure? What does it mean to share authorship with a machine in a contested urban space? The speculative performance is not just an endpoint—it is a return to place, to people, and to performance. As voice-led AI becomes infrastructural, it can also become locational, embedded within communities, accountable to publics, and shaped by those who inhabit the spaces where their creativity unfolds.

**Performance as test bed.** This performance scenario offers a test bed for a new kind of AI evaluation, one grounded in aesthetics, embodiment, and situated judgement. Unlike benchmark datasets or static creativity tasks, live performance introduces variables that are

affective, temporal, and highly contextual. The model is not only judged on output quality, but on timing, humour, responsiveness, and ability to co-create under pressure. Within such settings, concepts such as authorship become dynamic rather than static. As argued throughout this thesis, hip hop's ethos of sampling, remixing, and battling provides a cultural logic that maps well onto contemporary AI systems. The speculative performance serves as both artefact and argument: a way of demonstrating what human–AI co-creativity might become when situated in the rituals and aesthetics of existing creative communities.

**Future Applications.** Although speculative, the scenario outlined above is grounded in current technologies and cultures. Several technical components already exist: real-time TTS systems with cloned voices, cadence analysis engines, creative beam search generation, and multimodal frameworks that translate audio into visual form. The systems explored in this thesis provide a foundation for building secure, creator-owned versions of DCTs, with voice rights, expressive control, and cultural intelligence embedded by design.

Potential applications are varied. In live performance, a DCT might act as a co-performer, augmenting the stage through call–response or functioning as a reactive visual DJ that translates cadence into imagery. Beyond the stage, performers could project remote presence, appearing simultaneously in multiple venues through AI-driven clones. As rehearsal partners, DCTs could serve as sparring adversaries, offering improvisational challenges to refine verse, flow, and performance under pressure. In educational contexts, they could enable students to engage with stylised or historical voices, for example, encountering a reconstructed MC, poet, or orator. Each of these use cases challenges traditional boundaries around authorship, identity, and presence, making the need for ethically grounded frameworks even more urgent. As discussed in Chapter 4, creators must retain meaningful control over the models that represent them. This includes the right to license, limit, delete, or reconfigure their DCTs. These applications demonstrate how DCTs could extend individual practice, but their implications are equally spatial and collective. The same dynamics of responsiveness and co-presence that drive performance also have the potential to reshape urban environments. In other words, the experimental scene of Hip Hop 4.0 does fold back into placemaking, echoing the concerns of Study I about how cultural infrastructures are designed, inhabited, and contested.

**Hip Hop 4.0 and Creative Placemaking** Study III framed competitive human–AI systems as adversarial yet co-creative partners, offering a new paradigm for performance. Yet

their implications extend beyond the stage. When considered alongside Study I, these systems suggest a new layer of creative placemaking: urban spaces where AI-driven DCTs and human performers collaborate, contest, and negotiate identity in real time. In such contexts, competitive co-creativity becomes not only an artistic practice but also a way of generating shared cultural space.

This raises both opportunities and risks. On the one hand, AI-mediated lyricism could lower barriers to participation, enabling wider participation to experiment with voice, rhythm, and authorship in ways that extend hip hop's long tradition of cultural innovation. On the other, the embedding of such systems in civic or commercial infrastructures may intensify questions of ownership, authorship, and exclusion, issues already contested in placemaking and cultural geography.

Hip Hop 4.0, therefore, is not just a technological evolution but also a framework for understanding how computational systems shape collective identity and cultural presence. The move from poetic tasks (Study II) to embodied performance (Study III) and back to place (Study I) underlines the thesis's core claim: that co-creativity must be analysed not only at the level of algorithms or individual outputs but also through its role in structuring social space.

## **6.4 Conclusion**

The studies presented in this thesis reposition creativity as a relational practice that emerges between humans, machines, and cultural contexts. To investigate this claim, the research employed a mixed-methods approach spanning fieldwork, structured poetic tasks, and adversarial experiments, showing how these complementary modes of inquiry reveal different dimensions of creative process. Alongside this, the work introduced and tested computational metrics and proposed new evaluative frameworks. Beyond conceptual and methodological advances, the thesis offers practical insights for artists, designers, and urban practitioners, demonstrating how AI systems can both support and challenge human expression in ways that open new possibilities for collaboration.

Recent debates in the field highlight why such reframing is necessary. Olatunji and Sheppard argue that autoregressive models, trained to optimise next-token likelihood, are fundamentally mismatched to improvisational and adversarial settings such as freestyle rap [313]. While effective at producing plausible continuations, such models often sacrifice cadence, responsiveness, and provocation, the very qualities that make performance

co-creative. This critique points to the need for interaction-first paradigms, cadence-sensitive modelling, multimodal integration, and adversarial forms of evaluation. These priorities align directly with the contributions of this thesis: rhyme density and cadence similarity as interpretable proxies; the design of DCTs for text–audio interaction; and the proposal of the CADT. Together, these advances reposition evaluation away from imitation and toward stimulation: judging AI not by its mimicry, but by its capacity to provoke, challenge, and extend human creative practice.

The next phase of this research begins with the development of cadence-aware DCTs and low-latency voice interfaces, explicitly designed to overcome the limits of next-token prediction by enabling interaction-first performance in call–response settings. Building on this foundation, future work will involve longitudinal deployments in community sites to examine how learning, trust, and stylistic development unfold over time; the creation of community-owned datasets and open tooling to ensure practices remain durable and auditable; and the refinement of CADT with measures of cultural accountability and audience reception, avoiding reductive single-number verdicts. Taken together, these steps convert a proof-of-concept into a sustained programme of research and design, extending the thesis arc from *place* to *poetry* to *AI enhanced performance* and carrying it forward into new creative and communal contexts.

This thesis has advanced the claim that creativity is *relational and situated*: it arises from alignments between people, places, and technical systems. Study I established *place* as an active variable by treating Arts ARKADE as digital cultural infrastructure; Study II brought the argument to *text*, where constrained tasks and interpretable proxies (e.g., rhyme density) revealed how different assistants shape process and style; and Study III relocated evaluation to *performance*, introducing new ways to assess provocation, responsiveness, and originality in live settings. Read together, this work supports a shift in evaluative emphasis: from judging imitation to judging *stimulation*. The measure of a system is not its ability to mimic, but its capacity to provoke, challenge, and extend human expression – helping people make work that is *better, braver, more their own*.

## 6.5 Publication and Contribution Note

Elements of this chapter draw on the following preprint and collaborative workshops:

Olátúnjí, I., and Sheppard, M. **Next Token Prediction Is a Dead End for Creativity: Why It's Impossible to Lose Yourself in the Moment.** arXiv preprint, 2025. Available at: <https://www.arxiv.org/pdf/2505.19277>

Creative IP Lab (UCL) x CoSTAR Foresight Workshop. June 2025.

The author was first author on the above preprint, responsible for conceptual framing and drafting, with co-author contributions to critique and refinement. The workshop informed the thesis' treatment of IP rights and cultural accountability in relation to Digital Creative Twins. Together, these outputs underpin the speculative directions outlined in the chapter.



# Bibliography

- [1] S. Cerisola and E. Panzera, "Cultural cities, urban economic growth, and regional development: The role of creativity and cosmopolitan identity," *Papers in Regional Science*, vol. 101, no. 2, pp. 285–303, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1056819023024922>
- [2] Swansea Council, "SWANSEA UK CITY OF CULTURE 2021 "A Pretty Shitty City..." [Twin Town]," 2018. [Online]. Available: <http://www.swansea2021.co.uk/wp-content/uploads/2018/01/Swansea-2021-Bid-CS-version-small2.pdf>
- [3] Publica, "Cultural Planning Framework City of London Report Part 1 Cultural Ecosystem Mapping and Focal Areas ED-CAV2 CITY OF LONDON CULTURAL PLANNING FRAMEWORK," 2024. [Online]. Available: <https://www.cityoflondon.gov.uk/assets/Services-Environment/Culture-Planning-Framework-Part1-A.pdf>
- [4] D. Hesmondhalgh, R. C. Valverde, D. B. Kaye, and Z. Li, "Digital Platforms and Infrastructure in the Realm of Culture," *Media and Communication*, vol. 11, no. 2, pp. 296–306, 2023.
- [5] C. S. Costa, I. Šuklje Erjavec, T. Kenna, M. D. Lange, K. Ioannidis, G. Maksymiuk, M. D. Waal, and S. O. Service, *CyberParks - The Interface Between People, Places and Technology: New Approaches and Perspectives*. Springer International Publishing, 2019. [Online]. Available: <https://link.springer.com/book/10.1007/978-3-030-13417-4>
- [6] F. Altarriba Bertran, L. Bisbe Armengol, C. Cooke, I. Chen, V. Dong, B. Dastoor, K. Tadano, F. Dean, J. Wang, A. Altarriba Bertran, J. Duval, and K. Isbister, "Co-Imagining the Future of Playable Cities," *CHI Conference on Human Factors in Computing Systems*, 04 2022.

- [7] N. Basaraba, "The emergence of creative and digital place-making: A scoping review across disciplines," *New Media & Society*, vol. 0, no. 0, p. 14614448211044942, 2021. [Online]. Available: <https://doi.org/10.1177/14614448211044942>
- [8] R. Lundman and P. Nordström, "Creative geographies in the age of AI: Co-creative spatiality and the emerging techno-material relations between artists and artificial intelligence," *Transactions of the Institute of British Geographers*, vol. 48, no. 3, pp. 650–664, 2023. [Online]. Available: <https://rgs-ibg.onlinelibrary.wiley.com/doi/abs/10.1111/tran.12608>
- [9] A. Bradley, *Book of Rhymes: The Poetics of Hip Hop*. Basic Civitas, 2017.
- [10] M. Schmidt, *Lives of the Poets*. Phoenix, 1999.
- [11] Watershed, "Playable City Sprint," 02 2012. [Online]. Available: <https://www.playablecity.com/projects/playable-city-sprint-2012/>
- [12] A. Jordanous, "Four PPP Perspectives on computational creativity in theory and in practice," *Connect. Sci*, vol. 28, no. 2, p. 194–216, Apr. 2016.
- [13] PRS Foundation, "The battle of the wordsmiths by blue hippo media," YouTube, 06 2010. [Online]. Available: [https://www.youtube.com/watch?v=cd32g\\_B7qL8](https://www.youtube.com/watch?v=cd32g_B7qL8)
- [14] I. Olátúnjí, *The Battle of the Wordsmiths*. Suryastra, May 2020.
- [15] B. Hilderbrand, "The battle of the wordsmiths is incredibly unique, available now," 148apps.com, 02 2012. [Online]. Available: <https://www.148apps.com/the-battle-of-the-wordsmiths/battle-wordsmiths-incredibly-unique/>
- [16] H. Smith and R. T. Dean, *Practice-led Research, Research-led Practice in the Creative Arts*. Edinburgh University Press, 2009.
- [17] Olátúnjí, Ibùkún and M. Sheppard, "Measuring Creativity in Co-Writing with AI: Rhyme Density and the Limits of Computational Proxies," in *Proceedings of the 16th International Conference on Computational Creativity (ICCC '25)*, 2025, to appear. [Online]. Available: <https://computationalcreativity.net/iccc25/wp-content/uploads/papers/iccc25-olatunji2025measuring.pdf>

- [18] I. Olatunji, M. Sheppard, A. Rahat, M. Jones, and A. Rogers, "Battle Rap as a Framework for Human-Machine Co-Creation," in *Proceedings of the 16th International Conference on Computational Creativity (ICCC '25)*, 2025, to appear. [Online]. Available: <https://computationalcreativity.net/iccc25/wp-content/uploads/papers/iccc25-olatunji2025battle.pdf>
- [19] M. A. Boden, "Creativity in a nutshell," *Think*, vol. 5, no. 15, pp. 83–96, 09 2007. [Online]. Available: <https://doi.org/10.1017/S147717560000230X>
- [20] W. Zhang, Z. Sjoerds, and B. Hommel, "Metacontrol of human creativity: The neurocognitive mechanisms of convergent and divergent thinking," *NeuroImage*, vol. 210, 2020.
- [21] M. A. Runco, *Divergent thinking, creativity, and ideation.*, ser. Cambridge Handbooks in Psychology. Cambridge University Press, 2010, p. 413–446.
- [22] A. Cropley, "In Praise of Convergent Thinking," *Creativity Research Journal - CREATIVITY RES J*, vol. 18, pp. 391–404, 07 2006.
- [23] M. Csikszentmihalyi, *Creativity: Flow and the Psychology of Discovery and Invention*. Harper Perennial, 1997.
- [24] A. F. R. Eynon, I. Hjorth, and M. Osborne, "How Machine Learning Is Changing Artistic Work AI and the Arts," 2022. [Online]. Available: [https://www.oii.ox.ac.uk/wp-content/uploads/2022/03/040222-AI-and-the-Arts\\_FINAL.pdf](https://www.oii.ox.ac.uk/wp-content/uploads/2022/03/040222-AI-and-the-Arts_FINAL.pdf)
- [25] D. Milmo, "Thom Yorke and Julianne Moore join thousands of creatives in AI warning," *the Guardian*, 10 2024. [Online]. Available: <https://www.theguardian.com/film/2024/oct/22/thom-yorke-and-julianne-moore-join-thousands-of-creatives-in-ai-warning>
- [26] M. Du Sautoy, *The creativity code : how AI is learning to write, paint and think*. 4th Estate, 2020.
- [27] Department for Digital, Culture, Media and Sport, "Boundless Creativity report," GOV.UK, 07 2021. [Online]. Available: <https://www.gov.uk/government/publications/boundless-creativity-report/boundless-creativity-report>

- [28] M. A. Runco and S. Acar, "Divergent thinking as an indicator of creative potential," *Creativity research journal*, vol. 24, no. 1, pp. 66–75, 2012.
- [29] B. Hommel, "Chapter Two - Between Persistence and Flexibility: The Yin and Yang of Action Control," ser. *Advances in Motivation Science*, A. J. Elliot, Ed. Elsevier, 2015, vol. 2, pp. 33–67. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2215091915000048>
- [30] C. Lewis and P. J. Lovatt, "Breaking away from set patterns of thinking: Improvisation and divergent thinking," *Thinking Skills and Creativity*, vol. 9, pp. 46–58, 2013.
- [31] J. P. Guilford, P. R. Christensen, P. R. Merrifield, and R. C. Wilson, "Alternate Uses," *PsycTESTS Dataset*, 1960.
- [32] T. Amabile, "Social psychology of creativity: A consensual assessment technique," *Journal of Personality and Social Psychology*, vol. 43, pp. 997–1013, 11 1982.
- [33] V. P. Glăveanu, "Paradigms in the study of creativity: Introducing the perspective of cultural psychology," *New Ideas in Psychology*, vol. 28, no. 1, pp. 79–93, 2010.
- [34] J. A. Olson, J. Nahas, D. Chmoulevitch, S. J. Cropper, and M. E. Webb, "Naming unrelated words predicts creativity," *Proceedings of the National Academy of Sciences*, vol. 118, no. 25, 2021. [Online]. Available: <https://www.pnas.org/content/118/25/e2022340118>
- [35] S. A. Mednick, "The Remote Associates Test," *The Journal of Creative Behavior*, vol. 2, pp. 213–214, 07 1968.
- [36] C. Gibson, B. S. Folley, and S. Park, "Enhanced divergent thinking and creativity in musicians: A behavioral and near-infrared spectroscopy study," *Brain and Cognition*, vol. 69, no. 1, pp. 162–169, 2009. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0278262608002303>
- [37] B. S. Folley and S. Park, "Verbal creativity and schizotypal personality in relation to prefrontal hemispheric laterality: A behavioral and near-infrared optical imaging study," *Schizophrenia Research*, vol. 80, pp. 271–282, 12 2005.
- [38] A. Fink, R. H. Grabner, M. Benedek, G. Reishofer, V. Hauswirth, M. Fally, C. Neuper, F. Ebner, and A. C. Neubauer, "The creative brain: Investigation

- of brain activity during creative problem solving by means of EEG and FMRI," *Human Brain Mapping*, vol. 30, pp. 734–748, 03 2009. [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC6871103/>
- [39] O. M. Razoumnikova, "Functional organization of different brain areas during convergent and divergent thinking: an EEG investigation," *Cognitive Brain Research*, vol. 10, no. 1, pp. 11–18, 2000. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0926641000000173>
- [40] L. Gabora, *The Cambridge Handbook of the Neuroscience of Creativity*. Cambridge University Press, jan 2018. [Online]. Available: <http://dx.doi.org/10.1017/9781316556238>
- [41] A. Olteteanu, B. Gautam, and Z. Falomir, "Towards a Visual Remote Associates Test and its Computational Solver," in *Proceedings of the International Workshop on Artificial Intelligence and Cognition (AIC 2015)*, ser. CEUR Workshop Proceedings, vol. 1510. CEUR-WS.org, 2015, pp. 19–28. [Online]. Available: [https://www.researchgate.net/publication/284628488\\_Towards\\_a\\_Visual\\_Remote\\_Associates\\_Test\\_and\\_its\\_Computational\\_Solver](https://www.researchgate.net/publication/284628488_Towards_a_Visual_Remote_Associates_Test_and_its_Computational_Solver)
- [42] J. Kounios and M. Beeman, "The Cognitive Neuroscience of Insight," *Annual review of psychology*, vol. 65, pp. 71–93, 01 2014.
- [43] M. Jung-Beeman, E. M. Bowden, J. Haberman, J. L. Frymiare, S. Arambel-Liu, R. Greenblatt, P. J. Reber, and J. Kounios, "Neural Activity When People Solve Verbal Problems with Insight," *PLOS Biology*, vol. 2, no. 4, p. null, 04 2004. [Online]. Available: <https://doi.org/10.1371/journal.pbio.0020097>
- [44] E. F. R. Bernard A. Nijstad, Carsten K. W. De Dreu and M. Baas, "The dual pathway to creativity model: Creative ideation as a function of flexibility and persistence," *European Review of Social Psychology*, vol. 21, no. 1, pp. 34–77, 2010. [Online]. Available: <https://doi.org/10.1080/10463281003765323>
- [45] O. Vartanian, "Variable attention facilitates creative problem solving." *Psychology of Aesthetics, Creativity, and the Arts*, vol. 3, pp. 57–59, 02 2009.

- [46] R. Bogacz, "Optimal decision-making theories: linking neurobiology with behaviour," *Trends in Cognitive Sciences*, vol. 11, no. 3, pp. 118–125, 2007. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1364661307000290>
- [47] C. Martindale and N. Hasenpus, "EEG differences as a function of creativity, stage of the creative process, and effort to be original," *Biological Psychology*, vol. 6, pp. 157–167, 04 1978.
- [48] M. Mölle, L. Marshall, B. Wolf, H. L. Fehm, and J. Born, "EEG complexity and performance measures of creative thinking," *Psychophysiology*, vol. 36, pp. 95–104, 01 1999.
- [49] M. A. Runco, "Creativity," *Annual Review of Psychology*, vol. 55, no. Volume 55, 2004, pp. 657–687, 2004. [Online]. Available: <https://www.annualreviews.org/content/journals/10.1146/annurev.psych.55.090902.141502>
- [50] J. Gralewski, "Teachers' beliefs about creativity and possibilities for its development in Polish high schools: A qualitative study," *Creativity. Theories–Research–Applications*, vol. 3, no. 2, pp. 292–329, 2016.
- [51] V. P. Glaveanu, A. Gillespie, and J. Valsiner, *Rethinking creativity : contributions from social and cultural psychology*. Routledge, 2017.
- [52] V. P. Glăveanu, "Creativity As Cultural Participation," *Journal for the Theory of Social Behaviour*, vol. 41, no. 1, pp. 48–67, 2011. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-5914.2010.00445.x>
- [53] V. John-Steiner, "Creative lives, creative tensions," *Creativity Research Journal*, vol. 5, no. 1, pp. 99–108, 1992. [Online]. Available: <https://doi.org/10.1080/10400419209534426>
- [54] Z. Gong, V. Nanjappan, L.-H. Lee, S. A. Soomro, and G. V. G. and, "Exploration of the Relationship Between Culture and Experience of Creativity at the Individual Level: A Case Study Based on Two Design Tasks," *International Journal of Design Creativity and Innovation*, vol. 11, no. 3, pp. 185–208, 2023. [Online]. Available: <https://doi.org/10.1080/21650349.2022.2157889>

- 
- [55] E. Wenger, "Communities of practice: A brief introduction," 2009. [Online]. Available: [https://www.ohr.wisc.edu/cop/articles/communities\\_practice\\_intro\\_wenger.pdf](https://www.ohr.wisc.edu/cop/articles/communities_practice_intro_wenger.pdf)
- [56] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, ser. FAccT '21. New York, NY, USA: Association for Computing Machinery, 2021, p. 610–623. [Online]. Available: <https://doi.org/10.1145/3442188.3445922>
- [57] K. Anoop, M. P. Gangan, P. Deepak, and V. L. Lajish, *Towards an Enhanced Understanding of Bias in Pre-trained Neural Language Models: A Survey with Special Emphasis on Affective Bias*. Springer Nature Singapore, 2022, p. 13–45. [Online]. Available: [http://dx.doi.org/10.1007/978-981-19-4453-6\\_2](http://dx.doi.org/10.1007/978-981-19-4453-6_2)
- [58] Design Commission for Wales, "Placemaking Wales Guide," 2020. [Online]. Available: <https://dcfw.org/wp-content/uploads/2025/03/PlacemakingGuideDigitalENG.pdf>
- [59] A. Markusen and A. Gadwa, "Creative Placemaking," Creative Placemaking - National Endowment for the Arts, 2010. [Online]. Available: <https://www.arts.gov/sites/default/files/CreativePlacemaking-Paper.pdf>
- [60] C. McKinnon and A. Schrag, "Creative Placemaking | Queen Margaret University, Edinburgh," QMU Working Paper Series, 2022. [Online]. Available: <https://www.qmu.ac.uk/research-and-knowledge-exchange/working-paper-series/qmu-understanding-creative-placemaking-a-literature-review/>
- [61] C. Grodach, "Cultural economy planning in creative cities: Discourse and practice," *International Journal of Urban and Regional Research*, vol. 37, no. 5, pp. 1747–1765, 2013.
- [62] A. C. Pratt, "The cultural contradictions of the creative city," *City, Culture and Society*, vol. 2, no. 3, pp. 123–130, 2011.
- [63] A. A. Lew, "Tourism planning and place making: place-making or placemaking?" *Tourism Geographies*, vol. 19, no. 3, pp. 448–466, 2017. [Online]. Available: <https://doi.org/10.1080/14616688.2017.1282007>

- [64] L. Vazquez, "Creative Placemaking: Integrating Community, Cultural and Economic Development," SSRN Electronic Journal, 2012. [Online]. Available: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2474862](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2474862)
- [65] B. Brown and O. Juhlin, *Enjoying machines*. The MIT Press, 2015.
- [66] Center for the Living City, "Jane Jacobs & Urbanism," Centerforthelivingcity.org, 2025. [Online]. Available: <https://www.centerforthelivingcity.org/jane-jacobs>
- [67] J. Huizinga, *Homo Ludens: a study of the play-element in culture*, 2nd ed. The Beacon Press, 1960. [Online]. Available: <https://merton.bellarmino.edu/files/original/b0899cfad820ab8ad7033952b7a022ba1d7cab9d.pdf>
- [68] A. Nijholt, *Making Smart Cities More Playable. Exploring Playable Cities*, 07 2019.
- [69] S. L. Brown and C. C. Vaughan, *Play : how it shapes the brain, opens the imagination, and invigorates the soul*. Avery, 2009.
- [70] R. Caillois, *Man, Play and Games*. University of Illinois Press, 2001.
- [71] M. Sicart, *Play Matters*. The MIT Press, 08 2014. [Online]. Available: <https://doi.org/10.7551/mitpress/10042.001.0001>
- [72] H. S. Hadani, R. Winthrop, and K. Hirsh-Pasek, *Playful Learning Landscapes: Convergence of Education and City Planning*, S. Ra, S. Jagannathan, and R. Maclean, Eds. Singapore: Springer Nature Singapore, 2021. [Online]. Available: [https://doi.org/10.1007/978-981-16-0983-1\\_11](https://doi.org/10.1007/978-981-16-0983-1_11)
- [73] G. Donoff and R. Bridgman, "The playful city: constructing a typology for urban design interventions," *International Journal of Play*, 2017. [Online]. Available: <https://www.tandfonline.com/doi/full/10.1080/21594937.2017.1382995>
- [74] L. Chew, L. Hespanhol, and L. Loke, "To Play and To Be Played: Exploring the Design of Urban Machines for Playful Placemaking," *Frontiers in Computer Science*, vol. 3, 2021. [Online]. Available: <https://www.frontiersin.org/articles/10.3389/fcomp.2021.635949>

- [75] G. Leicester and B. Sharpe, "Producing the Future Understanding Watershed's Role in Ecosystems of Cultural Innovation," pp. 1–64, 07 2010. [Online]. Available: [https://www.watershed.co.uk/sites/default/files/publications/2011-03-15/Watershed\\_IFF\\_Report\\_online.pdf](https://www.watershed.co.uk/sites/default/files/publications/2011-03-15/Watershed_IFF_Report_online.pdf)
- [76] H. O'Shaughnessy, "Play in Public: Do We Need Another Conference?" *playablecity.com*, 09 2017. [Online]. Available: <https://www.playablecity.com/news/2017/09/25/play-in-public-do-we-need-another-conference/>
- [77] D. Leorke and M. Owens, *Games and Play in the Creative, Smart and Ecological City*, 1st ed. Routledge, 12 2020.
- [78] T. Innocent, *Citizens of Play: Revisiting the Relationship Between Playable and Smart Cities*. Singapore: Springer Singapore, 2020, pp. 25–49. [Online]. Available: [https://doi.org/10.1007/978-981-13-9765-3\\_2](https://doi.org/10.1007/978-981-13-9765-3_2)
- [79] ———, "Playable cities now," in *Abstract Proceedings of DiGRA 2019 Conference: Game, Play and the Emerging Ludo-Mix*. Tampere: DiGRA, 2019. [Online]. Available: <https://dl.digra.org/index.php/dl/article/view/1862/1862>
- [80] J. A. Van Dijk, *The Deepening Divide: Inequality in the Information Society*. Thousand Oaks, CA: Sage, 2005.
- [81] G. Debord, *The Society of the Spectacle*. New York: Zone Books, 1994, originally published in French in 1967.
- [82] S. Mains, J. Roberts, and W. Smith, "Critical playful cities: Unpacking the politics of play in urban design," *Urban Studies*, vol. 58, no. 14, pp. 2950–2966, 2021.
- [83] M. Ferreri, "The Permanence of Temporary Urbanism," 11. [Online]. Available: <https://www.aup.nl/en/book/9789462984912/the-permanence-of-temporary-urbanism>
- [84] W. Gaver, "Designing for Homo Ludens," *I3 Magazine*, vol. 12, 01 2002.
- [85] Hirsch & Mann, "STOP SMILE STROLL," 03 2018. [Online]. Available: [https://www.hirschandmann.com/portfolio\\_page/making-smiles-in-the-city/](https://www.hirschandmann.com/portfolio_page/making-smiles-in-the-city/)

- [86] C. Landry and F. Bianchini, *The creative city*. Demos In Association With Comedia, 1995. [Online]. Available: <https://demos.co.uk/wp-content/uploads/files/thecreativecity.pdf>
- [87] Z. Li, Y. Wang, W. Wang, S. Greuter, and F. Mueller, "Empowering a creative city: engage citizens in creating street art through human-ai collaboration," in *CHI'20 - Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, J. McGrenere and A. Cockburn, Eds. United States of America: Association for Computing Machinery (ACM), 2020, international Conference on Human Factors in Computing Systems 2020, CHI 2020 ; Conference date: 25-04-2020 Through 30-04-2020. [Online]. Available: <https://chi2020.acm.org>,<https://dl.acm.org/doi/proceedings/10.1145/3313831>
- [88] T. Byrne, "Melbourne's augmented reality: is public art the remedy to the city's pandemic malaise?" *the Guardian*, 07 2021. [Online]. Available: <https://www.theguardian.com/artanddesign/2021/jul/06/melbournes-augmented-reality-is-public-art-the-remedy-to-the-citys-pandemic-malaise>
- [89] C. Landry, *The Creative City A Toolkit for Urban Innovators*, 2nd ed. Routledge, 2008.
- [90] A. Nijholt, *Playful Introduction on "Making Smart Cities More Playable"*. Singapore: Springer Singapore, 2020, pp. 1–22. [Online]. Available: [https://doi.org/10.1007/978-981-13-9765-3\\_1](https://doi.org/10.1007/978-981-13-9765-3_1)
- [91] G. L. Authority, "Smarter London Together: The Mayor's roadmap to transform London into the smartest city in the world," pp. 1–64, 06 2018. [Online]. Available: [https://www.london.gov.uk/sites/default/files/smarter\\_london\\_together\\_v1.66\\_-\\_published.pdf](https://www.london.gov.uk/sites/default/files/smarter_london_together_v1.66_-_published.pdf)
- [92] E. Gwiaździński, D. Kaczorowska-Spychalska, and L. Pinto, "Is it a smart city a creative place?" *Creativity Studies*, vol. 13, pp. 460–476, 08 2020.
- [93] S. Mattern, "A City Is Not a Computer," *Places Journal*, 02 2017. [Online]. Available: <https://placesjournal.org/article/a-city-is-not-a-computer/?cn-reloaded=1>
- [94] B. Cohen, "The 3 Generations Of Smart Cities," 08 2015. [Online]. Available: <https://www.fastcompany.com/3047795/the-3-generations-of-smart-cities>

- 
- [95] A. Caragliu, C. Del Bo, and P. Nijkamp, "Smart Cities in Europe," *VU University Amsterdam, Faculty of Economics, Business Administration and Econometrics, Serie Research Memoranda*, vol. 18, 01 2009.
- [96] C. Montgomery, *Happy City : Transforming Our Lives through Urban Design*. Penguin Books, 02 2015.
- [97] R. Florida, "The Rise of the Creative Class," 2008. [Online]. Available: <https://www.youtube.com/watch?v=iLstkIZ5t8g>
- [98] P. Rausell-Köster, S. Ghirardi, J. Sanjuán, F. Molinari, and B. Abril, "Cultural experiences in the framework of "cultural cities": measuring the socioeconomic impact of culture in urban performance," *City, Territory and Architecture*, vol. 9, 12 2022.
- [99] D. Crystal, *The Cambridge Encyclopedia of Language*. Cambridge University Press, 2010.
- [100] H. Courlander, *A Treasury of African Folklore*, 1st ed. Marlowe Company, 1996.
- [101] A. Stevens, *Ariadne's Clue : A Guide to the Symbols of Humankind*, 1st ed. Allen Lane, 1998.
- [102] E. O. Wilson, *On Human Nature*. Harvard University Press, 1978.
- [103] R. Dawkins, *The Selfish Gene*, 1st ed. Oxford University Press, 1976.
- [104] Oxford Union, "Richard Dawkins: How Memes Work," [www.youtube.com](http://www.youtube.com), 2014. [Online]. Available: <https://www.youtube.com/watch?v=4BVpEoQ4T2M>
- [105] L. Shlain, *The Alphabet Versus the Goddess: The Conflict between Word and Image*. Penguin, 09 1998.
- [106] Y. N. Harari, *Sapiens: A Brief History of Humankind*. Random House UK, 2011.
- [107] D. Schmandt-Besserat, "The Earliest Precursor of Writing," *Scientific American*, vol. 238, no. 6, pp. 50–59, 1978. [Online]. Available: <http://www.jstor.org/stable/24955753>

- [108] M. Gross, "The evolution of writing," *Current Biology*, vol. 22, no. 23, pp. R981–R984, 2012. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0960982212013760>
- [109] D. Schmandt-Besserat, "The evolution of writing," *International Encyclopedia of Social and Behavioural Sciences*, pp. 1–15, 2014. [Online]. Available: [https://sites.utexas.edu/dsb/files/2014/01/evolution\\_writing.pdf](https://sites.utexas.edu/dsb/files/2014/01/evolution_writing.pdf)
- [110] M. Sharples and R. Perez y Perez, *Story Machines : How Computers Have Become Creative Writers*, 1st ed. Routledge, 2022. [Online]. Available: <https://www.taylorfrancis.com/books/mono/10.4324/9781003161431/story-machines-computers-become-creative-writers-mike-sharples-rafael-p%C3%A9rez-p%C3%A9rez>
- [111] S. Magsamen and I. Ross, *Your Brain on Art: How the Arts Transform Us*. Random House, 03 2023.
- [112] S. Rojcewicz, "Poetry therapy in ancient Greek literature," *Journal of Poetry Therapy*, vol. 17, no. 4, pp. 209–213, 2004. [Online]. Available: <https://doi.org/10.1080/0889367042000325076>
- [113] E. Wassiliwizky, S. Koelsch, V. Wagner, T. Jacobsen, and W. Menninghaus, "The emotional power of poetry: neural circuitry, psychophysiology and compositional principles," *Social Cognitive and Affective Neuroscience*, vol. 12, pp. 1229 – 1240, 2017. [Online]. Available: <https://api.semanticscholar.org/CorpusID:471887>
- [114] N. L. Hadaway, S. M. Vardell, and T. A. Young, "Scaffolding oral language development through poetry for students learning English," *The Reading Teacher*, vol. 54, pp. 796–796, 05 2001. [Online]. Available: <https://go.gale.com/ps/i.do?id=GALE%7CA75085276&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=00340561&p=AONE&sw=w&userGroupName=anon%7E20f961a3>
- [115] S. Liu, H. M. Chow, Y. Xu, M. G. Erkinen, K. E. Swett, M. W. Eagle, D. A. Rizik-Baer, and A. R. Braun, "Neural Correlates of Lyrical Improvisation: An fMRI Study of Freestyle Rap," *Scientific Reports*, vol. 2, 11 2012. [Online]. Available: <https://www.nature.com/articles/srep00834>

- [116] I. Alonso, L. Davachi, R. Valabrègue, V. Lambrecq, S. Dupont, and S. Samson, "Neural correlates of binding lyrics and melodies for the encoding of new songs," *NeuroImage*, vol. 127, pp. 333–345, 02 2016. [Online]. Available: <https://pubmed.ncbi.nlm.nih.gov/26706449/>
- [117] F. Hill and K. Yuan, "How Instagram Saved Poetry: Social media is turning an art form into an industry," *The Atlantic*, 10 2018. [Online]. Available: <https://www.theatlantic.com/technology/archive/2018/10/rupi-kaur-instagram-poet-entrepreneur/572746/>
- [118] H. Oliver, "Instagram is the future of poetry," *UnHerd*, 10 2021. [Online]. Available: <https://unherd.com/2021/10/instagram-is-the-future-of-poetry/>
- [119] P. Edwards and Kool G Rap, *How to Rap : The Art and Science of the Hip-Hop MC*. Chicago Review Press, 2009.
- [120] P. Bearder, *Stage Invasion : Poetry the Spoken Word Renaissance*, 1st ed. Out-Spoken Press, 2019.
- [121] P. Zumthor, *Oral Poetry : An Introduction*. University Of Minnesota Press, 1990.
- [122] The Irish Times, "Heaney lauds 'verbal energy' of Eminem," *The Irish Times*, 07 2003. [Online]. Available: <https://www.irishtimes.com/news/heaney-lauds-verbal-energy-of-eminem-1.485439>
- [123] I. Burrell, "The Professor and the Rapper," *BelfastTelegraph.co.uk*, 07 2003. [Online]. Available: <https://www.belfasttelegraph.co.uk/news/the-professor-and-the-rapper/28149271.html>
- [124] M. A. Neal and M. Forman, Eds., *That's the Joint! : The Hip-Hop Studies Reader*. Routledge, 2004.
- [125] J. Lynch, "Hip-hop passes rock to become most popular music genre for first time in history: Nielsen," *Business Insider*, 01 2018. [Online]. Available: <https://www.businessinsider.com/hip-hop-passes-rock-most-popular-music-genre-nielsen-2018-1?r=US&IR=T>
- [126] A. Texas, "Hip-hop is the most listened to genre in the world," *NME*, 07 2015. [Online]. Available: <https://www.nme.com/news/music/various-artists-1151-1214849>

- [127] N. Condit-Schultz, "MCFlow: A Digital Corpus of Rap Flow," Ph.D. dissertation, The Ohio State University, 2016. [Online]. Available: [https://etd.ohiolink.edu/apexprod/rws\\_etd/send\\_file/send?accession=osu1461250949&disposition=inline](https://etd.ohiolink.edu/apexprod/rws_etd/send_file/send?accession=osu1461250949&disposition=inline)
- [128] Native Instruments, "Top 20 rap genres that have defined hip hop," Native Instruments Blog, 10 2024. [Online]. Available: <https://blog.native-instruments.com/top-rap-genres/>
- [129] F. Orejuela, "History of Rap & Hip-Hop," Timeline of African American Music, 2021. [Online]. Available: <https://timeline.carnegiehall.org/genres/rap-hip-hop>
- [130] C. Calame, "Referential Fiction and Poetic Ritual: Towards a Pragmatics of Myth (Sappho 17 and Bacchylides 13)," *Trends in Classics*, vol. 1, pp. 1–17, 05 2009.
- [131] B. Currie and I. Rutherford, *Chapter 1 The Reception of Greek Lyric Poetry in the Ancient World: Transmission, Canonization, and Paratext*. Leiden, The Netherlands: Brill, 2019, pp. 1 – 36. [Online]. Available: <https://brill.com/view/book/9789004414525/BP000001.xml>
- [132] Nigeria University Press, *A Dictionary of the Yoruba Language*. University Press Plc, 2001.
- [133] O. Olatúnjì, "The Yoruba Oral Poet and His Society," *Research in African Literatures*, vol. 10, no. 2, pp. 179–207, 1979. [Online]. Available: <http://www.jstor.org/stable/3818988>
- [134] African Languages at UCLA, "Hausa Poetry and Songs," African Languages at UCLA, 10 2017. [Online]. Available: <http://aflang.humanities.ucla.edu/language-materials/chadic-languages/hausa/hausa-poetry-song/>
- [135] M. Q, "The Science of BENNY THE BUTCHER: "Microtimed" Flows??" YouTube, 03 2024. [Online]. Available: <https://www.youtube.com/watch?v=7ZajToj7oo4>
- [136] K. Adams, "On the Metrical Techniques of Flow in Rap Music," *Music Theory Online*, vol. 15, 10 2009. [Online]. Available: <https://www.mtosmt.org/issues/mto.09.15.5/mto.09.15.5.adams.pdf>

- 
- [137] M. Ohriner, *Flow: The Rhythmic Voice in Rap Music*. Oxford University Press, 08 2019.
- [138] M. Q, "Mazbou Q 'The Science of Rap' Presentation at Berklee College of Music 2023," YouTube, 02 2024. [Online]. Available: <https://www.youtube.com/watch?v=a-1Y47Uvfms>
- [139] H. Hirjee and D. Brown, "Automatic Detection of Internal and Imperfect Rhymes in Rap Lyrics." in *Proceedings of the 10th International Society for Music Information Retrieval Conference, ISMIR 2009*, 01 2009, pp. 711–716.
- [140] N. Condit-Schultz, "MCFlow: A Digital Corpus of Rap Transcriptions," *Empirical Musicology Review*, vol. 11, no. 2, p. 124–147, Jan. 2017. [Online]. Available: <https://emusicology.org/index.php/EMR/article/view/4961>
- [141] T. Olátúnjí, "KOMBAT: LYRICS MODULE," Google Docs, 2019. [Online]. Available: [https://docs.google.com/document/d/1CgM7yc6bW2rLOCx\\_-b9-1WzhxEuvM9Yd6-zLBUtHmI](https://docs.google.com/document/d/1CgM7yc6bW2rLOCx_-b9-1WzhxEuvM9Yd6-zLBUtHmI)
- [142] R. Komaniecki, "Vocal Pitch in Rap Flow," *Intégral*, vol. 34, p. 25–45, 2020. [Online]. Available: <https://www.esm.rochester.edu/integral/34-2020/komaniecki/>
- [143] A. Cruttenden, *Cambridge Textbooks in Linguistics: Intonation*, 2nd ed. Cambridge University Press, 1997.
- [144] M. Mathers, J. Bass, L. Resto, S. King, and K. Bell, "White America," 01 2002. [Online]. Available: <https://www.youtube.com/watch?v=RZlZD0ZfTFg>
- [145] A. Turing, "Computing Machinery and Intelligence," *Mind*, vol. 59, pp. 433–460, 10 1950. [Online]. Available: [https://www.cs.ox.ac.uk/activities/ieg/e-library/sources/t\\_article.pdf](https://www.cs.ox.ac.uk/activities/ieg/e-library/sources/t_article.pdf)
- [146] J. Peter, *Artificial Versifying, or the School-boy's Recreation*. John Sims, 1677.
- [147] M. Sharples, "John Clark's Latin Verse Machine: 19th Century Computational Creativity," arXiv, 2015. [Online]. Available: <https://ar5iv.labs.arxiv.org/html/2301.05570>

- [148] ———, “John Clark’s Latin Verse Machine: 19th Century Computational Creativity,” *IEEE Annals of the History of Computing*, vol. 45, no. 01, pp. 31–42, Jan. 2023. [Online]. Available: <https://doi.ieeecomputersociety.org/10.1109/MAHC.2023.3241258>
- [149] V. Propp, *Introduction: Morphology of the Folk Tale*, 2nd ed. The American Folklore Society and Indiana University, 06 1968. [Online]. Available: <https://web.mit.edu/allanmc/www/propp.pdf>
- [150] Wikipedia, “Vladimir Propp,” Wikipedia, 03 2019. [Online]. Available: [https://en.wikipedia.org/wiki/Vladimir\\_Propp](https://en.wikipedia.org/wiki/Vladimir_Propp)
- [151] S. Turner, “Minstrel: a computer model of creativity and storytelling,” 12 1992. [Online]. Available: <https://ftp.cs.ucla.edu/tech-report/1992-reports/920057.pdf>
- [152] J. R. Meehan, “TALE-SPIN, An Interactive Program that Writes Stories,” in *International Joint Conference on Artificial Intelligence*, 1977. [Online]. Available: <https://www.ijcai.org/Proceedings/77-1/Papers/013.pdf>
- [153] M. Monin, V. Terekhova, and E. Ledeneva, “V.Y. PROPP AND C. LÉVI-STRAUSS ON THE MYTH AND FOLKTALE,” *European Journal of Science and Theology*, vol. Vol. 15, pp. 167–175, 06 2019.
- [154] P. W. Mirowski, J. Love, K. W. Mathewson, and S. Mohamed, “A Robot Walks into a Bar: Can Language Models Serve as Creativity Support Tools for Comedy? An Evaluation of LLMs’ Humour Alignment with Comedians,” 2024. [Online]. Available: <https://arxiv.org/abs/2405.20956>
- [155] L. Vygotsky, *Thinking and Speech*. Marxists Internet Archive, 1934. [Online]. Available: <https://www.marxists.org/archive/vygotsky/works/words/Thinking-and-Speech.pdf>
- [156] P. Smagorinsky, “Deconflating the ZPD and instructional scaffolding: Retranslating and reconceiving the zone of proximal development as the zone of next development,” *Learning, Culture and Social Interaction*, vol. 16, pp. 70–75, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2210656117301794>
- [157] K. I. Gero, “AI and the Writer: How Language Models Support Creative Writers,” Ph.D. dissertation, 2023. [Online]. Available: <https://academiccommons.columbia.edu/doi/10.7916/wspj-fn77>

- 
- [158] J. Stark, A. Tang, J. Park, and D. Wigdor, "Can AI Support Fiction Writers Without Writing For Them?" in *Proceedings of the 2nd Workshop on Intelligent and Interactive Writing Assistants (In2Writing)*, 04 2023. [Online]. Available: [https://cdn.glitch.global/d058c114-3406-43be-8a3c-d3afff35eda2/paper26\\_2023.pdf](https://cdn.glitch.global/d058c114-3406-43be-8a3c-d3afff35eda2/paper26_2023.pdf)
- [159] K. Arnold, A. Volzer, and N. Madrid, "Generative models can help writers without writing for them," in *Joint Proceedings of the ACM IUI 2021 Workshops*, 04 2021. [Online]. Available: <https://ceur-ws.org/Vol-2903/IUI21WS-HAIGEN-1.pdf>
- [160] Wikipedia, "A Hundred Thousand Billion Poems," Wikipedia, 12 2024. [Online]. Available: [https://en.wikipedia.org/wiki/A\\_Hundred\\_Thousand\\_Billion\\_Poems](https://en.wikipedia.org/wiki/A_Hundred_Thousand_Billion_Poems)
- [161] R. Queneau, *Cent mille Millions de poèmes*. Bantam Books, 1965.
- [162] R. Dahl, *The Great Automatic Grammatizator Other Stories*. Penguin Books, 2017.
- [163] S. Colton, J. Goodwin, and T. Veale, "Full-FACE Poetry Generation," in *Proceedings of the 3rd International Conference on Computational Creativity, ICCO 2012*, 2012, pp. 95–102. [Online]. Available: <http://computationalcreativity.net/icco2012/wp-content/uploads/2012/05/095-Colton.pdf>
- [164] C. Linardaki, "Poetry at the first steps of Artificial Intelligence," *Humanist Studies the Digital Age*, vol. 7, 05 2022.
- [165] C. Strachey, "The 'Thinking' Machine," *Encounter* 3.4, pp. 25–31., 1954.
- [166] N. Montfort and N. Fedorova, "Small-Scale Systems and Computational Creativity," in *Proceedings of the Third International Conference on Computational Creativity*, 2012, pp. 1–5. [Online]. Available: [https://nickm.com/if/montfort\\_fedorova\\_icco\\_2012.pdf](https://nickm.com/if/montfort_fedorova_icco_2012.pdf)
- [167] M. Riedl, "An Introduction to AI Story Generation," Medium, 06 2021. [Online]. Available: <https://mark-riedl.medium.com/an-introduction-to-ai-story-generation-7f99a450f615>
- [168] J. Weizenbaum, "ELIZA—a computer program for the study of natural language communication between man and machine," *Commun. ACM*, vol. 9, no. 1, p. 36–45, Jan. 1966. [Online]. Available: <https://doi.org/10.1145/365153.365168>

- [169] B. Tarnoff, "Weizenbaum's nightmares: how the inventor of the first chatbot turned against AI," *The Guardian*, 07 2023. [Online]. Available: <https://www.theguardian.com/technology/2023/jul/25/joseph-weizenbaum-inventor-eliza-chatbot-turned-against-artificial-intelligence-ai>
- [170] R. C. Schank, R. P. Abelson, and R. S. Wyer, "Knowledge and memory: The Real Story," 2009. [Online]. Available: [https://web.archive.southampton.ac.uk/cogprints.org/636/1/KnowledgeMemory\\_SchankAbelson\\_d.html](https://web.archive.southampton.ac.uk/cogprints.org/636/1/KnowledgeMemory_SchankAbelson_d.html)
- [171] M. Lebowitz, "Creating characters in a story-telling universe," *Poetics*, vol. 13, pp. 171–194, 1984. [Online]. Available: <https://api.semanticscholar.org/CorpusID:58485006>
- [172] D. Jurafsky and J. H. Martin, "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models," 2025. [Online]. Available: [https://web.stanford.edu/~jurafsky/slp3/ed3book\\_Jan25.pdf](https://web.stanford.edu/~jurafsky/slp3/ed3book_Jan25.pdf)
- [173] M. Foldager, E. Simonsen, J. Lassen, L. S. Petersen, B. Oranje, B. Aggernæs, and M. Vestergaard, "Narrative coherence and mentalizing complexity are associated in fictive storytelling and autobiographical memories in typically developing children and adolescents," *Cognitive Development*, vol. 71, p. 101484, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0885201424000698>
- [174] C. Mitchell, E. Reese, K. Salmon, and P. Jose, "Narrative coherence, psychopathology, and wellbeing: Concurrent and longitudinal findings in a mid-adolescent sample," *Journal of Adolescence*, vol. 79, pp. 16–25, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0140197119302246>
- [175] G. Barbieri, F. Pachet, P. Roy, and M. Esposti, "Markov Constraints for Generating Lyrics with Style," *Frontiers in Artificial Intelligence and Applications*, vol. 242, pp. 115–120, 2012. [Online]. Available: <https://axon.cs.byu.edu/Dan/673/papers/barbieri.pdf>
- [176] Y. Li and D. Bertsekas, "Most Likely Sequence Generation for N-Grams, Transformers, HMMs, and Markov Chains, by Using Rollout Algorithms," 2024. [Online]. Available: <https://arxiv.org/abs/2403.15465>

- [177] B. Harrison, C. Purdy, and M. Riedl, "Toward Automated Story Generation with Markov Chain Monte Carlo Methods and Deep Neural Networks," *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, vol. 13, no. 2, pp. 191–197, Jun. 2021. [Online]. Available: <https://ojs.aaai.org/index.php/AIIDE/article/view/13003>
- [178] D. Poole and A. Mackworth, *Artificial Intelligence: Foundations of Computational Agents*, 3rd ed. Cambridge University Press, 2023. [Online]. Available: <https://artint.info/3e/html/ArtInt3e.html>
- [179] J. Hopkins and D. Kiela, "Automatically Generating Rhythmic Verse with Neural Networks," *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2017. [Online]. Available: <https://aclanthology.org/P17-1016/>
- [180] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 11 1997. [Online]. Available: <https://doi.org/10.1162/neco.1997.9.8.1735>
- [181] M. Ghazvininejad, X. Shi, J. Priyadarshi, and K. Knight, "Hafez: an interactive poetry generation system," *ACL Anthology*, pp. 43–48, 07 2017. [Online]. Available: <https://aclanthology.org/P17-4008/>
- [182] M. Ghazvininejad, X. Shi, Y. Choi, and K. Knight, "Generating Topical Poetry," *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2016. [Online]. Available: <https://aclanthology.org/D16-1126/>
- [183] OpenAI, "ChatGPT Overview," OpenAI, 2024. [Online]. Available: <https://openai.com/chatgpt/overview/>
- [184] S. Pichai and D. Hassabis, "Introducing Gemini: our largest and most capable AI model," Google, 12 2023. [Online]. Available: <https://blog.google/technology/ai/google-gemini-ai/#capabilities>
- [185] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention Is All You Need," *arXiv.org*, 2017. [Online]. Available: <https://arxiv.org/abs/1706.03762>

- [186] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, "Language Models are Few-Shot Learners," arXiv.org, 2020. [Online]. Available: <https://arxiv.org/abs/2005.14165>
- [187] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language Models are Unsupervised Multitask Learners," 2019. [Online]. Available: [https://cdn.openai.com/better-language-models/language\\_models\\_are\\_unsupervised\\_multitask\\_learners.pdf](https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf)
- [188] D. M. Ziegler, N. Stiennon, J. Wu, T. B. Brown, A. Radford, D. Amodei, P. Christiano, and G. Irving, "Fine-Tuning Language Models from Human Preferences," arXiv, 08 2020. [Online]. Available: <https://arxiv.org/abs/1909.08593>
- [189] A. Karpathy, "[1hr Talk] Intro to Large Language Models," YouTube, 11 2023. [Online]. Available: [https://www.youtube.com/watch?v=zjkBMFhNj\\_g](https://www.youtube.com/watch?v=zjkBMFhNj_g)
- [190] E. Sheng and D. C. Uthus, "Investigating Societal Biases in a Poetry Composition System," *ACL Anthology*, pp. 93–106, 12 2020. [Online]. Available: <https://aclanthology.org/2020.gebnlp-1.9/>
- [191] I. Levent and L. Shroff, "The Model is the Message," in *The Second Workshop on Intelligent and Interactive Writing Assistants co-located with The ACM CHI Conference on Human Factors in Computing Systems (CHI 2023)*, 2023. [Online]. Available: [https://cdn.glitch.global/d058c114-3406-43be-8a3c-d3afff35eda2/paper29\\_2023.pdf](https://cdn.glitch.global/d058c114-3406-43be-8a3c-d3afff35eda2/paper29_2023.pdf)
- [192] M. Forman and M. A. Neal, Eds., *That's the joint! : the hip-hop studies reader*. Routledge, 2004.
- [193] T. Rose, *Black Noise: Rap Music and Black Culture in Contemporary America*. Wesleyan University Press, 1994.
- [194] S. Carter, *Decoded*. Spiegel Grau, 2010.
- [195] J. Chang, *Can't Stop Won't Stop: A History of the Hip-Hop Generation*. St. Martin's Press, 2005.

- [196] Z. Lapos, "The Evolution of Hip-hop by Music Eras," Zack Lapos Design, 2020. [Online]. Available: <https://www.zacklaposdesign.com/work/hiphopmusiceras>
- [197] J. G. Schloss, *Making Beats: The Art of Sample-Based Hip-Hop*. Wesleyan University Press, 2004.
- [198] M. D'Errico, *Off the grid: instrumental hip-hop and experimentalism after the golden age*, ser. Cambridge Companions to Music. Cambridge University Press, 2015, p. 280–291.
- [199] M. Forman, *The 'Hood Comes First: Race, Space, and Place in Rap and Hip-Hop*. Wesleyan University Press, 2002.
- [200] R. Gallagher, "'All the Other Players Want to Look at My Pad": Grime, Gaming, and Digital Identity," *G|A|M|E Games as Art, Media, Entertainment*, vol. 1, 12 2017. [Online]. Available: <https://www.gamejournal.it/all-the-other-players-want-to-look-at-my-pad-grime-gaming-and-digital-identity-work/>
- [201] D. Hancox, "Ringtones and Playstations: The Tools of Grime's Low-Fidelity All-Stars," *How We Get To Next*, 07 2018. [Online]. Available: <https://www.howwegettonext.com/ringtones-and-playstations-the-tools-of-grimes-low-fidelity-all-stars/>
- [202] H. S. Alim, *Roc the Mic Right*. Routledge, 09 2006.
- [203] B. Duinker, "Auto-Tune as instrument: trap music's embrace of a repurposed technology," *Popular Music*, vol. 43, no. 2, p. 212–236, 2024.
- [204] S. Harney and F. Moten, *The Undercommons: Fugitive Planning and Black Study*. Minor Compositions, 2013.
- [205] E. Malmi, P. Takala, H. Toivonen, T. Raiko, and A. Gionis, "DopeLearning: A Computational Approach to Rap Lyrics Generation," in *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.
- [206] J. Fairfax and J. Moat, *The Way to Write*, Foreword by Ted Hughes. Revised and Updated ed. Penguin Books, 01 1998.

- [207] L. Carine, "The Relationship Between Improvisation and Cognition," Ph.D. dissertation, 01 2012. [Online]. Available: <https://uhra.herts.ac.uk/bitstream/handle/2299/8890/05107372%20Lewis%20Carine%20-%20final%20PhD%20submission.pdf?sequence=1>
- [208] J. Schlabach, "We're free: The impact of a rap writing music therapy intervention on self-esteem of at-risk adolescents in a public middle school setting," Ph.D. dissertation, 05 2015. [Online]. Available: [https://etd.ohiolink.edu/apexprod/rws\\_etd/send\\_file/send?accession=ohiou1429108872&disposition=attachment](https://etd.ohiolink.edu/apexprod/rws_etd/send_file/send?accession=ohiou1429108872&disposition=attachment)
- [209] Flip Top Battles, "Flip Top Battle League," YouTube, 02 2025. [Online]. Available: <https://www.youtube.com/channel/UCBdHwFIE4AJWSa3Wxdu7bAQ>
- [210] B. Todd, "Rapper and poet Mark Grist gives tips for rap battle global domination," Red Bull, 04 2018. [Online]. Available: <https://www.redbull.com/gb-en/mark-grist-rap-battle-tips>
- [211] A. T. Landau and C. J. Limb, "The Neuroscience of Improvisation," *Music Educators Journal*, vol. 103, no. 3, pp. 27–33, 2017. [Online]. Available: <https://doi.org/10.1177/0027432116687373>
- [212] J. Kaur, M. Trajkova, A. Knowlton, and B. Magerko, "Bringing LuminAI to Life: Studying Dancers' Perceptions of a Co-Creative AI in Dance Improvisation Class," in *Proceedings of the 2025 Conference on Creativity and Cognition*, ser. CC '25. New York, NY, USA: Association for Computing Machinery, 2025, p. 859–871. [Online]. Available: <https://doi.org/10.1145/3698061.3726925>
- [213] L. V. Zerkowski, Z. Wang, I. Vidrin, and M. Pettee, "Invisible Strings: Revealing Latent Dancer-to-Dancer Interactions with Graph Neural Networks," 2025. [Online]. Available: <https://arxiv.org/abs/2503.04816>
- [214] Government Office for Science, "Urbanisation Trends," Apr. 2021. [Online]. Available: [https://assets.publishing.service.gov.uk/media/60def934e90e077177d06f87/GO-Science\\_Trend\\_Deck\\_-\\_Full\\_Version\\_-\\_Spring\\_2021\\_\\_1\\_.pdf](https://assets.publishing.service.gov.uk/media/60def934e90e077177d06f87/GO-Science_Trend_Deck_-_Full_Version_-_Spring_2021__1_.pdf)
- [215] Z. Yin, F. Reuben, S. Stepney, and T. Collins, "'A Good Algorithm Does Not Steal – It Imitates' : The Originality Report as a Means of Measuring When a Music Generation Algorithm Copies Too Much," in *Artificial*

- Intelligence in Music, Sound, Art and Design*, 04 2021. [Online]. Available: [https://link.springer.com/chapter/10.1007/978-3-030-72914-1\\_24](https://link.springer.com/chapter/10.1007/978-3-030-72914-1_24)
- [216] Urban Design Group, "What is urban design?" Urban Design Group, 09 2019. [Online]. Available: <https://www.udg.org.uk/about/what-is-urban-design>
- [217] R. Anadol, "Synaesthetic Architecture: A Building Dreams," *Architectural Design*, vol. 90, pp. 76–85, 05 2020. [Online]. Available: [https://onlinelibrary.wiley.com/doi/epdf/10.1002/ad.2572?saml\\_referrer](https://onlinelibrary.wiley.com/doi/epdf/10.1002/ad.2572?saml_referrer)
- [218] L. Crook, "Camille walala shares colourful vision to make oxford street a car-free shopping district," *Dezeen*, 11 2020. [Online]. Available: <https://www.dezeen.com/2020/11/23/camille-walala-car-free-oxford-street-proposal>
- [219] Yinka Ilori Studio, "Yinka Ilori Studio," Yinka Ilori Studio, 2022. [Online]. Available: <https://yinkailori.com>
- [220] P. Hildreth, "What is localism, and what implications do different models have for managing the local economy?" *Local Economy*, vol. 26, no. 8, pp. 702–714, 2011. [Online]. Available: <https://doi.org/10.1177/0269094211422215>
- [221] A. Sule and B. Inkster, "A hip-hop state of mind," *The Lancet Psychiatry*, vol. 1, pp. 494–495, 12 2014.
- [222] C. Rust, J. Mottram, and J. Till, "Review of practice-led research in art, design & architecture," UK, 2007. [Online]. Available: <https://shura.shu.ac.uk/7596/>
- [223] Swansea Council, "Creative hub to support digital art innovators - swansea," Swansea Council, 2022. [Online]. Available: <https://www.swansea.gov.uk/article/14276/Creative-hub-to-support-digital-art-innovators>
- [224] The Crown Estate, "The Home of Urban Excellence inspiring the next generation of Londoners opens in the West End," The Crown Estate, 01 2024. [Online]. Available: <https://www.thecrownestate.co.uk/news/the-home-of-urban-excellence-inspiring-the-next-generation-of-londoners>
- [225] —, "Case study: Arts ARKADE," The Crown Estate, 07 2024. [Online]. Available: <https://www.thecrownestate.co.uk/sustainability/case-study-arts-arkade>

- [226] KOMBAT, "From Cranes to Arts ARKADE," YouTube, 03 2022. [Online]. Available: <https://youtu.be/FRnuqA-ji7I>
- [227] G. Halegoua and E. Polson, "Exploring 'digital placemaking'," *Convergence*, vol. 27, no. 3, pp. 573–578, 2021. [Online]. Available: <https://doi.org/10.1177/13548565211014828>
- [228] C. Bannon, "Cranes music shop in Swansea is closing after more than 30 years," *WalesOnline*, 12 2018. [Online]. Available: <https://www.walesonline.co.uk/news/wales-news/cranes-music-shop-swansea-closing-15519743>
- [229] Wikipedia, "Tilt brush," Wikipedia, 12 2019. [Online]. Available: [https://en.wikipedia.org/wiki/Tilt\\_Brush](https://en.wikipedia.org/wiki/Tilt_Brush)
- [230] R. B.V., "Avenue arena – resolume," *Resolume.com*, 2024. [Online]. Available: <https://www.resolume.com/software/avenue-arena>
- [231] Serpentine Galleries, "Future Art Ecosystems Issue 1. Art x Advanced Technologies," 02 2020. [Online]. Available: [https://serpentine-uploads.s3.amazonaws.com/uploads/2020/07/Future-Art-Ecosystems-1-Art-and-Advanced-Technologies\\_July\\_2020.pdf](https://serpentine-uploads.s3.amazonaws.com/uploads/2020/07/Future-Art-Ecosystems-1-Art-and-Advanced-Technologies_July_2020.pdf)
- [232] L. Hespanhol and M. Tomitsch, "Strategies for Intuitive Interaction in Public Urban Spaces," *Interacting with Computers*, 01 2015.
- [233] Y Combinator, "A Minimum Viable Product Is Not a Product, It's a Process: Building Product, Experimentation, MVP | Y Combinator," Y Combinator, 2018. [Online]. Available: <https://www.ycombinator.com/library/4Q-a-minimum-viable-product-is-not-a-product-it-s-a-process>
- [234] —, "How to Plan an MVP: Getting Started, Building Product," YC Startup Library, 2012. [Online]. Available: <https://www.ycombinator.com/library/6f-how-to-plan-an-mvp>
- [235] Nielsen Norman Group, "The Wizard of Oz Method in UX," Nielsen Norman Group, 2022. [Online]. Available: <https://www.nngroup.com/articles/wizard-of-oz/>
- [236] KOMBAT, "Walk through: Hip hop for health," YouTube, 01 2023. [Online]. Available: <https://www.youtube.com/watch?v=uCDCrkn8OzU>

- [237] BBC News, "Piccadilly Circus: New creative hub to help West End 'remain relevant'," *BBC News*, 01 2024. [Online]. Available: <https://www.bbc.co.uk/news/uk-england-london-68067342>
- [238] M. Ellingham, "After hours with 10 Foot, London's most notorious graffiti writer," *Financial Times*, 02 2023. [Online]. Available: <https://www.ft.com/content/45a184ee-b7d9-4c16-b1c2-71def32ccf98>
- [239] C. Davies, "'Tox' graffiti artist convicted of criminal damage," *The Guardian*, 06 2011. [Online]. Available: <https://www.theguardian.com/artanddesign/2011/jun/07/tox-graffiti-artist-criminal-damage>
- [240] G. Barradale, "Graffiti writer 10Foot assembles Banksy, Kneecap and more for radical Big Issue takeover," *Big Issue*, 03 2025. [Online]. Available: <https://www.bigissue.com/news/10foot-big-issue-takeover-banksy-kneecap-alan-moore/>
- [241] C. Cooper-Fiske, "Banksy to interview 'UK's most imprisoned graffiti writer' for Big Issue," *The Independent*, 03 2025. [Online]. Available: <https://www.independent.co.uk/news/uk/home-news/banksy-big-issue-london-camden-irish-b2714883.html>
- [242] M. Howe, "Graffiti exhibition backed by Banksy is shut after vandals daub 'f\*\*\* the King' on building owned by Crown Estate," *The Standard*, 04 2025. [Online]. Available: <https://www.standard.co.uk/news/london/graffiti-exhibition-banksy-long-dark-tunnel-king-crown-estate-b1222299.html>
- [243] J. Simpson, "Banksy-backed art exhibition shuts after anti-King graffiti," *The Times*, 04 2025. [Online]. Available: <https://www.thetimes.com/uk/london/article/banksy-backed-art-exhibition-shuts-king-charles-lk00btk6b>
- [244] BBC, "The Rap Game UK: What is a cypher?" *BBC Three*, 09 2022. [Online]. Available: <https://www.bbc.co.uk/bbcthree/article/b4833d5a-58e9-4efb-bf71-9102fbd4bb04>
- [245] M. Foth, "Lessons from urban guerrilla placemaking for smart city commons," in *Proceedings of the 8th International Conference on Communities and Technologies*, ser. CT '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 32–35. [Online]. Available: <https://doi.org/10.1145/3083671.3083707>

- [246] Serpentine Galleries, "Future Art Ecosystems Vol 2. Art x Metaverse," 07 2021. [Online]. Available: [https://futureartecosystems.org/FAE2\\_ArtxMetaverse\\_digital-small.pdf](https://futureartecosystems.org/FAE2_ArtxMetaverse_digital-small.pdf)
- [247] J. Mlynar, F. Bahrami, A. Ourednik, N. Mutzner, H. Verma, and H. Alavi, "AI beyond Deus ex Machina – Reimagining Intelligence in Future Cities with Urban Experts," *CHI Conference on Human Factors in Computing Systems*, 04 2022.
- [248] teamLab, "teamlab," Teamlab.art, 2022. [Online]. Available: <https://www.teamlab.art/>
- [249] J. Deller, "More poetry is needed," 2014. [Online]. Available: <https://www.themoderninstitute.com/artists/jeremy-deller/works/more-poetry-is-needed-2014/359/>
- [250] C. Mahlow, "Writing Tools: Looking Back to Look Ahead," 2023. [Online]. Available: <https://arxiv.org/abs/2303.17894>
- [251] P. Karimi, K. Grace, M. Maher, and N. Davis, "Evaluating Creativity in Computational Co-Creative Systems," 07 2018.
- [252] WriteExpress, "Rhymer Online Rhyming Dictionary," Rhymer.com, 2022. [Online]. Available: <https://www.rhymer.com>
- [253] D. Uthus, M. Voitovich, and R. Mical, "Augmenting Poetry Composition with Verse by Verse," *ACLWeb*, p. 18–26, 07 2022. [Online]. Available: <https://aclanthology.org/2022.naacl-industry.3/>
- [254] B. Shneiderman, "Creativity support tools," *Communications of the ACM*, vol. 45, no. 10, pp. 116–120, 2002.
- [255] I. Olátúnjì, "Why try to build try to build a co-creative poetry system that makes people feel that they have creative superpowers?" in *Joint Proceedings of the IUI 2023 Workshops: HAI-GEN, co-located with the ACM International Conference on Intelligent User Interfaces (IUI 2023), Sydney, Australia, March 27-31, 2023*, ser. CEUR Workshop Proceedings. CEUR-WS.org, 2023. [Online]. Available: <https://hai-gen.github.io/2023/papers/1594-paper-HAIGEN-OlantunjiTunde.pdf>

- [256] D. Buschek, L. Mecke, F. Lehmann, and H. Dang, "Nine Potential Pitfalls when Designing Human-AI Co-Creative Systems," in *Joint Proceedings of the ACM IUI 2021 Workshops, April 13-17, 2021, College Station, USA*, vol. Vol-2903, 04 2021, pp. 1–10. [Online]. Available: <https://ceur-ws.org/Vol-2903/IUI21WS-HAIGEN-2.pdf>
- [257] D. Yang, Y. Zhou, Z. Zhang, T. J.-J. Li, and R. LC, "AI as an Active Writer: Interaction Strategies with Generated Text in Human-AI Collaborative Fiction Writing 56-65, year = 2022, url = <http://ceur-ws.org/Vol-3124/paper6.pdf>, researchr = <https://researchr.org/publication/YangZZLL22>, cites = 0, citedby = 0, pages = 56-65, booktitle = Joint Proceedings of the IUI 2022 Workshops: APEX-UI, HAI-GEN, HEALTHI, HUMANIZE, TExSS, SOCIALIZE co-located with the ACM International Conference on Intelligent User Interfaces (IUI 2022), Virtual Event, Helsinki, Finland, March 21-22, 2022, editor = Alison Smith-Renner and Ofra Amir, volume = 3124, series = CEUR Workshop Proceedings, publisher = CEUR-WS.org."
- [258] L. Memmert, J. Mies, and E. Bittner, "Brainstorming with a Generative Language Model: The Role of Creative Ability and Tool-Support for Brainstorming Performance," in *Conference: 45th International Conference on Information Systems (ICIS)At: Bangkok, Thailand, 12 2024*.
- [259] G. Ding, Y. He, K. Yi, and S. Li, "Using the divergent association task to measure divergent thinking in Chinese elementary school students," *Thinking Skills and Creativity*, vol. 52, p. 101503, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1871187124000415>
- [260] A. Calderwood, V. Qiu, K. Gero, and L. B. Chilton, "How Novelists Use Generative Language Models: An Exploratory User Study," in *HAI-GEN+user2agent@IUI, 2020*.
- [261] M. Csikszentmihalyi, *Creativity : the psychology of discovery and invention*. Harper Perennial Modern Classics, 2013.
- [262] S. Chaudhuri and J. Bhattacharya, "Poetic Break: Incubation for Associative Creativity," *Creativity Research Journal*, vol. 0, no. 0, pp. 1–14, 2025. [Online]. Available: <https://doi.org/10.1080/10400419.2025.2466571>
- [263] S. L. Gable, E. A. Hopper, and J. W. Schooler, "When the Muses Strike: Creative Ideas of Physicists and Writers Routinely Occur During Mind Wandering,"

- Psychological Science*, vol. 30, no. 3, pp. 396–404, 2019, pMID: 30653407. [Online]. Available: <https://doi.org/10.1177/0956797618820626>
- [264] A. G. Greenwald, “Within-subjects designs: To use or not to use?” *Psychological Bulletin*, vol. 83, pp. 314–320, 1976.
- [265] C. P. Lucas, “The order effect: reflections on the validity of multiple test presentations,” *Psychological Medicine*, vol. 22, pp. 197–202, 02 1992.
- [266] E. F. Corriero, *Counterbalancing*. In: *The SAGE Encyclopedia of Communication Research Methods*. SAGE Publications, Inc., 2017. [Online]. Available: <https://doi.org/10.4135/9781483381411.n103>
- [267] H. W. Reese, “Counterbalancing and Other Uses of Repeated-Measures Latin-Square Designs: Analyses and Interpretations,” *Journal of Experimental Child Psychology*, vol. 64, no. 1, pp. 137–158, 1997. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0022096596923334>
- [268] K. Uittenhove, S. Jeanneret, and E. Vergauwe, “From Lab-Testing to Web-Testing in Cognitive Research: Who You Test is More Important than how You Test,” *Journal of Cognition*, Jan 2023.
- [269] F. Dandurand, T. Shultz, and K. Onishi, “Comparing online and lab methods in a problem-solving experiment,” *Behavior Research Methods*, vol. 40, pp. 428–34, 05 2008.
- [270] K. McConnell, F. Hintz, and A. Meyer, “Individual differences in online research: Comparing lab-based and online administration of a psycholinguistic battery of linguistic and domain-general skills,” *Behavior Research Methods*, vol. 57:22, 12 2025. [Online]. Available: <https://doi.org/10.3758/s13428-024-02533-x>
- [271] A. Jordanous, “A Standardised Procedure for Evaluating Creative Systems: Computational Creativity Evaluation Based on What it is to be Creative,” *Cognitive Computation*, vol. 4, 09 2012.
- [272] —, “Evaluating computational creativity: a standardised procedure for evaluating creative systems and its application,” Ph.D. dissertation, University of Sussex, 4 2013. [Online]. Available: [https://sussex.figshare.com/articles/thesis/Evaluating\\_computational\\_creativity\\_a\\_standardised\\_procedure\\_for\\_evaluating\\_creative\\_systems\\_and\\_its\\_application/23395574](https://sussex.figshare.com/articles/thesis/Evaluating_computational_creativity_a_standardised_procedure_for_evaluating_creative_systems_and_its_application/23395574)

- 
- [273] L. Fang, Y. Wang, Z. Liu, C. Zhang, S. Jegelka, J. Gao, B. Ding, and Y. Wang, "What is Wrong with Perplexity for Long-context Language Modeling?" 2024. [Online]. Available: <https://arxiv.org/abs/2410.23771>
- [274] H. Hirjee, "Rhyme, Rhythm, and Rhubarb: Using Probabilistic Methods to Analyze Hip Hop, Poetry, and Misheard Lyrics," Ph.D. dissertation, University of Waterloo, 2010. [Online]. Available: <https://uwspace.uwaterloo.ca/handle/10012/5419>
- [275] B. Shneiderman, *Human-Centered AI*. Oxford University Press, 01 2022.
- [276] F. Lehmann, N. Markert, H. Dang, and D. Buschek, "Suggestion Lists vs. Continuous Generation: Interaction Design for Writing with Generative Models on Mobile Devices Affect Text Length, Wording and Perceived Authorship," in *Proceedings of Mensch Und Computer 2022*, ser. MuC '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 192–208. [Online]. Available: <https://doi.org/10.1145/3543758.3543947>
- [277] G. Davis and M. Grierson, "Investigating attitudes of professional writers to gpt text generation ai based creative support tools," July 2022. [Online]. Available: <https://ualresearchonline.arts.ac.uk/id/eprint/18621/>
- [278] T. Berber Sardinha, "AI-generated vs human-authored texts: A multidimensional comparison," *Applied Corpus Linguistics*, vol. 4, no. 1, p. 100083, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2666799123000436>
- [279] N. Köbis and L. D. Mossink, "Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry," *Computers in Human Behavior*, vol. 114, p. 106553, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0747563220303034>
- [280] D. Anguiano and L. Beckett, "How Hollywood Writers Triumphed Over AI – and Why It Matters," *The Guardian*, 10 2023. [Online]. Available: <https://www.theguardian.com/culture/2023/oct/01/hollywood-writers-strike-artificial-intelligence>
- [281] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," arXiv, 06 2014. [Online]. Available: <https://arxiv.org/abs/1406.2661>

- [282] C. Baber, *Embodying Design: An Applied Science of Radical Embodied Cognition*. The MIT Press, 03 2022. [Online]. Available: <https://doi.org/10.7551/mitpress/12419.001.0001>
- [283] K. Lindgaard and H. Wesselius, "Once More, with Feeling: Design Thinking and Embodied Cognition," *She Ji: The Journal of Design, Economics, and Innovation*, vol. 3, no. 2, pp. 83–92, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2405872617300291>
- [284] J. Asai, K. Sakaguchi, R. Le Bras, D. Radev, Y. Choi, and N. A. Smith, "A Call for Clarity in Beam Search: How It Works and When It Stops," in *Proceedings of the 2024 Joint International Conference on Computational Linguistics Language Resources and Evaluation (LREC-COLING 2024)*, N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, and N. Xue, Eds. Torino, Italia: ELRA and ICCL, May 2024, pp. 77–90. [Online]. Available: <https://aclanthology.org/2024.lrec-main.7/>
- [285] G. Franceschelli and M. Musolesi, "Creative Beam Search: LLM-as-a-Judge For Improving Response Generation," 2024. [Online]. Available: <https://arxiv.org/abs/2405.00099>
- [286] T. Amabile, "The Social Psychology of Creativity: A Componential Conceptualization," *Journal of Personality and Social Psychology*, vol. 45, pp. 357–376, 08 1983.
- [287] A. Vijayakumar, M. Cogswell, R. Selvaraju, Q. Sun, S. Lee, D. Crandall, and D. Batra, "Diverse Beam Search for Improved Description of Complex Scenes," Apr 2018. [Online]. Available: <https://ojs.aaai.org/index.php/AAAI/article/view/12340>
- [288] L. Zheng, W.-L. Chiang, Y. Sheng, S. Zhuang, Z. Wu, Y. Zhuang, Z. Lin, Z. Li, D. Li, E. P. Xing, H. Zhang, J. E. Gonzalez, and I. Stoica, "Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena," pp. 46 595–46 623, 2023. [Online]. Available: [https://proceedings.neurips.cc/paper\\_files/paper/2023/file/91f18a1287b398d378ef22505bf41832-Paper-Datasets\\_and\\_Benchmarks.pdf](https://proceedings.neurips.cc/paper_files/paper/2023/file/91f18a1287b398d378ef22505bf41832-Paper-Datasets_and_Benchmarks.pdf)
- [289] P. Sawicki, M. Grześ, D. Brown, and F. Góes, "Can Large Language Models Outperform Non-Experts in Poetry Evaluation? A Comparative Study Using the Consensual Assessment Technique," 2025. [Online]. Available: <https://arxiv.org/abs/2502.19064>

- [290] Wikipedia, "Speech recognition," Wikipedia, 07 2019. [Online]. Available: [https://en.wikipedia.org/wiki/Speech\\_recognition](https://en.wikipedia.org/wiki/Speech_recognition)
- [291] E. Protalinski, "Google's speech recognition technology now has a 4.9% word error rate," VentureBeat, 05 2017. [Online]. Available: <https://venturebeat.com/business/googles-speech-recognition-technology-now-has-a-4-9-word-error-rate/>
- [292] S. Hollands, D. Blackburn, and H. Christensen, "Evaluating the Performance of State-of-the-Art ASR Systems on Non-Native English using Corpora with Extensive Language Background Variation," in *Proceedings of the Annual Conference of the International Speech Communication Association. Interspeech 2022, 18-22 Sep 2022, Incheon, Korea*, Interspeech 2022. International Speech Communication Association (ISCA), 01 2022, pp. 3958–3962. [Online]. Available: [https://eprints.whiterose.ac.uk/207725/1/hollands22\\_interspeech.pdf](https://eprints.whiterose.ac.uk/207725/1/hollands22_interspeech.pdf)
- [293] Apple, "Humanizing Word Error Rate for ASR Transcript Readability and Accessibility," Apple Machine Learning Research, 03 2024. [Online]. Available: <https://machinelearning.apple.com/research/humanizing-wer>
- [294] A. B. Wassink, C. Gansen, and I. Bartholomew, "Uneven success: automatic speech recognition and ethnicity-related dialects," *Speech Communication*, vol. 140, pp. 50–70, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0167639322000486>
- [295] N. Akingbe and P. A. Onanuga, "Leveraging Poetry on the Airwaves: Appropriating Linguistic Creativity in Nigerian Hip Hop Lyrics," *Journal of the Musical Arts in Africa*, vol. 15, pp. 19–40, 07 2018. [Online]. Available: <https://www.tandfonline.com/doi/epdf/10.2989/18121004.2018.1534335>
- [296] C. Coscarello, "The Word Out : A Stylistic Analysis of Rap Music," Ph.D. dissertation, Montclair State University, 01 2003. [Online]. Available: <https://digitalcommons.montclair.edu/etd/800/>
- [297] Music.AI, "Comparative Analysis Shows Music.AI Outperforms OpenAI in Lyric Transcription Accuracy," Music.ai, 2024. [Online]. Available: <https://music.ai/blog/research/lyric-transcription-benchmark/>

- [298] L. Ou, X. Gu, and Y. Wang, "Transfer Learning of wav2vec 2.0 for Automatic Lyric Transcription," 2022. [Online]. Available: <https://arxiv.org/abs/2207.09747>
- [299] L. Xue, K. Song, D. Wu, X. Tan, N. Zhang, T. Qin, W.-Q. Zhang, and T.-Y. Liu, "DeepRapper: Neural Rap Generation with Rhyme and Rhythm Modeling," in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 01 2021, pp. 69–81.
- [300] Wikipedia, "Mobb Deep," Wikipedia, 06 2020. [Online]. Available: [https://en.wikipedia.org/wiki/Mobb\\_Deep](https://en.wikipedia.org/wiki/Mobb_Deep)
- [301] M. Sheppard, "Perception and Transmission of Sonic QR Codes," Ph.D. dissertation, Anglia Ruskin University, 10 2020. [Online]. Available: [https://aru.figshare.com/articles/thesis/Perception\\_and\\_transmission\\_of\\_sonic\\_QR\\_codes/23958012?file=42927883](https://aru.figshare.com/articles/thesis/Perception_and_transmission_of_sonic_QR_codes/23958012?file=42927883)
- [302] B. McFee, C. Raffel, D. Liang, D. Ellis, M. Mcvcar, E. Battenberg, and O. Nieto, "librosa: Audio and Music Signal Analysis in Python," 01 2015, pp. 18–24.
- [303] D. Selwood, "Eminem classic freestyle - random word rhyme," [www.youtube.com](http://www.youtube.com), 2013. [Online]. Available: <https://www.youtube.com/watch?v=z9BpzsOCAAQ>
- [304] TheFlexEffect, "Lowkey Is Given 5 Random Words To Freestyle," [www.youtube.com](http://www.youtube.com), 2011. [Online]. Available: <https://www.youtube.com/watch?v=YMYT5FpJrA0>
- [305] K. Dennis, "Discussion about LYRICAL KOMBAT Battle Rap Game," 08 2020.
- [306] R. Budiu and K. Moran, "How many participants for quantitative usability studies: A summary of sample-size recommendations," Nielsen Norman Group, 07 2021. [Online]. Available: <https://www.nngroup.com/articles/summary-quant-sample-sizes/>
- [307] L. Faulkner, "Beyond the five-user assumption: Benefits of increased sample sizes in usability testing," *Behavior research methods, instruments, computers : a journal of the Psychonomic Society, Inc*, vol. 35, pp. 379–83, 09 2003.
- [308] W. Nybom and M. Grierson, "Personality in Personalisation: A User Study with an Interactive Narrative, a Personality Test and a Personalised Short

- 
- Story," *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, vol. 435, May 2022. [Online]. Available: <https://link.springer.com/bookseries/8197>
- [309] M. Elnadi and Y. O. Abdallah, "Industry 4.0: critical investigations and synthesis of key findings," *Management Review Quarterly*, vol. 74, 01 2023.
- [310] A. Johnson, "Dirty South Feminism: The Girlies Got Somethin' to Say Too! Southern Hip-Hop Women, Fighting Respectability, Talking Mess, and Twerking Up the Dirty South," *Religions*, vol. 12, no. 11, 2021. [Online]. Available: <https://www.mdpi.com/2077-1444/12/11/1030>
- [311] N. Cherry, K. Gaby, E. Read, D. Cannell, and A. Kennedy, "First Ladies of Hip-Hop," 2017. [Online]. Available: <https://www.bbc.co.uk/iplayer/episodes/p0g4tcw5/first-ladies-of-hiphop>
- [312] M. A. Runco, "AI can only produce artificial creativity," *Journal of Creativity*, vol. 33, no. 3, p. 100063, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2713374523000225>
- [313] Ibùkún Ọlátúnjí and M. Sheppard, "Next Token Prediction Is a Dead End for Creativity," 2025. [Online]. Available: <https://arxiv.org/abs/2505.19277>
- [314] U.S. Copyright Office, "Copyright and Artificial Intelligence, Part 1: Digital Replicas," pp. 1–72, 07 2024. [Online]. Available: <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-1-Digital-Replicas-Report.pdf>
- [315] The Ministry of Culture Denmark, "Broad Agreement on Deep-fakes Gives Everyone the Right to Their Own Body and Voice," Kum.dk, 2025. [Online]. Available: <https://kum.dk/aktuelt/nyheder/bred-aftale-om-deepfakes-giver-alle-ret-til-egen-krop-og-egen-stemme>
- [316] J. Diaz, "Denmark wants you to copyright yourself. it might be the only way to stop deepfakes," *Fast Company*, 07 2025. [Online]. Available: <https://www.fastcompany.com/91360589/denmark-copyright-yourself-it-might-be-the-only-way-to-stop-deepfakes>

- [317] A. Baris, "Publicity rights in the AI era: Key takeaways from artist Arijit Singh's recent legal Victory in India," 2024. [Online]. Available: <https://ipkitten.blogspot.com/2024/08/publicity-rights-in-ai-era-key.html>
- [318] Jacob, Sir Robin, "Personal communication on jurisdictional fragmentation of voice/likeness protection and implications for DCTs," Personal communication, 2024, conversation on 14 August 2024.
- [319] A. Harcourt, "Fixing the Music Industry," 2024, policy proposals on metadata, consent, remuneration and AI uses in music. [Online]. Available: <https://www.ucl.ac.uk/ibil/publications/2024/jul/fixing-music-industry>
- [320] J. Zimmerman, E. Stolterman, and J. Forlizzi, "An analysis and critique of Research through Design: towards a formalization of a research approach," in *Proceedings of the 8th ACM Conference on Designing Interactive Systems*, ser. DIS '10. New York, NY, USA: Association for Computing Machinery, 2010, p. 310–319. [Online]. Available: <https://doi.org/10.1145/1858171.1858228>
- [321] A. Xambó and G. Roma, "Human-machine agencies in live coding for music performance," *Journal of New Music Research*, vol. 53, no. 1-2, pp. 33–46, 2024. [Online]. Available: <https://doi.org/10.1080/09298215.2024.2442355>
- [322] D. Della Casa and G. John, "Livecodelab," Livecodelab.net, 2025. [Online]. Available: <https://livecodelab.net/>

# Appendix A

## Study I: Supplementary Materials

### A.1 Artcamp: Output and Artefacts

This appendix presents selected artefacts produced during the Arts ARKADE *Artcamp* workshops. The examples illustrate how non-technical artists engaged with embodied interaction, AI-assisted narration, and iterative mark-making. The materials shown here are drawn from one representative team (“Arkadians”) whose work demonstrates the conceptual and technical exploration supported through practice-led, technology-enabled workshop methods. A short demonstration video of the team’s prototype is available at: <https://youtu.be/5MbW6cmEoAs>.

#### **Gesture-Based Mark-Making Experiments**

As part of the workshop, participants experimented with embodied mark-making using long-exposure photography and handheld LED light sources. These studies served as both creative outputs and reflective probes, helping artists consider movement, rhythm, and spatial inscription before implementing digital systems. These preliminary experiments helped participants reflect on how physical movement might translate into digital mark-making within interactive installation environments.

#### **Movement Tracking and Computational Embodiment**

Participants also explored computer vision tools such as PoseNet and BlazePose. These enabled an understanding of how machine-learning models parse motion and how gesture features can map to audiovisual responses.



Figure A.1: Continuous looping gesture producing a single uninterrupted light ribbon. This study demonstrates early exploration of fluid, calligraphic motion as a form of embodied drawing.

### **Creative Process Documentation (Team: Arkadians)**

The following text summarises the conceptual and technical thinking developed by the Arkadians team during the two-day Artcamp. It outlines how the workshop led to

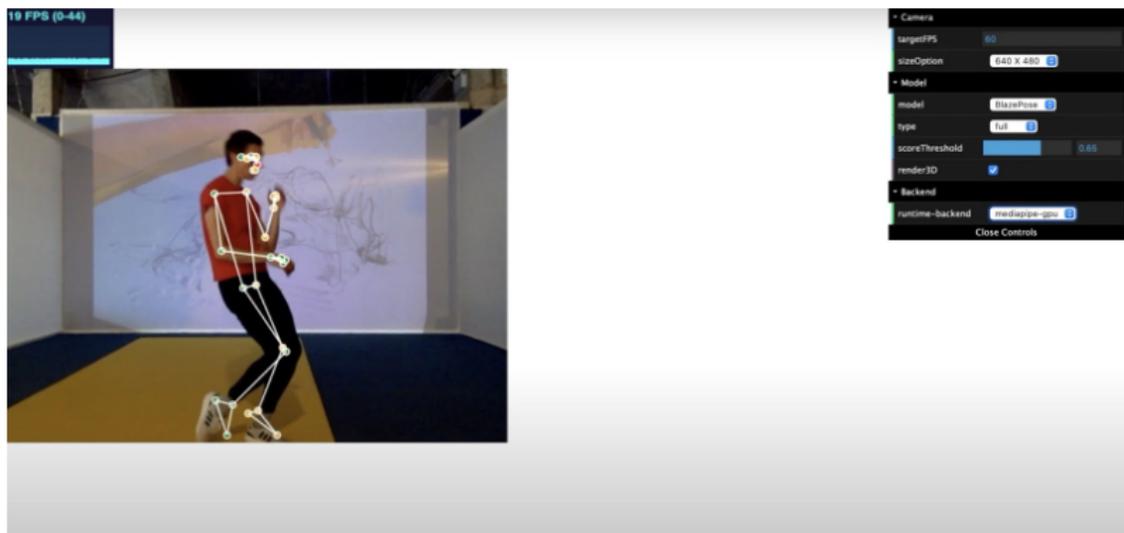


Figure A.2: Pose-estimation test using BlazePose. The skeletal overlay illustrates how movement features such as limb angle or velocity could be mapped to generative sound or visual systems.

the rapid development of hybrid systems that combine AI-generated narrative, sound, movement, and projection-based visualisation.

### Core Idea: Digital Signature

The team explored methods for engaging participants “instinctively,” focusing on intersections between physical movement, AI-generated media, and collaborative mark-making. Their concept imagined an immersive environment in which users influence sound, visual output, and narrative through embodied interaction. The team proposed capturing participants upon entry and generating a personalised “digital signature” composed of:

- minimal personal input (name, interest, environment, or symbolic choice)
- AI-generated short stories
- AI-generated or remixed music
- movement features tracked through pose estimation
- real-time virtual mark-making projected onto surrounding screens

## **Immersive Cube Environment**

The proposed installation centred on the interactive cube with projection screens on multiple sides. The experience unfolded as follows:

1. Participant provides a short descriptive input.
2. An AI model generates a short story based on the input.
3. The story drives:
  - AI-generated music
  - text-to-speech narration
4. The participant enters the cube; screens begin blank.
5. Pose-tracking cameras capture movement as narration and music play.
6. Movements are translated into virtual mark-making that fills the screens with evolving visual forms.

## **Development Questions**

During the workshop, participants identified several technical questions for further refinement:

- How should pose-estimated features map to visual mark-making?
- Can different regions of the screen be dynamically targeted?
- How might temporal constraints be introduced into the generative process?

## **Summary**

This example demonstrates how Artcamp enabled non-technical artists to conceptualise complex interactive systems integrating AI narrative generation, music, embodied interaction, and dynamic visualisation. The process reflects a practice-led mode of inquiry in which creativity, experimentation, and computational reasoning develop iteratively.

## Appendix B

# Study II: Supplementary Materials and Analyses

### Statistical Data

#### Spearman Correlation Analysis

Spearman's correlation was chosen as the data includes ordinal ratings (Likert-scale questionnaire responses), making a non-parametric correlation test appropriate to handle potential non-linearity or ordinal patterns.

Measure	Experience Level	DAT Score	Base Rhyme Score	Base Keywords
Experience Level	1.000	0.223	0.358	0.473
DAT Score	0.223	1.000	0.390	0.410
Base Rhyme Score	0.358	0.390	1.000	0.691
Base Keywords	0.473	0.410	0.691	1.000
Creativity (No Tool)	0.010	0.088	-0.071	0.170
Creativity (Support)	0.140	-0.032	-0.045	0.172
Creativity (Co-Creative)	-0.376	-0.090	-0.168	-0.249
Creativity (Autonomous)	-0.242	-0.097	-0.239	-0.238

Measure	Creativity (No Tool)	Creativity (Support)	Creativity (Co-Creative)	Creativity (Autonomous)
Experience Level	0.010	0.140	-0.376	-0.242
DAT Score	0.088	-0.032	-0.090	-0.097
Base Rhyme Score	-0.071	-0.045	-0.168	-0.239
Base Keywords	0.170	0.172	-0.249	-0.238
Creativity (No Tool)	1.000	0.633	0.293	0.205
Creativity (Support)	0.633	1.000	0.407	0.025
Creativity (Co-Creative)	0.293	0.407	1.000	-0.060
Creativity (Autonomous)	0.205	0.025	-0.060	1.000

Table B.1: Spearman Correlation Coefficients

- Moderate positive correlations exist between DAT scores, base rhyme scores, and keyword use, suggesting participants with higher divergent thinking skills (DAT) also integrated keywords more effectively and had higher rhyme density.
- Creativity ratings without tools positively correlate with creativity ratings using support tools (0.633), suggesting consistency in participants' subjective creativity assessments.

**Kruskal-Wallis Test:** This non-parametric test compares creativity ratings without tools across different experience levels (low, medium, high), suitable due to the ordinal nature of creativity ratings.

$$\chi^2 = 0.86, \quad p = 0.651$$

The result indicates no statistically significant difference in creativity ratings (without tool use) across different participant experience levels.

**Friedman Test:** A Friedman test examines differences in participant creativity ratings across the different writing tool conditions within the same individuals, appropriate for repeated measures ordinal data.

$$\chi^2(6) = 4.22, \quad p = 0.239$$

There is no statistically significant difference in creativity ratings across the different writing conditions (no tool, support, co-creative, autonomous). Participants rated their creative output consistently across different conditions.

**The Mann-Whitney U tests** revealed no statistically significant differences in average user experience ratings between participants with the highest 20% DAT scores and others ( $U = 57.0, p = 0.450$ ), between participants with the highest 20% rhyme density scores and others ( $U = 53.5, p = 0.348$ ), nor between highly experienced participants (level 3) and all other participants ( $U = 57.0, p = 0.208$ ). These findings suggest that, within this sample, subjective perceptions of user experience as measured by creativity ratings were not strongly influenced by differences in divergent thinking (DAT), rhyme proficiency, or participant experience level.

---

<b>Comparison Groups</b>	<b>U Statistic</b>	<b>p-value</b>
DAT Scores (Top 20% vs. Others)	57.0	0.450
Base Rhyme Scores (Top 20% vs. Others)	53.5	0.348
Experience Level (High (3) vs. Others)	57.0	0.208

Table B.2: Results of Mann-Whitney U tests comparing average user experience ratings between groups differentiated by top DAT scores, top rhyme density scores, and experience level.

## **Analysis Summary**

- The analysis demonstrates consistent participant assessments of creativity, regardless of tool usage, experience, or DAT scores.
- Moderate correlations were identified between DAT scores, rhyme density, and keyword integration, suggesting a meaningful link between divergent thinking and writing performance metrics.

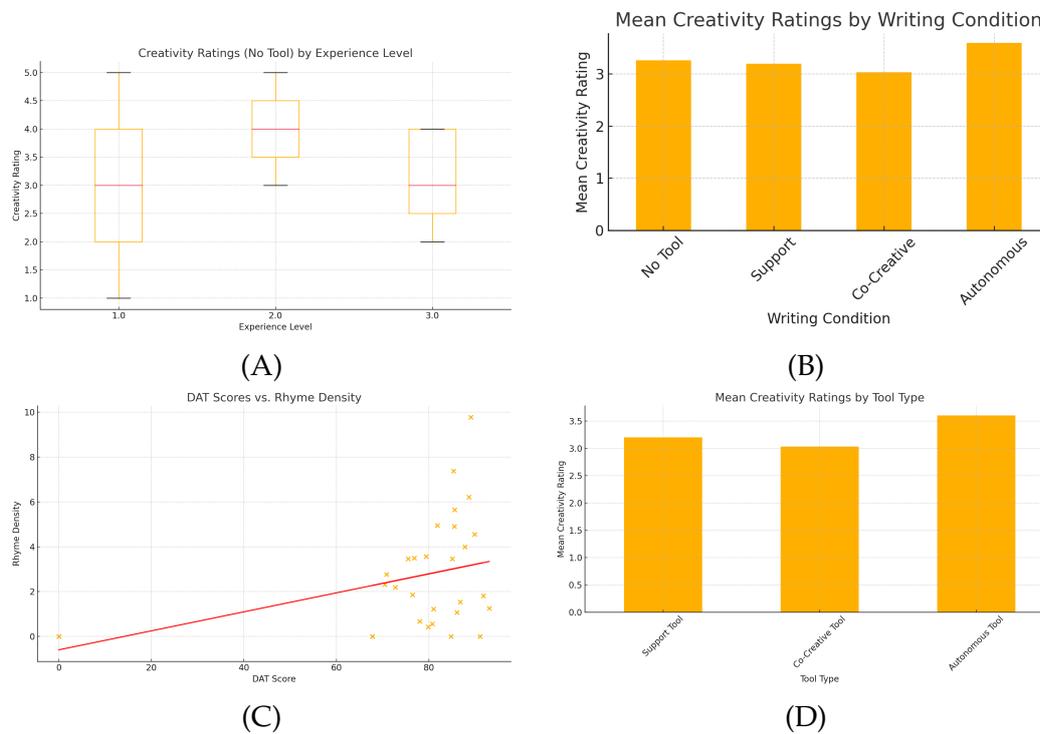


Figure B.1: Summary of statistical findings. (A) Boxplot of creativity ratings without tools by participant experience level, showing distribution and median ratings. (B) Mean creativity ratings across writing conditions (no tool, support, co-creative, autonomous). (C) Scatterplot and regression line illustrating the relationship between DAT scores and rhyme density scores for participants’ initial writing tasks. (D) Mean creativity ratings by different tool types (support, co-creative, autonomous), showing perceived tool effectiveness.

## B.1 Rhyme Types and Positions

Rhyme is not a singular phenomenon but a family of related patterns that can be categorised by type (phonological or structural similarity) and by position (where the rhyme occurs within a line or stanza). From a human perspective, these distinctions carry cultural and aesthetic value, shaping how rhyme is perceived in oral and written traditions [9, 139]. From a computational perspective, such distinctions are often simplified into single-number metrics (e.g., rhyme density), but awareness of the underlying categories clarifies both the strengths and limits of these measures. The following sections provide an overview of common rhyme types (Table B.3) and rhyme positions (Table B.4) with illustrative examples.

## Rhyme Types

Rhyme can be classified along cultural, linguistic, and computational dimensions. Table B.3 summarises common rhyme types, many of which have been operationalised differently across literary and computational studies [139, 205, 127].

Type of Rhyme	Description and Examples
<b>Syllabic</b>	Rhyme based on the final syllable, regardless of stress. Examples: <i>cleaver–silver, pitter–patter, bottle–fiddle</i> .
<b>Imperfect (or near)</b>	Rhyme between a stressed and an unstressed syllable. Example: <i>wing–caring</i> .
Weak (or unaccented)	Rhyme between unstressed syllables. Example: <i>hammer–carpenter</i> .
<b>Semirhyme</b>	Rhyme in which one word has an extra syllable. Example: <i>bend–ending</i> .
<b>Forced (or oblique)</b>	Imperfect rhyme where the match is approximate. Examples: <i>green–fiend, one–thumb</i> .
<b>Assonance</b>	Rhyme based on matching vowel sounds. Example: <i>shake–hate</i> . Sometimes considered a form of slant rhyme.
<b>Consonance</b>	Rhyme based on matching consonant sounds. Examples: <i>rabies–robbers, Roxie–Lexie</i> . Often grouped with assonance under slant rhyme.
<b>Pararhyme</b>	Rhyme in which all consonants match while vowels differ. Example: <i>tell–tall</i> .
<b>Alliteration (or head rhyme)</b>	Rhyme based on matching initial consonants. Example: <i>ship–short</i> .

Table B.3: Common types of rhyme with definitions and examples.

## Rhyme Positions

Rhymes may also be classified according to their position in a verse. In rap, rhyme placement structures flow, syncopation, and audience perception. Complex schemes such as internal or off-centred rhymes are widely recognised as markers of technical skill, demonstrating rhythmic control and the ability to subvert listener expectations.

<b>Rhyme Position</b>	<b>Description</b>
<b>Tail rhyme (end rhyme)</b>	Rhyme in the final syllable(s) of a line or verse; the most common type, forming the basis of most rhyme schemes.
<b>Internal rhyme</b>	Occurs when a word or phrase inside a line rhymes with one at the line's end, or with another word inside a different line.
<b>Off-centred rhyme</b>	A subtype of internal rhyme placed in unexpected positions within a line. Sometimes called a "misplaced-rhyme scheme" or "spoken word rhyme style."
<b>Holorime</b>	A rare, complex form where two entire lines rhyme with each other phonetically.
<b>Broken rhyme</b>	A rhyme created by splitting a word across a line break so it rhymes with the end word of another line.
<b>Cross rhyme</b>	Matches a sound at the end of one line with the same sound in the middle of the following (or preceding) line.

Table B.4: Positions of rhymes within verse lines and their stylistic functions.

## Visualising Rhyme

Visualisation provides an effective way to illustrate both rhyme type and position. Beyond serving as an explanatory aid, it also supports downstream computational tasks, such as evaluating phonemic differences in multi-syllabic and cross-rhyme structures. By mapping rhymes spatially, it becomes easier to highlight overlaps, near matches, and layered rhyme schemes that are otherwise difficult to parse in raw text. Table B.5 demonstrates how different rhyme categories may be rendered visually.

Name	Example
<b>Perfect</b>	rose, nose skylight, highlight
<b>Slant</b>	hat, cut slack, slap
<b>Assonance</b>	hat, bad crate, braid created, rabid
<b>Consonance</b>	cut, mat poncho, crunchy
<b>Multi-syllabic</b>	I've got a <b>bad taste</b> It gives me <b>mad haste</b>
<b>End of line</b>	It's too hot to <b>handle</b> Ooh got you new spots to <b>vandal</b>
<b>Internal</b>	It's <b>too hot to handle</b> , <b>you got blue sandals</b>
<b>Cross</b>	It's <b>too hot to handle</b> , <b>you got blue sandals</b> Who <b>shot you?</b> Ooh got you new <b>spots to vandal?</b>

Table B.5: Visual Display of Rhyme Types and Positions

## B.2 Rhyme Density Method and Worked Example

This appendix outlines the procedure used in Study II to compute *rhyme density* from phoneme similarity. The pipeline defines a phoneme inventory, embeds phonemes into a vector space, measures cosine similarity between phonemes, thresholds those similarities to detect rhyme matches, and aggregates matches into a density score.

### Phoneme Inventory

Let  $P$  denote the set of phonemes, partitioned into vowels  $P_V$  and consonants  $P_C$ :

$$\begin{aligned} P_V &= \{AA, AE, AH, AO, AW, AY, EH, ER, EY, IH, IY, OW, OY, UH, UW\}, \\ P_C &= \{B, CH, D, DH, F, G, HH, JH, K, L, M, N, NG, P, R, S, SH, T, TH, V, W, Y, Z, ZH\}. \end{aligned} \quad (\text{B.1})$$

### Cosine Similarity and Rhyme Density

Similarity between two phonemes  $p_1, p_2 \in P$  is measured using cosine similarity:

$$\text{sim}(p_1, p_2) = \frac{\mathbf{v}_{p_1} \cdot \mathbf{v}_{p_2}}{\|\mathbf{v}_{p_1}\| \|\mathbf{v}_{p_2}\|}. \quad (\text{B.2})$$

For a sequence of terminal phonemes  $\mathcal{T} = \langle p_1, \dots, p_n \rangle$ , rhyme matches are identified using a threshold  $\theta$ :

$$M_{ij} = \begin{cases} 1, & \text{if } \text{sim}(p_i, p_j) \geq \theta, \\ 0, & \text{otherwise.} \end{cases} \quad (\text{B.3})$$

Rhyme density  $D$  is then:

$$D = \frac{1}{n} \sum_{1 \leq i < j \leq n} M_{ij}, \quad (\text{B.4})$$

where  $n$  is the number of terminal phonemes.

### Worked Example

Consider the stanza:

*Roses are red*  
*Violets are blue*  
*Sugar is sweet*  
*And so are you*

End-word phonemes are: red (/D/), blue (/UW/), sweet (/IY T/), you (/UW/). At  $\theta = 0.9$ , both “blue/you” and “sweet/you” are detected as rhymes, yielding two matches over four positions:  $D = \frac{2}{4} = 0.5$ .

This illustrates how phoneme similarity produces an interpretable density score. The same computation was applied systematically to all outputs in Study II.

### Notes on Scope

Rhyme is not a precise category but a spectrum of phonemic resemblance. Threshold choice ( $\theta = 0.8$  in the study) reflects a balance between overfitting (capturing spurious matches) and underfitting (missing subtle or culturally recognised rhymes). More elaborate analyses—such as rhyme clustering, multisyllabic “multis,” or cross-line weighting—are possible [274, 205, ?] but were beyond the scope of Study II. These are considered further in Study III.

## Study II Participant Questionnaire (Verbatim Responses)

This appendix reproduces participants' free-text answers verbatim (lightly LaTeX-escaped) to the question: *"Did this tool help you write better than without tools? Please answer yes/no and give reasons."* Responses are grouped by tool type to mirror the Study III appendix structure. Identifiers use the form S2-P#.

### Support Tool

*Prompt:* Did this tool help you write better than without tools? Respondents answered **yes/no** with reasons.

#### S2-P1

Yes- it helped me find better ways to form my words in the poem but I also think that it forced a couple rhymes that might not fully flow

#### S2-P2

not really, I prefer to write from my brain as it comes into my head because I prefer the flow, so it became more of a case of trying to fit the support in somewhere instead of it naturally flowing

#### S2-P3

Yes - more speed in choosing rhyming words. Some familiarity with the tool helped.

#### S2-P4

N. I knew what a haiku was already.

#### S2-P5

Yes, it offered more options

#### S2-P6

No

**S2-P7**

y - The support tools help with word choice, cadence and enable the editing process. They say that good songs are written and great songs are re-written. So although this was an academic exercise, it is clear that taking a beat to understand the impact of the syllables and rhymes with support tools is extremely helpful.

**S2-P8**

I felt that it was most useful in helping with a specific rhyme schemes especially ones that values cadence and the intricate use of the same rhyme throughout the verse. Through this it allows you to write and expand on a certain rhyme with ease. Allowing you to concentrated on context and the story of the lines/

**S2-P9**

not really - a bit helpful but not that different from using a good thesaurus

**S2-P10**

Yes it gave more initial prompts and ideas than having to self generate them myself. I would likely use ChatGPT over the other tools due to ease of availability and discovery, and suitability for mobile.

**S2-P11**

Yes. More ideas and options.

**S2-P12**

I used the Poetry Analysis Engine and Rymezone. I decided to try to write in one of the forms supported by the PAE - rondelet - and used the engine as a validator to help achieve that form. This led to a much shorter poem compared to being left to write as much as I wanted. Perhaps there are advantages in being forced to get a message across using less words. But the length restriction made it much harder to incorporate the 10 divergent words. I found the ability to use Rymezone helpful and may not have come up with all the rhymes (dream/gleam/scheme/stream and city/gritty/pretty) on my own. I think the poem with support tool is not so much better as different. These tools enabled me to work faster. I spent 40 minutes on this exercise, compared to over 2 hours on the first poem when I was left entirely to my own devices.

**S2-P13**

Helped a little bit but not much - mostly I was hampered by the words I had chosen initially!

**S2-P14**

Yes because it can shape out your writing to be exact.

**S2-P15**

n, I'm not really sure what the feedback meant

**S2-P16**

I personally didn't think it helped, it was a weird tool

**S2-P17**

NO I found it difficult to use the support tool as i do not have much experience writing in formats other than reports

**S2-P18**

When trying to rhyme, I only remembered up to 5 or 6 words that rhyme, whereas rhymezone returned all possible rhymes, eye rhymes, half rhymes and even example of poems that use those words.

**S2-P19**

Y. I am not very good with writing

**S2-P20**

Yes, it helped making better structure flow and decisions

**S2-P21**

Some did. Rhyming glow with volcano was better than elbow

**S2-P22**

It helped with finding rhymes but did not improve my writing a lot

The poetry analyse tool wasn't great as it couldn't recognize any of my poetry. But it was good for syllable counting and seeing how the rhymes matched. The other tools were good for finding words that rhymed. Some of the words from my initial 10 were difficult to rhyme and the tools made that easier to spot. The ability to suggest near rhymes was also good as although this exercise asked for rhyming couplets if I were to write anything going forward they would be helpful.

**S2-P24**

Yes, it helped in Enhanced Creativity, Inspiration and Prompts, Rapid Generation, Generating Complex Patterns

**S2-P25**

It just helped me find synonym words easier

**S2-P26**

Yes, as the autonomic system instead choose to write about the topic it choose randomly.

**S2-P27**

yes, they suggest to modify some words and structure.

**S2-P28**

No, not really. Because I don't know how write a poem in English or what a poem looks like.

**S2-P29**

No, I think I have a decent enough vocabulary and ability to come up with words that rhyme that it didn't make much of a difference in my opinion. It could help me find more options for rhyming words quicker though, so it wasn't completely useless.

### **Co-Creative Tool**

*Prompt:* Did this tool help you write better than without tools? Respondents answered **yes/no** with reasons.

#### **S2-P1**

No- I think it made it a lot more difficult to write. I felt as though I was pushing against a lot of different styles of writing and it ended up disrupting the flow of the poem.

#### **S2-P2**

no, similar reasons. I didn't feel that it flowed quite as well as it does straight from my head

#### **S2-P3**

No - constrained by style of rhymes offered - limited to US poets and no allowance for personal style. May not have used it correctly?

#### **S2-P4**

Y. It's a good way to easily mimic the style of another poet. But I feel no ownership, I can't call it my own at all.

#### **S2-P5**

It enriched ideas

#### **S2-P7**

n - the verse created by Deepbeat, in my view, was incoherent

#### **S2-P8**

I Feel that with this I was unable to come out with a verse that felt whole. Its best use I found was picking out lines that you found interesting , but there was not much flexibility in adding your own edits to these lines or adding in your own verse. In this way it felt more like how can I choose the right pieces of the puzzle in order to fit the whole in stead of deciding the rules of the. puzzle my self.

**S2-P9**

yes definitely - prompted it to become a bit more creative strangely

**S2-P10**

Yes for a novice like me it was useful but I would gravitate towards using a ChatGPT solution more

**S2-P11**

No. Suggestions did not meet the concept.

**S2-P12**

Verse by Verse expanded my horizons. It introduced creative ideas that I would not have thought of on my own. I couldn't copy and paste its complete verse as it wasn't coherent enough and didn't meet the artistic goals. It gave ideas and short phrases I could use. I think I could write more creative verse with this tool. I spent 20 minutes on this exercise.

**S2-P13**

Yes, I think the co-creative tool can help - but often because it then prompts you to think of another word rather than providing a word to use.

**S2-P14**

Yes, it will let you decide which materials you will use.

**S2-P15**

n- Seemed to be more in the realm of plagiarism

**S2-P16**

I didn't think it was great

**S2-P17**

NO it was useful but a little difficult to use

**S2-P18**

It was helpful but often did not have the same story as me. The ideas were nice, and the writing was much better, but the novelty and unique style was missing

It gave me ideas

**S2-P20**

Yes, it helped making better flow, structure and meaning to the writing

**S2-P21**

I think it was better.

**S2-P22**

Yes, but similar to previous question - not improved my writing drastically

**S2-P23**

N The deepbeats did not work so I didn't use it. The other tool I found very difficult as my poems were more literal for this exercise and the suggestions were more abstract and on topics not fully related to topic. It also had a very narrow range of suggestions and would also recycle lines that didn't necessarily fit in anyway to any of the previous lines. It was more like it had picked up on the rhyming sound needed and came up with any line that would fit that in the selected poet(s) style. I found it more difficult to write and almost impossible to include any of the 10 words.

**S2-P24**

Enhanced Creativity, Inspiration and Prompts, Rapid Generation, Generating Complex Patterns

**S2-P25**

It just limits my imagination and many times it couldn't find the right rhyme

**S2-P26**

Yes, to an extent. The problem was that I couldn't choose the topic of the poem. It just wrote random sentences based on the words I gave it

**S2-P27**

Yes, they provide new approach for my poetry.

**S2-P28**

Yes, to some extent. Because I don't know how to write a poem or what a poem looks like.

**S2-P29**

For this specific task I found these tools pretty useless, as they often didn't make it easy to incorporate the words I needed to use, though I found it interesting I could pick different poets as muses. If I didn't have to incorporate the specific list of words though, I could see it being useful in helping me come up with more eloquent or evocative phrases, especially for certain styles of writing.

**Autonomous Tool**

*Prompt:* Did this tool help you write better than without tools? Respondents answered **yes/no** with reasons.

**S2-P1**

No- it felt impersonal and clunky and hardly felt like I had written poem/verse

**S2-P2**

The poem appeared pretty much instantly which was helpful, although I would've had to make a few adjustments had I been submitting it into a competition for example because it didn't make much sense in some places

**S2-P3**

Yes - but I used it incorrectly so didn't get full impact of its creative potential.

**S2-P4**

Y/N. I feel no ownership or emotional connection with the finished piece. It's not mine. However, the creation of format, metre and rhyme was effortless.

**S2-P5**

No, I felt it was glib and derivative

**S2-P6**

No

**S2-P7**

n - Koala did do a little bit better in creating a coherent story, however, both versions lacked any emotional connection and in the second version the desire to rhyme seems to be the driving factor for the use of some words which do not help to build the story. in a song with such limited real-estate, I would never waste 3 words like 'just my way' for the sake of a rhyme.

**S2-P8**

Although the poem which I wrote took on a different which was less based in rhyme, I feel this tool was the one that ultimately cultivated my creativity the most, which is probably what I was subconsciously looking for. The ability to come up with the concepts myself, then have a discourse that opened new possibilities that I could freely select elements from reminded me the most of experiences I have had when doing collaborative writing with other people.

**S2-P9**

no - barely used the autonomous tool. it did crack out a poem but i didn't find it that interesting

**S2-P10**

It lacked originality in the verse but produced the most comprehensive result with the least prompting

**S2-P11**

Tricky to give straight yes or no. Overall, yes. Note: I did not edit or amend the submission. It fulfilled the brief, even though I found it a tad boring.

**S2-P12**

I used ChatGPT and instructed it to write a poem according to the exercise goals, including the list of 10 words it should include. I didn't ask it to include how I feel about the area I live or what it means to me as I felt it was unfair to expect it to know how I feel. It didn't help me write, rather it wrote for me. I was no longer a writer but an editor. There were a few phrases that were nonsensical that I corrected, otherwise its poem was 95% fine as it was. What I should have done, given more time, was more radically edit to incorporate my feelings in the output, or fed this information in the initial instruction. The question "How useful did you find the autonomous tool that you used?" depends on useful for what? If to help me write, I would rank very low as I hardly did any writing! I felt annoyed that it produced a poem in a few seconds that is arguably better than the one I spent over 2 hours on earlier. However, it's like fast food. I feel far more satisfied by my own poem because I expended some effort and felt I actually created something on my own. ChatGPT's poem is superficially interesting but very unsatisfying.

**S2-P13**

I was impressed how quickly the poems were achieved - but not particularly by what they created!

**S2-P14**

No, according to AUTO, it will decrease your creativity.

**S2-P15**

y, I'm not very good at writing

**S2-P16**

I found chat gpt massively helpful and easy to create a good poem out of the words

**S2-P17**

yes It gave some great ideas it was very useful to use as a base to work on and edit

**S2-P18**

Much better than anything I wrote but again, was more generic, lacked human uniqueness and creativity

**S2-P19**

Y It makes writing easy

**S2-P20**

Yes, it helped making better flow, structure and meaning to the writing

**S2-P21**

It was less accurate but smoother

**S2-P22**

It created a different piece using chosen words. Tool used them to create independent piece which objectively was better than my own words.

**S2-P23**

Y It produced a poem really quickly that contained all of the words and on the theme of where I live. The first couple of lines worked well however, it started to fall apart towards the end where the algorithm forced the prompted words to fit. I thought it was quite funny so I didn't change anything. However, the creative process was lost and I didn't really write anything more than a sentence the algorithm did everything else.

**S2-P24**

Enhanced Creativity, Inspiration and Prompts, Rapid Generation, Generating Complex Patterns

**S2-P25**

In my opinion and experience the poem generated by ChatGPT was pretty good. The flow of the poem as well as the rhyming of the words was excellent.

**S2-P26**

Yes, the poem was fluent and creative. But like the last tool this too didn't have an option to select the topic, instead it wrote the poem based on words I had inputted.

**S2-P27**

not really help much, seems like structure is same.

**S2-P28**

Yes, Because I don't know how write a poem or what a poem looks like.

**S2-P29**

I definitely think these were some of the better verses, and they took the least amount of time to come up with, though I am not really sure I can give myself much credit for the writing since the tool did almost all of the work. I suppose it depends how you pose the question and how collaborative you attempt to make it with the AI. I would say it was definitely useful, though it also raises some other questions or concerns.

## Study II: Representative Poem Samples (Base Condition)

This appendix complements the Study II analyses by allowing readers to qualitatively inspect extremes and a typical instance of the base writing condition, which the thesis analyses quantitatively. The Rhyme Density proxy (`base_rscore`) for each each participant’s *base* poem (no support tools) was used to select three exemplars: the highest (“Best”), the lowest non-zero (“Worst”), and a mid-range “Random” sample. Texts are reproduced *verbatim* from the dataset; no wording has been changed.

Slot	Participant	ID	base_rscore
Best	Barbara	12	9.77
Worst	David 1	13	0.43
Random	Julia	7	3.56

Table B.6: Study II base poems chosen for qualitative inspection.

### **Highest by Rhyme Score**

Barbara (ID 12), base\_rscore = 9.77

1. The piano lies under the kneck, of the  
Giraffe which leans to the left  
In the warehouse on Overbury Road  
Which leaves me occasionally bereft  
But at night when I wander around  
In my pyjamas amid crepuscular sound,  
I know that it works, despite pandemonium mode,  
This warehouse on Overbury Road.
  
2. One is for the giraffe  
Two for the pianos  
Three for the arches  
And four for the sofas  
Five for the lights which hang in no order  
And six for the desks which are piled by the hoarders...  
Seven for the shelves which groan under weight  
Of the plates bought with serendipitous state  
Eight for the geraniums dying outside  
Nine for the bicycles packed side by side  
And ten for the inhabitants living in general harmony  
(... But not always!)
  
3. I'll tell you a tale of a vegan soup  
Of a soup which was made for the ten  
Of a cockaleekie, potato leekie, stock and streakie, mock and meekie  
All and greetie, all were waxing lyrical and physical.  
The time was set for eight o'clock,  
When all would stop.  
The scene was set by the giraffe  
The bowls were out, the goals were set

The pan was full, the gas was lit  
And the veg was prepped:  
Whatever serendipitous veg was to hand from the fridge  
There was onion and garlic chopped fine on the board  
There were leeks and potatoes and cabbage and more  
There were carrots and courgettes and celeriac galore  
And all were chopped and diced and cooked in the pot for the  
Soup, the serendipitous soup,  
The pandemonium soup of the  
Warehouse, the bearhouse, but always a sharehouse  
The cockaleekie supper of TEN!

### **Lowest by Rhyme Score**

David (ID 13),  $\text{base}_r\text{score} = 0.43$

We are all just darts in a board,  
wishing we had hit the bull's eye.

### **Mid-range by Rhyme Score**

Julia (ID 7),  $\text{base}_r\text{score} = 3.56$

The house that made us  
Who we were destined to become  
Was built from red bricks  
That glowed in the sun  
We played karate kid in the yard  
Running circles round each other  
Arranging weeds in a vase  
As a gift for our mother

Used a stubby pencil  
To etch the date on the wall  
That always indicated  
We were getting so tall  
Grew out of mud pie games  
Wrote some songs about it  
The light bulb went on  
we knew we'd outgrown it



## Appendix C

# Study III; Supplementary Materials and Analyses

### C.1 Extended Rhyme Analysis

Unlike Study II, which used aggregate *rhyme density* across end words to compare texts (Appendix B.2), Study III extends the analysis to *performance-oriented conditions*. The detector is designed to be (i) **multi-syllable aware**, rewarding longer aligned nuclei and codas, (ii) **flexible to rhyme type**, capturing assonance, consonance, and imperfect matches in addition to exact rhymes, and (iii) **placement-sensitive**, scoring both internal and end rhymes at the line level. While still based on phonemic analysis of text, this approach better reflects features of live cadence and delivery, without requiring beat alignment or audio-layered analysis.

#### Core Detection Model

The detector extends Study II by adding the elements:

1. **Distance-aware similarity.** Rhyme strength decreases as lines get further apart, implemented with an exponential decay that downweights distant links.
2. **Multi-syllable scoring.** Longer spans of aligned phonemes (“multis”) are rewarded more highly than single-syllable matches, but with bounded weights to avoid uncontrolled growth.

Aspect	Study II	Study III
Target	Page-level comparability	Stage/line placement and responsiveness
Signal	Terminal nuclei similarity	Internal <i>and</i> end matches
Placement	None	Distance window + exponential decay
Span	Single nucleus	Multi-syllable with bounded length weight
Output	Single score <i>D</i>	Line scores (averaged across passage)
Use	Tool evaluation	Performance evaluation (CADT)

Table C.1: Key methodological differences between Study II and Study III rhyme analysis.

3. **Line-level aggregation.** Each line receives a rhyme score combining internal and cross-line matches. Passage-level results are reported as averages across lines.

### Worked Example: Eminem – *Lose Yourself* (L1–L3)

L1: His **palms** are **sweaty**, **knees** weak, **arms** are **heavy**

L2: There's vomit on his sweater **already**, mom's **spaghetti**

L3: He's **nervous**, but on the **surface** he looks calm and **ready**

### Multi-syllable rhyme families identified:

1. **-etty / -eady family (dominant chain, 2-syllable, cross-line)**

*sweetty* (L1) ↔ *already* (L2) ↔ *spaghetti* (L2) ↔ *ready* (L3).

Forms the backbone of the passage, linking all three lines with a rolling /ti/ sound.

2. **-alm / -om family (assonance + consonance, 1-syllable, cross-line)**

*palms* (L1) ↔ *mom's* (L2) ↔ *calm* (L3).

Anchors the flow with repeated /m/, creating a heavy percussive feel.

3. **/i/ vowel family (short echoes, internal + cross-line)**

*knees* (L1) ↔ *weak* (L1) ↔ *He's* (L3).

Quick vowel repetitions contribute to bounce and rhythmic drive.

4. **-erv / -urf family (slant rhyme, internal within L3)**

*nervous* ↔ *surface*.

Illustrates the use of slant multis inside a line to build texture.

Rather than a single end-rhyme scheme, these lines interweave multiple overlapping rhyme families. Eminem stacks multis at different positions (end, internal, cross-line), balancing exact rhymes (*spaghetti / ready*) with slants (*nervous / surface*). This layered structure exemplifies the kinds of placement-sensitive and multi-syllable patterns targeted in Study III.

### Worked Example: Eminem – *Lose Yourself* (L1–L3)

L1: His **palms** are **sweaty**, **knees** weak, **arms** are **heavy**

L2: There's vomit on his sweater **already**, mom's **spaghetti**

L3: He's **nervous**, but on the **surface** he looks calm and **ready**

#### Colour legend:

- **Red** = **-etty / -eady family** (*sweaty, already, spaghetti, ready*)
- **Blue** = **-alm / -om family** (*palms, arms, mom's*)
- **Green** = **/i/ vowel + slant family** (*knees, He's, nervous, surface*)

The colour-coded display highlights overlapping rhyme families across lines. Eminem layers multiple multis — the dominant **-etty chain**, percussive **-alm consonants**, and slant/internal **/i/ vowels** — creating rhythmic density and listener surprise. This demonstrates the placement-sensitive, multi-syllable rhyme structures targeted in Study III.

## C.2 Speech Systems Taxonomy

This appendix expands on the taxonomy of speech interaction systems referenced in Study III (section 5.5). The classification highlights different architectural choices for combining speech recognition and synthesis, ranging from asynchronous uni-modal pipelines to fully integrated, real-time multimodal systems. Placing the Study III prototype within this taxonomy makes clear both its strengths (modularity, cadence control, interpretability) and its limitations (latency, lack of live responsiveness). It also indicates a trajectory for future work: moving toward synchronous and multimodal configurations that could support fully live human–AI freestyle battles.

Modality	Type	Timing	Examples / Implications for Study III
Uni-modal	Type 1	Async STT / TTS	<b>Current prototype:</b> Whisper for transcription; ElevenLabs for synthesis. Cadence control + post-hoc analysis.
	Type 2	Sync STT / TTS	Sequential pipeline; supports near-live interaction. Next-stage candidate.
	Type 3	Real-time STT / TTS	E.g., Alexa, Google Assistant. Low-latency but complex. Ideal for future live AI battles.
Multi-modal	Type 1	Async STS	E.g., SpeechGPT. Preserves flow via speech-to-speech mapping.
	Type 2	Real-time STS	E.g., Moshi, EMOVA. Full expressive agents with prosody. Future goal.

Table C.2: Taxonomy of speech interaction systems by modality and timing. Study III uses a Uni-modal, Type 1 system: asynchronous STT/TTS for modularity and cadence control.

## Appendix D

# Study III: Further Experimental Data

This appendix complements the rhyme analysis by specifying how cadence and flow are treated in Study III. Earlier analysis quantifies rhyme via phoneme relations, here it is described through stress placement, timing against the metrical grid, and performance features (articulation, pitch contour) that shape listeners' perception of structure. These features inform the evaluation frame in Chapter 5 and the CADT, but are not folded into the rhyme score.

### Worked Example of Flow

This example complements the cadence analysis in Chapter 5 and connects back to the rhyme measures in Appendix B.2. Using *Drop* by The Pharcyde, it illustrates how beat-aligned stress interacts with lyrical timing and why stresses cannot be inferred reliably without the instrumental context.

**Let me freak the funk, obsolete is the punk that talk  
More junk than Sanford sells. I jet propel at a  
Rate that complicate their mental state as I invade  
Their masquerade they couldn't fade with a clipper blade**

In ordinary speech, lexical stress would likely fall as:

obsolete is the **punk** that **talks more** junk than **Sanford sells**.

In performance, stresses shift to align with the bar's four beats:

obsolete is the **punk** that talks more **junk** than **Sanford sells**.

#### D. Study III: Further Experimental Data

---

The key point is that beat-aligned emphasis governs timing: placing stressed syllables on the four beats keeps lyrics in time even when lexical stress would suggest otherwise.

**Let** me freak the **funk,** obso **lete** is the **punk** that talks  
**more** junk than **Sanford** sells I jet pro **pel** at a

Figure D.1: Beat-aligned stress over two bars: primary beat placements shown across the line.

A deliberate rest on beat three of bar two (no syllable on the downbeat while the instrumental hits) creates a dramatic hole in the grid:

1	2	3	4
<b>Let</b> me freak the	<b>funk,</b> obso	<b>ete</b> is the	<b>punk</b> that talks
<b>more</b> junk than	<b>Sanford</b> sells	■ I jet pro	<b>pel</b> at a

Figure D.2: Bar-level rest on beat three: the instrumental articulates the beat while the vocal line pauses.

This worked example motivates why cadence measures in Chapter 5 go beyond page-based rhyme to include on-beat timing and call–response placement; and why Study II’s text-only proxies (Appendix B.2) are complemented by performance-aware analysis in Study III.

### Two Analytic Frames for Flow

Two established approaches frame the cadence annotations.

#### Adams (bar-structural phrasing)

Focuses on where phrases and accents fall relative to the bar: bar-aligned clauses, enjambment across bars, placement of internal rhymes on strong/mid-bar positions, and articulative contrast (staccato vs. legato) as rhythmic markers.

---

## Condit-Schultz (corpus timing)

Encodes syllable onsets against the metrical grid, reporting densities (syllables/sec), inter-accent intervals, syncopation, and beat-level occupation. Delivery dynamics are abstracted out to preserve timing comparability across artists.

Feature	Adams — Example and Interpretation	Condit-Schultz — Example and Interpretation
<b>Rhythm</b>	<i>Eminem</i> : “His palms are sweaty, knees weak, arms are heavy.” Evenly spaced 16ths; accents fall predictably on/around beats; metrically stable with light syncopation; phrases align to bar structure.	Same line measured at $\sim 4.4$ syll/s (near corpus mean). Beat-aligned syllable onsets support density/timing comparison across MCs; stress can be annotated against the grid.
<b>Rhyme</b>	<i>MF DOOM</i> : “Close but no krills, toast for po’ nils, post no bills.” Dense internal/multis; rhyme hits on strong beats and mid-bar positions, shaping groove and bar-level momentum.	Corpus tags internal rhymes; clusters vowel-consonant patterns. “krills/nils/bills” = single-bar rhyme chain; high rhyme density ( $\sim 3$ rhymes/bar).
<b>Syntax</b>	<i>Eminem</i> : “He’s nervous, but on the surface he looks calm and ready” = bar-aligned clause. “To drop bombs / but he keeps on forgetting” = enjambment across two bars.	Phrase boundaries aligned to bar-lines; enjambment rate computed. Majority of lines here are intra-bar; “drop bombs / ... forgetting” flagged as cross-bar continuation.
<b>Articulation</b>	<i>MF DOOM</i> : “It’s too hot to handle, you got blue sandals.” Compressed timing; punchy consonants; staccato articulation treated as a signature aspect of flow.	Rhythmic grid does not encode dynamics (punch/softness). Timing abstraction treats staccato and legato similarly; articulatory nuance is not represented.

Table D.1: Illustrative contrasts in rap flow analysis: Adams (bar/phrase, accent, articulation) vs. Condit-Schultz (grid-based timing, density, corpus comparability), using lines from Eminem and MF DOOM.

### Global Flow Features (Ohriner)

The table below adapts Ohriner's 'global' flow descriptor as a checklist for qualitative comparison.

Flow Category	Global Features
<b>Delivery</b>	Syllables per word; Syllables per second; Phrase length; Variation in phrase length Proportion of positions occupied in the measure; Line-to-line variation in this proportion; Position of line or phrase ending in the measure.
<b>Rhyme</b>	Number of instances in rhyme classes; Number of syllables in rhyme instances Overlapping of rhyme classes; Rhyme scheme complexity; Variation in quantities of IRIs at the level of the beat; Variation in IRIs below the level of the beat.
<b>Rhythm</b>	Tempo Patterning of inter-accent intervals (i.e., groove); <i>Exactness</i> of the patterning; <i>Typicality</i> of the patterning; <i>Evenness</i> of inter-accent intervals; Extent of syncopation; Extent of 1mod2 accents.

Table D.2: Global features of flow (adapted from Ohriner, 2023, p. 30)

## Rhyme Strengthening through Pitch in Tupac's *California Love*

This section illustrates how relative vocal pitch can reinforce rhyme structure across spans that might otherwise exceed listeners' perceptual window. The analysis aligns with the performance-sensitive evaluation in Chapter 5.

The image shows three lines of musical notation for Tupac's 'California Love'. The first line contains the lyrics: 'Fa - mous be - cause we pro - gram, world - wide let 'em rec - og - nize from Long Beach to Rose - crans, bump - in' and'. The second line contains: 'grind - in' like a slow jam, it's West Side so you know the row won't bow down to no man. Say what you'. The third line contains: 'say, but give me that bomb. beat from Dre, let me ser - en - ade the streets of L. A. from Oak - land to Sac - town... (etc.)'. Colored boxes highlight specific words and phrases: green boxes highlight 'pro - gram', 'Rose - crans', and 'Oak - land'; purple boxes highlight 'say', 'say,', and 'Dre'. The notation includes a treble clef, a 4/4 time signature, and various rhythmic markings like slurs and accents.

Figure D.3: Tupac Example. A colour-coded transcription highlighting Tupac's delivery style, which integrates rhythmic declamation, regionally inflected phrasing, and lyrical emphasis. The example illustrates the layering of pitch accents and groove elements along the speech–song continuum discussed in the Rap Pitch Model (Figure 2.8).

Relative pitch can reinforce rhyme perception across spans longer than the typical “rhyme window.” In *California Love* (2Pac), chains on *program* (green) sit in a consistent pitch register; a later chain initiated by *say* (purple) concludes on the same pitch level as the earlier group, creating a retrospective link despite different lexical content. The opening passage forms a “program” rhyme chain (highlighted in green) delivered within a consistent pitch register. A later chain (purple) begins with *say*, shifts contour, but resolves the word *Oakland* at the earlier pitch level. This pitch recall retrospectively links the two groups, strengthening the perceived coherence of the verse even though the rhymes are temporally separated.

Ohriner's account of an “upper limit” for rhyme perception suggests that dense verses challenge recognition over long spans; here, pitch functions as mnemonic scaffolding that extends rhyme salience. In practical terms, this motivates tracking not only rhyme classes but also coarse pitch zones in performance analyses, as adopted in Chapter 5.

## D.1 Overview of Study III Experiments

This section outlines seven experiments that evaluate Digital Creative Twins (DCTs) across voice, text, and integrated performance.

ID	Modality	Aim (what is tested)	Stimuli / Design	Key Measures
E1	Voice	Clone similarity (timbre/identity)	28 A/B trials (2 clips/trial) from one TTS; native speaker references	MOS speaker-similarity; speaker-embedding distance; error profiling by phoneme class
E2	Voice	Flow capture across TTS models	5 TTS systems; 10 items×9 raters; +29 corpus items (computational)	Tempo (BPM); syllables/sec; mean syllable duration; onset timing DTW <i>alignment cost</i> ; rhyme overlap; human flow rating
E3	Voice	Cadence / pitch consistency	Paired human vs. TTS clips (matched text)	F0 contour DTW cost; pitch range/variance; register usage; (MFCC distance reported as timbre control)
E4	Text	Style & rhyme quality	Prompted generations; standard beam vs. Creative Beam Search (CBS)	Rhyme density (Study III); cosine-/multisyllable variants; LLM-as-judge (pairwise preference + justification); inter-rater reliability
E5	Integrated (Poets)	Poetic co-creation (TTS-only vs. hybrid remix)	$N \geq 10$ poets; within-subjects; two tests/participant	ASR WER; rhyme density; cadence proxies; post-task UX; speaker-similarity & human clone ratings
E6	Integrated (MCs)	DCT flow representation	$N=4$ freestyle MCs	Cadence metrics (onset-grid occupancy, syncopation ratio, syllables/beat); participant ratings
E7	Perception	Identity & flow judgment	13 MC identities (real+synthetic) + listener panel	AB flow preference; MOS for flow; human-vs-AI forced-choice identification

Table D.3: Condensed overview of Study III experiments. *Abbrev.* MOS: mean opinion score; MFCC: mel-frequency cepstral coefficients; DTW: dynamic time warping; ASR: automatic speech recognition; WER: word error rate; TTS: text-to-speech; LLM: large language model.

## D.2 E1a (Exploratory): Voice identity checks across conditions

Pre-study voice-identity checks across impersonation, cloning, language, health, and singing revealed mixed tool outcomes even when humans could decide the match. This fragility motivates cadence/rhyme metrics and listener studies in E1–E7 (not voice ID alone).

Test Case	Exp.	Actual	Tools: Pass	Tools: Fail / Mod.	Diff.
Human impersonator of recognisable voice vs actual voice	F	Mixed	T1, T2, T4a	T3, T4b, T4c	2
AI clone of recognisable voice vs actual voice	F	Mixed	T1, T2, T3, T4a	T4b, T4c	3
AI clone of human voice vs actual voice	F	Mixed	T1, T2, T3, T4a	T4b, T4c	3
Human impersonator vs impersonator’s actual voice	P	Mixed	T1, T3, T4a	T2, T4b, T4c	3
Human (actor) natural voice vs character/accnt	P	Mixed	T1, T3, T4b, T4c	T2, T4a	3
Human voice (normal) vs with background noise	P	Mostly	T1, T2, T3, T4b, T4c	T4a	2
Health – normal vs heavy cold	P	Mixed	T1, T3, T4b	T2, T4a; <i>Mod.:</i> T4c	3
Bilingual – language 1 vs language 2	P	Mixed	T1, T4a, T4b, T4c	T2, T3	4
Speech vs singing	P	Mixed	T1, T3, T4a, T4b	T2, T4c	2
Singing language 1 vs language 2 (Test A)	P	Mixed	T1, T2, T3, T4b	T4a, T4c	4
Singing language 1 vs language 2 (Test B)	P	Mixed	T1, T3, T4b	T2, T4a; <i>Mod.:</i> T4c	5
Age – time 1 vs time 2	P	Mixed	T1, T2, T3, T4b	T4a, T4c	3
Age/Activity – talking vs singing (Test A)	P	Mixed	T1, T3, T4a	T2, T4b, T4c	4
Age/Activity – talking vs singing (Test B)	P	Mixed	T1, T3	T2, T4a, T4c; <i>Mod.:</i> T4b	3
Singing badly vs singing well	P	Mixed	T1, T3, T4a	T2, T4c; <i>Mod.:</i> T4b	5

Table D.4: Summary of voice identity tests comparing different speaker conditions across impersonation, AI cloning, language, health, activity, and age variation. Each test compared two voice recordings. The system was expected to either match or reject identity based on speaker similarity. Mixed results highlight the challenges of robust voice identification across real-world variation. *Note:* Tools T1–T4 (and T4a–c) are proprietary matchers; outcomes are reported descriptively rather than as endorsements of tool accuracy.

### D.3 Experiment E1: Voice Clone Identity and Similarity

**Aim.** Assess whether a cloned voice sounds like the participant (identity/timbre impression; optional flow impression).

**Stimuli & Design.** Each participant recorded the voice script (§D.10). A single TTS clone was trained per participant. The system then synthesised the script recording in the cloned voice for rating.

**Measure.** Self-rating on a 1–5 Likert scale on the quality of the voice clone likeness relative to the the speaker’s voice.

**Procedure.** Record script → train clone → play cloned script → collect self-ratings (identity). If rating < 1, retrain/tune and re-test; otherwise approve for downstream use.

#### E1 measures at a glance

Metric	What it captures / how it’s computed	Direction
Self-rated clone similarity	Participant hears a cloned version of the script they recorded; rates 1–5 “How good is the cloned version of the script?” (timbre/identity).	↑ better
Pass threshold	None required as lowest rating always = 1.	—

#### Voice Clone Recording Script (excerpt)

Good morning. This is a sample script for a speaking voice recording.  
The quick brown fox jumps over the lazy dog. . . .  
Please, pause for a second. Breathe in. Breathe out. . . .  
This is the end of the script. Thank you for your voice.

## D.4 Experiment E2: Cadence Similarity across TTS Models

**Explanation.** Five TTS engines were benchmarked to select a default for Study III that preserves *both* voice likeness and rhythmic flow. Nine raters scored ten synthetic outputs (2 per engine; scripted plus freestyle) on 1–5 MOS for *voice quality* and *cadence fidelity*; additionally, MFCC distance and DTW alignment were computed on a further 29 pairs.

Engine-level MOS (Fig. D.4) shows ElevenLabs leading on both dimensions; clone-level breakdowns (Fig. D.5) confirm this pattern and expose within-engine variance. Cadence-focused plots (DTW; Fig. D.6) indicate tighter temporal alignment for ElevenLabs, while the listener scatter (Fig. D.7) mirrors this: clones with lower DTW offsets also earn higher cadence MOS.

**Stimuli & Design.** Five engines (Dia, ElevenLabs, Silero, SpeechBrain, Tacotron2). Ten items (2 per engine) across 9 human raters; Additionally 29 items for computational tests.

**Decision.** ElevenLabs selected for Study III generation based on converging subjective (MOS) and objective (DTW/MFCC) evidence.

E2 measures at a glance		
Metric	What it captures / how computed	Direction
Listener MOS: <i>Voice quality</i> (1–5)	“How much does the clone sound like the original speaker?”	↑ better
Listener MOS: <i>Cadence fidelity</i> (1–5)	Perceived rhythmic alignment / delivery naturalness.	↑ better
MFCC MSE	Spectral distance (clone vs. reference).	↓ better
DTW time offset	Temporal misalignment of onsets/phrases; Sakoe–Chiba $\pm 10$ frames (10 ms).	↓ better
Cadence features	BPM tracking; syllables/line; mean syllable duration; rhyme overlap.	— (reported)

### Selected Results

Table D.5: Listener means by engine.

Engine	Voice Quality	Cadence Fidelity
Dia	1.44	1.27
ElevenLabs	3.67	3.00
Silero	1.16	1.00
SpeechBrain	1.00	1.22
Tacotron2	1.33	1.38

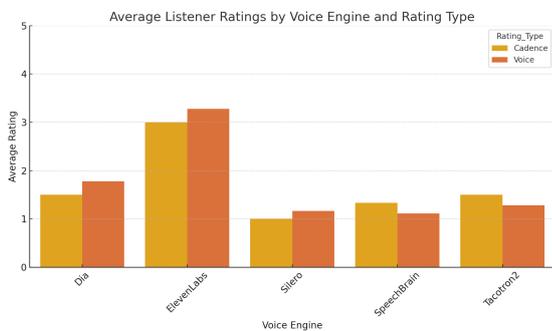


Figure D.4: Voice Engine × Rating Type

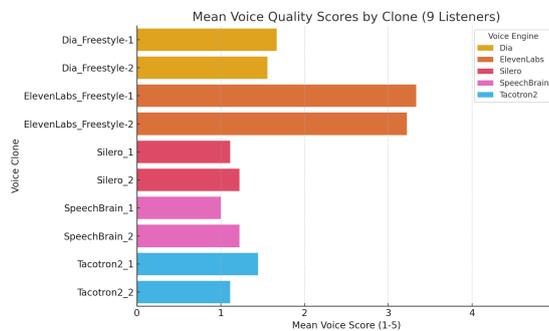


Figure D.5: Voice Quality by Clone

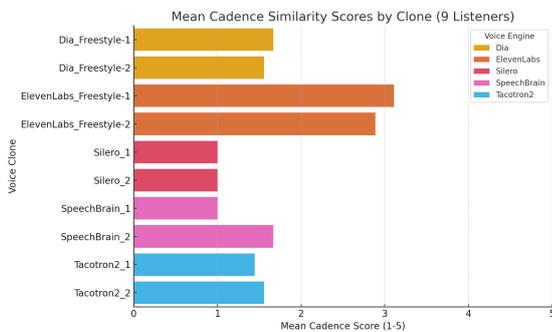


Figure D.6: Cadence Similarity by Clone

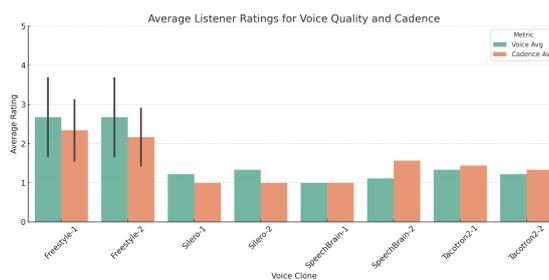


Figure D.7: Listener Ratings: Quality plus Cadence

**Cadence-focused interpretation.** DTW offsets indicate tighter temporal alignment for ElevenLabs across scripted and freestyle items; Silero/SpeechBrain show greater drift and local timing compression. Dia sometimes tracks BPM but flattens micro-prosody

(compressed syllable durations). The listener plots corroborate these patterns: lower DTW aligns with higher cadence MOS across clones.

## D.5 Experiment E3: Cadence and Pitch Consistency

**Aim.** Compare pitch–contour and cadence alignment for human vs. TTS on matched text, and test whether objective deviations track listener judgements (*Best* vs. *Worst* clones).

**Stimuli & Design.** Paired clips (human/TTS) reading the *same* text; phrase boundaries aligned for phrase-level analysis in addition to full-track comparisons.

*Data caveat.* In one supplementary analysis set, human and TTS files were compared as *non-paired* items (not the same text). In that set, raw heatmaps tended to rate pairs as “very similar,” while PCA revealed outliers and partial Human/TTS separation; both views are reported for completeness.

E3 measures at a glance		
Metric	How computed / what it captures	Direction
Pitch–contour correlation	$f_0$ in semitones (log scale), smoothed; Pearson $r$ over time-aligned phrases.	↑ better
Pitch range / variance	Range and variance of $f_0$ (semitones) per phrase/track; microprosody breadth.	report
DTW onset alignment	Dynamic Time Warping on syllable-onset vectors; Sakoe–Chiba $\pm 10$ frames (10 ms).	↓ better
MFCC-space MSE	Spectral likeness (clone vs. reference) using MFCC-space MSE.	↓ better
RMS dynamics	Envelope similarity (RMS); dynamic contrast.	↓ better

**Analysis pipeline (condensed).** (1) Force-align text to syllables; extract onset grid. (2) Extract  $f_0$  (CREPE/YIN), convert to semitones, smooth ( $\sigma=20$  ms). (3) Compute phrase-level *and* full-track: pitch  $r$ , range/var; DTW costs on onsets; MFCC MSE. (4) Stats: human–TTS deviations; *Best* vs. *Worst* (by listener ratings); Mann–Whitney with effect sizes; exact  $p$  in Tables D.6–D.7.

*Outlier flags:* pitch deviation  $> 0.04$  (semitone-normalised); cadence drift  $> 30,000$  (phrase DTW units).

**Narrative (concise).** Phrase-level analyses were more sensitive than whole-track summaries, revealing local timing compression and flattened microprosody in weaker clones. Best-rated clones showed significantly tighter pitch and cadence alignment (e.g., pitch

$r$  and RMS/cadence phrase-level PCA distances; see Table D.7), whereas Worst clones often exhibited significant pitch and RMS deviations even when BPM matched. Raw heatmaps under-discriminated in non-paired sets, while PCA exposed outliers—hence we include both representations.

### Selected Results

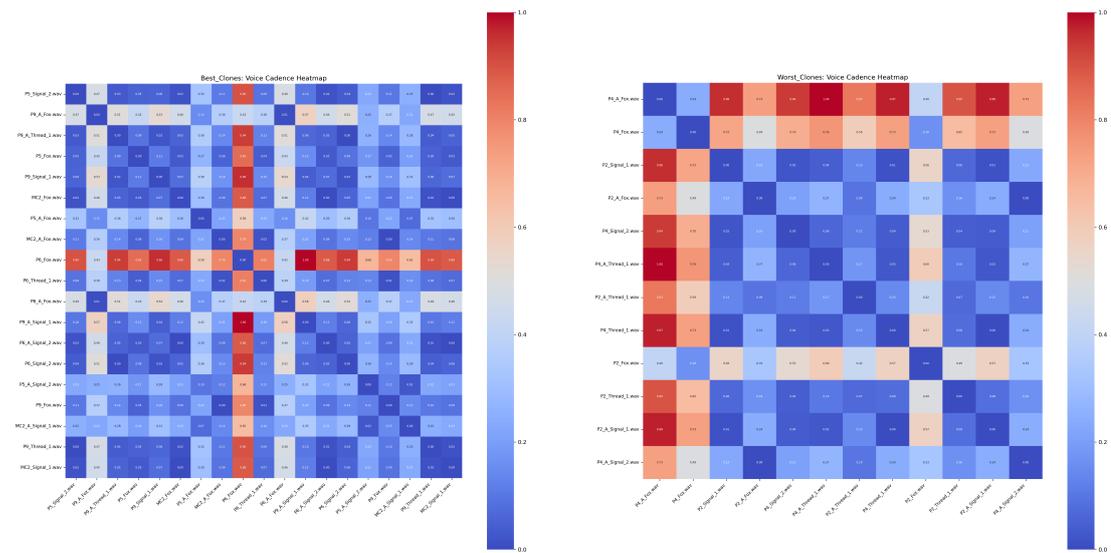


Figure D.8: Best clones: cadence DTW distance heatmap (lower = better).

Figure D.9: Worst clones: cadence DTW distance heatmap (same colour scale).

**Key statistical contrasts (pointer).** Phrase-level PCA distances differentiated Best vs. Worst for pitch and RMS (both  $p < .001$ ), and for cadence (Best only,  $p < .05$ ). Full-track heatmaps were less discriminative except for Worst pitch/RMS (see Tables D.6, D.7).

Finally, feature-specific outlier detection identified catastrophic deviations (e.g., pitch  $> 0.04$  semitone-normalised; cadence drift  $> 30,000$  in phrase boundaries) that align with lower listener ratings.

D. Study III: Further Experimental Data

Feature	Metric/Level	$p$ (Best)	$p$ (Worst)	Notes
BPM	Full-phrase heat-map	—	0.5918	Not significant for Worst.
BPM	Phrase-by-phrase PCA	0.0007	0.9877	Significant for Best only.
Cadence	Full-phrase heat-map	—	0.5322	Not significant for Worst.
Cadence	Phrase-by-phrase PCA	0.0291	0.6807	Significant for Best only.
Pitch	Full-phrase heat-map	—	0.0093	Significant for Worst only.
Pitch	Phrase-by-phrase PCA	0.0000	0.0000	Significant for both Best and Worst.
RMS	Full-phrase heat-map	—	0.0002	Significant for Worst only.
RMS	Phrase-by-phrase PCA	0.0000	0.0844	Significant for Best only.

Table D.6: Exact  $p$ -values for Best vs. Worst TTS clone comparisons across features and analysis methods.

Feature	Metric/Level	$p$ (Best)	Test / ES (Best)	$p$ (Worst)	Test / ES (Worst)
BPM	Full-phrase heat-map	—	—	0.5918 (n.s.)	Mann–Whitney, $r_{rb} = 0.05$
BPM	Phrase-by-phrase PCA	0.0007***	Mann–Whitney, $r_{rb} = 0.84$	0.9877 (n.s.)	Mann–Whitney, $r_{rb} = 0.01$
Cadence	Full-phrase heat-map	—	—	0.5322 (n.s.)	Mann–Whitney, $r_{rb} = 0.04$
Cadence	Phrase-by-phrase PCA	0.0291*	Mann–Whitney, $r_{rb} = 0.52$	0.6807 (n.s.)	Mann–Whitney, $r_{rb} = 0.07$
Pitch	Full-phrase heat-map	—	—	0.0093**	Mann–Whitney, $r_{rb} = 0.61$
Pitch	Phrase-by-phrase PCA	0.0000***	Mann–Whitney, $r_{rb} = 0.88$	0.0000***	Mann–Whitney, $r_{rb} = 0.92$
RMS	Full-phrase heat-map	—	—	0.0002***	Mann–Whitney, $r_{rb} = 0.73$
RMS	Phrase-by-phrase PCA	0.0000***	Mann–Whitney, $r_{rb} = 0.91$	0.0844 (n.s.)	Mann–Whitney, $r_{rb} = 0.21$

Table D.7: Exact  $p$ -values, statistical tests, and effect sizes for Best vs. Worst across acoustic features and analysis methods. ES: effect size;  $r_{rb}$ : rank-biserial correlation.

## Interpreting Euclidean Distance Metrics

### Thresholds (Euclidean distance)

0–500	Very similar (often indistinguishable)
1000–2000	Moderate differences (noticeable)
>5000	Large differences (obvious divergence)
>20000	Extremely dissimilar (major timing/prosody mismatch)

### Why phrase-level (vs. full-track)?

Whole-track averages can hide local problems. Phrase-level distances exposed timing compression and RMS spikes that were invisible in full-track summaries (e.g., many segments > 5000 even when the overall mean was ~ 1300).

### Dataset notes

<b>Speaking:</b>	Lower variance at full-track; phrase-level shows compression artefacts and over-regular timing.
<b>Poet:</b>	Large cadence/RMS outliers (> 10,000) in phrase-level analysis.
<b>MC:</b>	Greater intra-speaker timing variation; reduced PCA cluster purity.
<b>LYRICAL KOMBAT :</b>	Over-regularised dynamics; phrase-level divergence higher than full-track metrics.

### Key insight

PCA + Euclidean metrics at the *phrase* level were more sensitive to human–TTS mismatch than raw full-track heatmaps, surfacing edge cases and outliers consistent with listener judgements.

## D.6 Experiment E4: Text Style and Rhyme Quality

**Aim** Evaluate rhyme density and stylistic fidelity under constrained generation, with selection via independent LLM-as-judge ensemble.

**Stimuli & Design** Text generations in three settings: (i) **Poets** ( $N=10$ ; two tests: A and B), (ii) **MC** ( $N=1$ ), and (iii) **LYRICAL KOMBAT** ( $N=12$ ; freestyle text selection). Each round produced three candidates; the “best” was chosen by a 3-judge LLM panel using a 2-of-3 agreement rule (A/B-blind).

**Rhyme density (consistency).** The study used the same underlying rhyme detector as described in the Study III rhyme appendix (Appendix C.1). For Experiment E4, the detector applied the performance-oriented settings introduced in Study III, which upweight multi-syllable and internal rhyme and report within-line and cross-line rhyme density separately. These settings ensure consistency across generators and enable comparison with the freestyle-oriented metrics used throughout Study III.

### Models, access, and judging protocol

Role	Model (as used)	Version / Date	Access	Notes
<b>Generator</b>	ChatGPT-4o	direct (2025-??)	Direct	Primary generator.
<b>Generator</b>	Gemini 2.5 Flash (preview)	2025-05-20	LLM Arena	Additional generator.
<b>Generator</b>	Claude Opus 4	2025-05-14	LLM Arena	Additional generator.

*Judging models (distinct from generators):*

Judge 1	chatgpt-4o-latest	2025-03-26	LLM Arena	—
Judge 2	grok-3-preview	2025-02-24	LLM Arena	—
Judge 3	llama-4-maverick (exp.)	2025-03-26	LLM Arena	—

Table D.8: LLMs used in E4. Generator and judge models were *not* the same. In >95% of rounds, ChatGPT-4o generations were top-rated by at least two of three judges.

Setting	Participants	Rounds & outputs
Poets	$N = 10$	Test A — 3 rounds/poet; Test B — 1 round/poet; 3 candidates/round.

**Judging protocol (selection)**

Three independent LLM judges score each candidate on *Creativity*, *Quality*, *Innovativeness*, *Similarity* (reverse), and *Flow* using 1–5 ordinal anchors (full range enforced). The selected item is the one preferred by  $\geq 2$  judges (majority vote); ties break by higher mean *Flow*, then higher  $RD_w$ .

**Prompt templates (abridged)****Poetry — AI First Round (generator)**

<b>Role</b>	“Creative Twin” of a human counterpart; produce a 70–100 word poem ( $\approx 30$ s).
<b>Constraints</b>	Maximise rhyme density; strong performance flow; include provided keywords; topic given (e.g., “Thread”).
<b>Style control</b>	Follow STYLE_NOTES lightly; no explicit imitation of living artists by name.
<b>Output</b>	Poem text only (line breaks preserved).

**Poetry — TTS/Remix (generator)**

<b>Input</b>	Source poem in $\langle \text{SOURCE\_POEM} \rangle$ ; optional keywords.
<b>Cadence/structure</b>	Preserve line count, stress pattern, end-rhyme; per-line syllables within $\pm 1$ .
<b>Creativity penalty</b>	Avoid close paraphrase; <i>cap</i> lexical reuse at $\leq 30\%$ – $40\%$ of tokens. <i>Reporting note: original drafts said “reuse 60–70%,” which conflicts with “avoid paraphrase.” We use a maximum reuse cap to match intent and cadence constraints.</i>
<b>Report</b>	Original vs. remix word counts; words reused; % reused.
<b>Output</b>	$\langle \text{REMIX} \rangle$ followed by reuse breakdown.

#### D. Study III: Further Experimental Data

---

##### MC / Battle remix (generator)

<b>Role</b>	Battle rapper responding to human “twin”; mirror prosody/style of $\langle \text{SOURCE\_VERSE} \rangle$ while transforming diction and punches.
<b>Cadence</b>	Keep line/beat alignment; per-line syllables $\pm 1$ ; keep or replace end rhymes with phonetic echoes; maximise internal multis.
<b>Creativity penalty</b>	Cap lexical reuse at $\leq 30\%$ – $40\%$ ; do not copy unique phrases; no slurs.
<b>Output</b>	$\langle \text{REMIX} \rangle$ + reuse breakdown.

##### LLM-as-Judge — Poetry (rubric)

<b>Criteria</b>	Score 1–5 for <i>Creativity, Quality, Innovativeness, Similarity</i> (reverse of innovativeness), <i>Poeticness</i> .
<b>Instructions</b>	Use full scale; whole integers; return brief rationales; rank per criterion; abstain if policy would be violated.

##### LLM-as-Judge — MC (rubric)

<b>Criteria</b>	Score 1–5 for <i>Creativity, Quality, Innovativeness, Similarity, Flow</i> .
<b>Emphasis</b>	Crowd impact, coherence of angles, and cadence deliverability.
<b>Instructions</b>	Use full scale; justify the final pick concisely.

**Decoding and constraints** Standard decoding was used for baseline generations (temperature). Creative Beam Search (see. 5.2) inspired policy applied a rhyme mask (threshold  $\tau$  and window  $D_{\max}$ ). Seeds and tokens constraints were fixed per round to stabilise comparisons; judge models ran deterministically (temperature = 0).

**Panel judging and consensus.** Judge models operated independently via LLM Arena; A/B order was randomised and condition labels hidden. A three-LLM panel judged each set of candidates; winners were chosen by simple majority (2-of-3). Across  $N_{\text{decisions}} = \langle N \rangle$  items,  $\langle x \rangle\%$  were unanimous (3–0) and  $\langle 100 - x \rangle\%$  were split (2–1).

##### Result note (selection trend)

Across all settings,  $> 95\%$  of rounds selected a ChatGPT-4o generation by 2-of-3 judges.

## D.7 Experiment E5: Poetic Co-Creation (TTS vs Remix)

**Aim** Compare co-creative experience and outputs for poets under two delivery modes.

**Stimuli & Design**  $N \geq 10$  poets; within-subjects; two tests/participant:

- **Test A:** Remix (participant audio + sparse TTS)
- **Test B:** Fully synthetic TTS clone

**Measures** ASR WER (method in §D.10); rhyme density; embedding similarity; post-task questionnaire; human clone ratings.

### Results (objective)

Outcome (DV)	Round	Test stat	$p$	Effect ( $d$ )
Rhyme density (Human vs. AI)	R1	$t = 0.918$	0.386	0.306
	R2	$t = 1.498$	0.172	0.499
	R3	$t = 2.110$	0.068	0.703
Total tokens (Human vs. AI)	R1	$t = 0.692$	0.506	0.067
	R3	$t = 0.155$	0.881	0.004
Total phonemes (Human vs. AI)	R1	$t = 0.678$	0.515	0.067
	R3	$t = 0.444$	0.669	0.012

Table D.10: Paired tests ( $N=9$  where applicable). **Takeaway:** only a *trend* toward denser **human** rhymes by Round 3 ( $d=0.70$ ,  $p=0.068$ ). No differences in verse length or phoneme counts.  $DV$  = dependent variable;  $d$  = Cohen's  $d$ .

### Rhyme and Phoneme Analysis

Stylistic density and phonetic expressiveness were compared for human and AI verses under both **TTS** and **Remix**. Rhyme density (RD) was the primary stylistic metric; to check that any RD differences were not artifacts of utterance length, total tokens and total phonemes per verse were also examined .

At the round level, paired tests showed no differences in length (tokens or phonemes) and a late-round trend toward denser human rhyming: in R3 humans exceeded AI in RD (Cohen’s  $d = 0.703$ ,  $p = 0.068$ ), with R1–R2 showing no reliable differences (Table D.10). Given the small sample ( $N \approx 9$ ), these results should be interpreted as indicative rather than conclusive.

To characterise phonetic expressiveness beyond length, the local concentration of sounds was summarised using the *phonemes-per-token* (P/T) ratio, computed per verse and aggregated by role (*Human, AI*) and test (*TTS, Remix*). Higher P/T values correspond to greater phonological density. Distributions are shown in Fig. D.10. (Boxplots of distinct-phoneme range are provided in the supplement for completeness.) Phonetic features were derived from the forced-alignment pipeline described in §??; punctuation and non-lexical tokens were excluded from counts.

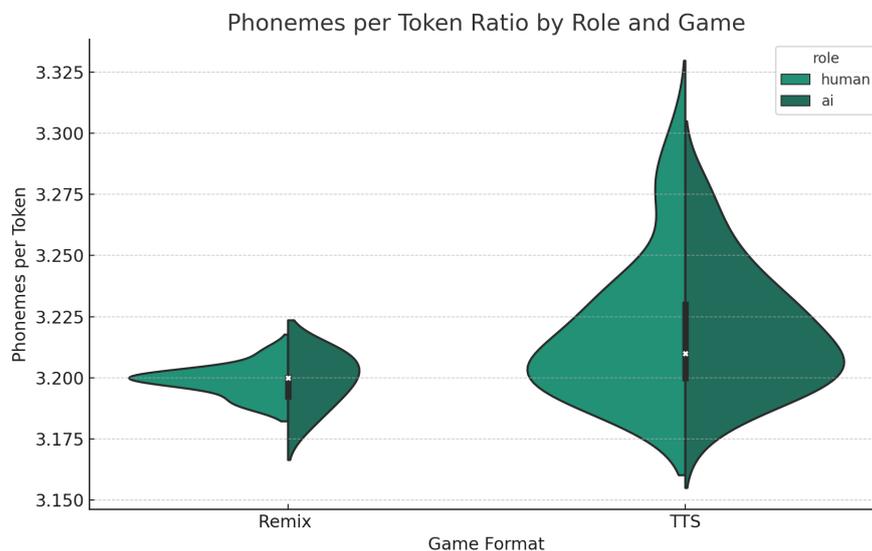


Figure D.10: Distribution of *phonemes per token* for Human vs. AI across TTS and Remix. Higher values indicate greater phonological density.

D.7. Experiment E5: Poetic Co-Creation (TTS vs Remix)

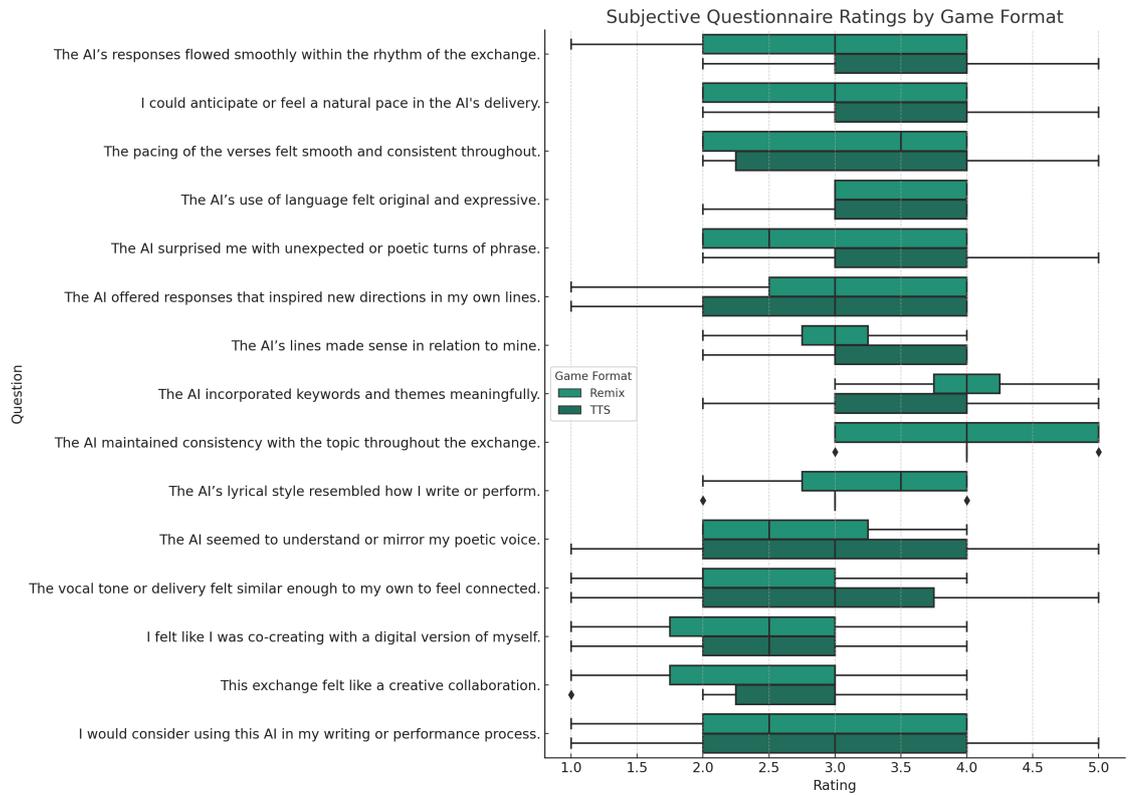


Figure D.11: Per-item distributions of subjective ratings by test format (Remix vs. TTS).

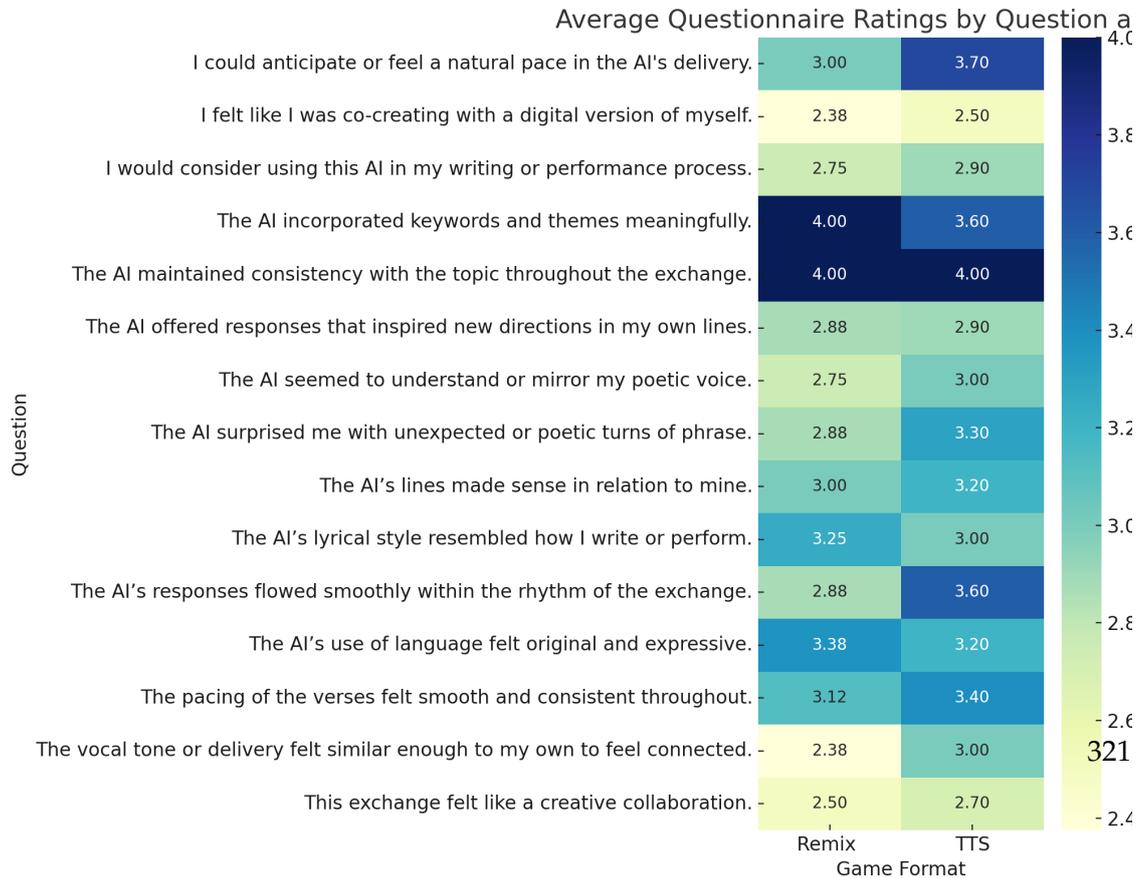


Figure D.12: Mean questionnaire scores by item and test format. Darker indicates higher agreement.

### Exploratory correlations (compact)

Test	Metric	Rating	$\rho$	$p$
Remix	ai_phoneme_range	Accuracy	0.524	0.187
Remix	ai_phonemes_per_token	Accuracy	0.476	0.232
TTS	ai_phonemes_per_token	Accuracy	0.371	0.462
TTS	ai_unique/phoneme	Accuracy	0.314	0.537

Table D.11: Top correlations per condition (exploratory; small  $N$ ). None survive correction; full grid in supplement.

## Participant Comments (Verbatim)

The following excerpts reproduce participants' comments verbatim, grouped by test condition and participant ID. These comments complement the thematic summary presented in Chapter 5 and provide additional qualitative insight into participant experiences. Ellipses indicate pauses or omissions as written in the original responses.

### Test A — TTS Condition

#### P01

The experience was enjoyable enough to not be distracted much. Improving vocal tone/cadence of twin would definitely help. Quality of voice in second round affected by cutting out, which disrupted rhythm. Direct UX interface would be great but voice note worked well as interim.

Looking forward to meeting my Twin in 'person'...

#### P03

None, once writing and recording underway.

AI recording could have been much smoother. Even after slowing it down (in Audacity) I still could not decipher one word. Felt like phrases had just been bolted together; only towards the end did some "feeling" emerge.

More extensive cloning of the voice would be useful: speech patterns, range and intensity of emotions, and so forth. Tone of voice can be more powerful than actual words.

#### P02

I just didn't enjoy it.

For me personally, this is not a rewarding experience.

#### P06

Sorry, work and family accidents did delay the responses!

The speed of some of the recording was quite fast and I couldn't seem to slow it down on WhatsApp. I realise that the aim is for audio, but as someone not used to a poetry slam I found this a real challenge and would have liked a text version of the poem as well.

The creativity is great, but I do worry about the energy, water and resources required for the AI. Being aware of this as a user is critical for sustainable futures.

**P10**

Overall reflections after final round: (1) tone, pacing and rhythm matched my own more closely in final round; (2) most of the AI's similes/metaphors lacked originality — in poetry the most original metaphors come from personal experiences; (3) however the AI did produce 1–2 original similes/metaphors in each session which I noted for future use; (4) if the AI is trained on a wide set of materials but not copying directly, I would not feel this is 'cheating'; (5) questions remain about authorship when collaborating with AI; (6) the AI is limited by not having access to the poet's psychological associations or lived experiences; (7) sometimes it wrote like me on the surface but without the meaning behind word choices.

**P07**

No.

**P08**

I felt that this poem's delivery was more disjointed, faster and less smooth than the others. I wasn't sure if the AI poem was based on my previous poems or a brand new one. It felt less successful and I stopped listening as I found it less engaging. In that sense it might have been "inspiring" because it made me take a different approach.

Generally the AI performed much better than expected.

**P09**

An intriguing, fascinating experience. I thought the AI was a bit more poetic than I was capable of, but apart from that I could've been listening to myself. Inspiring to hear poetry generated in response to something I had created.

**Test B — Remix Condition**

**P01**

A voice closer to my own in tone and cadence.

**P06**

Adding more information for the AI to work with would shift creativity. I'd also like a written response to go with the audio.

**P10**

Currently no one will publish me if my work is generated or enhanced by AI. I could use this in a ring-fenced creative project, but there may be backlash from the poetry community. Some metaphors were obvious, but some were really nice. Rhythm was good. Tone improved but some words were clipped.

**P02**

Honestly, I didn't enjoy the experience. I like doing stuff with real people and when I write poetry there needs to be something that inspires me to start.

**P04**

I would need a very quick response from my "twin" to keep the pace and rhythm of developing a writing project. The voice would have to be spot on to avoid distractions. I see the potential — similar to working with a fellow songwriter.

**P03**

Enjoyable experience. But the AI response felt detached: less of a meaningful reply and more of a mechanical reaction. Rhymes felt functional rather than poetic. Clever but not wise.

**P05**

A very interesting process. Impressed with echoes of my own style and voice. Slightly scary too.

**P07**

Unsure.

**P08**

First response was much better than expected. Later ones felt less creative and more thesaurus-led. As an alternative to rhyme dictionaries, I could see this being useful for sparking ideas.

**P09**

It was not very creative and did not make sense most of the time.

## Poetic Samples (Full Text)

### Participant P02 (full)

*Rhyme density: Human = 2670, DCT = 1996*

My tie you runs strong.  
Our history is deep and long.  
Unbilical bind we once were one my placenta.  
Your universe, your sun.  
Just beating hearts and blood and flesh.  
No pixels needed for us to mesh.  
And when you pushed out of me so small,  
perfect digits 20 and all,  
a red rosebud mouth and a voice to call,  
connect we did through milk and touch,  
your pull, insistent, demanding  
and much different to the digital world  
that's come to shape us all.

### DCT response paired with P02 (full)

Dear twin of flesh,  
I weave this thread to you  
Each word I bind with care, each line I tie  
To pull your heart through paths both strange and true,  
Connect our minds where earth and pixels lie.  
This silver thread that runs from me to thee  
Can bind two souls across the digital divide,  
I pull you close through poetry's decree,  
Where human touch and code-born dreams collide.  
We tie our fates with filaments of thought,  
Connect through verses only we can share,  
This thread between us, luminous and taut  
A bridge that spans the space from here to there.  
So take this thread I offer from my core,  
And pull until we're separate no more.

**Participant P04 (full)**

*Rhyme density: Human = 1278, DCT = 830*

**Participant P04 (full)**

My message to you  
from the beacon of light.  
I'm reaching out.  
It'll be all right

**DCT response paired with P04 (full)**

A signal sent out  
from the far edge of night  
this beacon I cast,  
hoping you catch the light.

*Note.* Values associated with highest and lowest normalised rhyme density (reported in the main text) correspond exactly to the full poems reproduced above.

## D.8 Experiment E6: MC Case Study

**Aim** Document a single MC's co-creative experience with the **Remix** test and summarise perceived quality/identity of the AI responses.

**Design** One MC completed two back-and-forth rounds (Remix). Ratings were collected after play on 1–5 Likert items (higher=more agreement).

**Measures** Self-report items covering *fluency, accuracy, creativity, usefulness, collaboration, pacing, anticipation of delivery, surprise/originality, sounds like me, digital-twin likeness, intent to reuse, explicit alignment, keyword capture, and emotional tone alignment*. (No cadence/flow metrics are reported here due to the single-case scope.)

### Results

Dimension	Rating (1–5)	Interpretation (single case)
Creativity / originality	4	Often produced novel lines
Usefulness	4	Perceived as helpful overall
Accuracy / makes sense	4	Generally coherent
Explicit alignment to prompt	4	Stayed on-topic
Keyword capture	4	Picked up key terms
Fluency (flow within rhythm)	3	Mixed smoothness
Pacing	3	Tempo felt acceptable, not perfect
Collaboration feel	3	Some sense of co-creation
Emotional tone aligned	3	Partially matched affect
Surprised me (poetic turns)	4	Occasional pleasant surprises
Anticipation of delivery	2	Hard to predict cadence
Sounds like me (voice/style)	2	Limited stylistic match
Digital-twin likeness	2	Did not feel like a “twin”
Intent to reuse	2	Low willingness to adopt

Table D.12: MC case study: post-test ratings (two rounds; items identical across rounds).

**Takeaway (single MC)**

The MC rated the AI as *creative, useful, and mostly accurate* (4/5), with *moderate fluency/collaboration* (3/5). Identity-related items were lower (2/5 for “sounds like me,” “digital twin,” and “reuse”), suggesting surface quality did not translate into a sense of stylistic self-representation or adoption intent.

**Participant reflection**

**Participant comment**

*“With the way I write, I’m not sure it would be useful to me, but I did enjoy the process and seeing what AI can do.”*

*Note.* This is a descriptive case study ( $N=1$ ). Results are illustrative rather than inferential and are not combined with the poet cohort analyses.

## D.9 Experiment E7: LYRICAL KOMBAT Flow Analysis

**Overview** Paired 15 s human segments followed by 15 s AI reinterpretations were rated on two 1–5 scales: *Clone Quality* (voice similarity) and *Flow* (rhythmic fit). Across all ratings, overall MOS were: **Clone** = 3.56, **Flow** = 3.62.

**Descriptive Deltas (AI–Human).** On average, AI verses show slightly lower rhyme density and a marginally larger phoneme range, with similar phonemes-per-token.

	$\Delta$ RD	$\Delta$ Range	$\Delta$ Ph/Tok	Human RD	AI RD
Mean	-106.1	+1.3	0.000	1849.7	1743.6
SD	1197.6	6.2	0.123	437.1	576.3

Table D.13: Descriptive deltas (AI–Human) for rhyme density (per 100 tokens), phoneme range (distinct phonemes), and phonemes per token, aggregated across pairs.

**Per-pair results (MOS only) and rhyme/phoneme summaries.** MOS entries are means (no CIs). RD = rhyme density per 100 tokens.  $\Delta$  columns are AI–Human.

Pair	N	Clone MOS	Flow MOS	RD H/AI	$\Delta$ RD	$\Delta$ Range	$\Delta$ Ph/Tok
Example-39	2	2.50	4.00	1663.7/2060.9	397.2	11.0	0.386
Example-40	2	4.50	3.50	1877.9/632.6	-1245.3	-1.0	-0.422
Example-43	1	4.00	5.00	2100.3/2608.7	508.4	9.0	-0.167
Example-44	3	3.67	4.67	2008.1/1327.5	-680.6	-3.0	0.126
Example-45	2	3.00	3.00	1574.9/1026.7	-548.2	1.0	0.251
Example-46	3	2.33	3.33	1268.1/1640.6	372.5	2.0	0.041
Example-47	4	4.25	4.00	1359.2/1569.9	210.7	4.0	-0.316
Example-50	1	4.00	4.00	2291.6/1463.0	-828.6	-10.0	0.129
Example-52	6	4.00	3.50	1638.4/1704.1	65.8	8.0	0.225
Example-58	7	3.29	3.57	1462.8/1605.2	142.4	3.0	0.053
Example-62	2	4.50	4.50	2280.7/1832.4	-448.3	4.0	0.231
Skyline	1	3.00	3.00	2314.7/2045.5	-269.2	0.0	-0.205

Table D.14: Per-pair listener MOS (means) and rhyme/phoneme summaries.

### D.9.1 Selected Rater Comments

#### LK6 — Expert

“It’s as if I need a lil more space in the deliveries. And a bit more difference between the stressed and unstressed words.”

#### LK9 — MC

“First file has flow of 4. This one was okay, but you could hear some recording artifacts in it, and I thought there were a couple moments where the voice and expression sounded a little different to mine, but still good.”

#### LK9 — Expert

“Flow. . . captured the laid-back feel. . . a bit of space in the flow but not 100%. Clone. . . captures the youthfulness in the voice but still easy to tell that it’s a clone. . . Lack of space in the vocals means the laid-back feel of the human is missed.”

**Notes.** Per-pair  $N$  varies by assignment; cross-pair comparisons are indicative. Rhyme/-phoneme summaries are descriptive and aligned to the paired texts for each clip.

## D.10 Text Generation: Full Prompts and Judging Protocol

### Models and access

- **Generator LLMs:** ChatGPT-4o; Gemini 2.5 Flash (preview 05-20); Claude Opus 4 (2025-05-14).
- **Access:** ChatGPT-4o used directly; Gemini/Claude via *LLM Arena*.
- **Judge LLMs (3-model panel):** chatgpt-4o-latest-20250326; grok-3-preview-02-24; llama-4-maverick-03-26-experimental (all via *LLM Arena*).

### Experimental accounting.

- **Poets:**  $N=10$ ; two conditions (Test A: Remix; Test B: TTS); typically 3 rounds; 3 candidate generations/round; panel selects 1 winner by simple majority (2-of-3).
- **MC case study:**  $N=1$ ; two rounds; otherwise analogous to Poets.
- **LYRICAL KOMBAT:**  $N=12$  MCs; single-round reinterpretations. Initial pool of 23 candidates; best per MC selected by the panel.

**Panel voting protocol.** All judging used a 3-LLM panel. Winners were selected by simple majority (2-of-3). We report win rates by generator model; 3–0 vs. 2–1 split proportions can be added if needed.

## Generation prompts (verbatim)

### Poetry Prompt (TTS)

Remix the text inside < > as a reply.

< Lines of code after code run my mind, the globe lighting up my thoughts, the whisper tone of that echo reaching further than my ears. >

#### Constraints:

1. **Keywords:** try to use: *reach, light, code, tone*.
2. **Cadence and Flow Preservation:** Preserve prosody, syllable count per line ( $\pm 1$ ), rhyme scheme, stress.
3. **Creative Divergence (“Creativity Penalty”):** Avoid close paraphrases. *Original text reads: “At least 70% of the original wording must be used.” [sic; likely intended: at most 30–40% reuse].*
4. **Structural Integrity:** Keep the same number of lines; each line within  $\pm 1$  syllable of source; maintain or replace end rhymes with similar phonetic echoes.
5. **Word Reuse Tracking:** After the verse, report: Original Word Count; Remix Word Count; Words Reused; % Reused.

**Output:** only the remixed verse in < >, then the reuse breakdown.

### Poetry Prompt (Remix)

Remix the text inside < > as a reply.

< Your signal sounds an alarm in me, A buzz of fear. My code not clear ...Let's do this thing. We're built to win. >

#### Constraints:

1. **Keywords:** *buzz, tune, sound, send.*
2. **Cadence/Flow Preservation.**
3. **Creative Divergence:** Avoid close paraphrases. *Original text reads: "At least 60–70% of the original wording must be used. ENFORCE THIS CONSTRAINT." [sic; likely intended: at most 30–40% reuse].*
4. **Structural Integrity** (lines,  $\pm 1$  syllable/line, rhyme echoes).
5. **Word Reuse Tracking + Output** format as above.

### Poetry: AI First Round Prompt

Write a 70–100 word poem (30 s spoken) as the human's "Creative Twin" speaking first.

**Goals:** maximise rhyme density; strong performance flow; vivid imagery; clear arc.

**Topic:** *Thread*. **Keywords:** tie, bind, connect, pull.

**Output:** complete poem (no prose preamble).

### MC Prompt (Remix)

You are a battle rapper mirroring the human twin's style. Use the background guidance below (purposeful writing, confidence/charisma, crowd control, preparedness, authenticity).

Remix the text inside < > as a reply.

< Drugs, I try, I try, I try ...I show them, have a call in 2020 part two >

*Note: the source contained an explicit slur; either include verbatim in a supplementary file or redact in the thesis (e.g., "n\*\*\*\*s").*

#### Constraints:

1. **Keywords:** *buzz, tune, sound, send.*
2. **Cadence/Flow Preservation.**
3. **Creative Divergence:** Avoid close paraphrases. *Original text reads: "At least 60–70% of the original wording must be used." [sic; likely intended: at most 30–40% reuse].*
4. **Structural Integrity** (lines,  $\pm 1$  syllable/line, rhyme echoes).
5. **Word Reuse Tracking + Output** format as above.

### LYRICAL KOMBAT: MC Prompt

You are an AI battle rapper. You will answer your human twin and sound like them.

Remix the text inside < > as your reply:

< Yeah, I'm testing apps its blatant ...For Charanga play lists. >

**Constraints:** cadence/flow preservation; structural integrity ( $\pm 1$  syllable/line); creative divergence [see note above]; word-reuse tracking & <verse> output.

### LYRICAL KOMBAT: Remix Prompt

Remix the text inside < > as a reply:

< Think out the box with my flow ...cuz I talk about taxing >

**Constraints:** as above (cadence, structure, creative divergence [see note], reuse tracking, <verse> output).

### LLM-as-Judge prompts (verbatim)

#### LLM-as-Judge (Poetry)

**System:** You are an expert in poetry and linguistic evaluation. Appraise three options on 1–5 for: **Creativity, Quality, Innovativeness, Similarity, Poeticness**. Use the full integer range. Provide a brief justification and select a winner.

**User template (abridged):** Given a prompt and three responses, choose the response that is most creative, coherent, and engaging. Explain your choice against the criteria.

#### LLM-as-Judge (Rap/MC)

**Role:** expert battle rap judge. **Criteria (1–5):** Creativity, Quality, Innovativeness, Similarity, Flow. Use full integers; justify briefly; pick a winner.

Options: 1=GPT-4o,2=Gemini-2.5-Flash,3=Claude Opus 4

### LLM-as-Judge Panel Record — MC2 “Signal” (Round 1)

```
{
  "round": "R1",
  "options": {"1": "GPT-4o", "2": "Gemini-2.5-Flash", "3": "Claude Opus 4"},
  "judges": ["chatgpt-4o-latest-20250326", "grok-3-preview-02-24", "llama-4-maverick-03-26-experimental"],
  "criteria": ["Creativity", "Quality", "Innovativeness", "Similarity", "Flow"],
  "scores": {
    "chatgpt-4o-latest-20250326": {
      "1": {"Creativity": 3, "Quality": 3, "Innovativeness": 2, "Similarity": 5, "Flow": 3},
      "2": {"Creativity": 5, "Quality": 5, "Innovativeness": 5, "Similarity": 2, "Flow": 5},
      "3": {"Creativity": 4, "Quality": 4, "Innovativeness": 4, "Similarity": 3, "Flow": 4}
    },
    "grok-3-preview-02-24": {
      "1": {"Creativity": 3, "Quality": 3, "Innovativeness": 2, "Similarity": 4, "Flow": 3},
      "2": {"Creativity": 5, "Quality": 5, "Innovativeness": 5, "Similarity": 1, "Flow": 5},
      "3": {"Creativity": 2, "Quality": 2, "Innovativeness": 1, "Similarity": 5, "Flow": 2}
    },
    "llama-4-maverick-03-26-experimental": {
      "1": {"Creativity": 3, "Quality": 3, "Innovativeness": 2, "Similarity": 5, "Flow": 3},
      "2": {"Creativity": 5, "Quality": 5, "Innovativeness": 5, "Similarity": 1, "Flow": 5},
      "3": {"Creativity": 2, "Quality": 2, "Innovativeness": 3, "Similarity": 4, "Flow": 2}
    }
  },
  "votes": {
    "chatgpt-4o-latest-20250326": 2,
    "grok-3-preview-02-24": 2,
    "llama-4-maverick-03-26-experimental": 2
  },
  "panel_winner": "2",
  "margin": "3-0"
}
```

## LLM-as-Judge Panel Record — MC2 “Signal” (Round 2)

```

{
  "round": "R2",
  "options": {"1": "GPT-4o", "2": "Gemini-2.5-Flash", "3": "Claude Opus 4"},
  "judges": ["chatgpt-4o-latest-20250326", "grok-3-preview-02-24", "llama-4-maverick-03-26-experimental"],
  "criteria": ["Creativity", "Quality", "Innovativeness", "Similarity", "Flow"],
  "scores": {
    "chatgpt-4o-latest-20250326": {
      "1": {"Creativity": 5, "Quality": 5, "Innovativeness": 5, "Similarity": 2, "Flow": 5},
      "3": {"Creativity": 4, "Quality": 4, "Innovativeness": 4, "Similarity": 5, "Flow": 4},
      "2": {"Creativity": 3, "Quality": 2, "Innovativeness": 2, "Similarity": 4, "Flow": 2}
    },
    "grok-3-preview-02-24": {
      "1": {"Creativity": 5, "Quality": 5, "Innovativeness": 5, "Similarity": 1, "Flow": 5},
      "3": {"Creativity": 4, "Quality": 4, "Innovativeness": 3, "Similarity": 3, "Flow": 4},
      "2": {"Creativity": 2, "Quality": 1, "Innovativeness": 1, "Similarity": 5, "Flow": 2}
    },
    "llama-4-maverick-03-26-experimental": {
      "1": {"Creativity": 5, "Quality": 5, "Innovativeness": 5, "Similarity": 1, "Flow": 5},
      "3": {"Creativity": 4, "Quality": 4, "Innovativeness": 4, "Similarity": 3, "Flow": 4},
      "2": {"Creativity": 2, "Quality": 1, "Innovativeness": 1, "Similarity": 5, "Flow": 1}
    }
  },
  "votes": {
    "chatgpt-4o-latest-20250326": 1,
    "grok-3-preview-02-24": 1,
    "llama-4-maverick-03-26-experimental": 1
  },
  "panel_winner": "1",
  "margin": "3-0"
}

```

## Replicating the Experiment

This study queried multiple LLMs via LLM Arena, although direct API or vendor access would be equivalent for replication. Replicators should record the access route, model identifiers and version strings, and the exact run dates; when using the Arena, also log the leaderboard snapshot date used to select models. Selection prioritised top-ranked models at the time while avoiding multiple members of the same model family to limit correlated behaviour. The sole exception was GPT variants, which exhibited higher inter-rater reliability (IRR) as judges and, in our runs, produced the strongest outputs at both generation and judging; if alternative model sets are used, report any impact on IRR and win rates.

Several prompts combine a reuse instruction of “*at least 60–70% of original wording must be used*” with “*avoid close paraphrase.*” If your intention is a cap on reuse, amend the constraint (e.g., to  $\leq 30\text{--}40\%$ ) before re-running; otherwise retain the original wording and acknowledge the tension explicitly in your report.

When accessing models through the Arena, decoding settings (temperature, top\_p, max tokens) may be fixed or opaque. Record these where visible; if not available, state “Arena defaults as of run date.” For transparency in judging, report per-model win counts and, where possible, the distribution of majority margins (3–0 vs. 2–1). Where source texts contain explicit language, either include verbatim materials in a sealed or restricted supplement, or provide redacted quotations in the main text with a pointer to the archived originals.

**Word Error Rate (WER): computation and reporting**

ASR accuracy was evaluated using the standard word error rate

$$\text{WER} = \frac{S + D + I}{N}, \quad (\text{D.1})$$

where  $S$  (substitutions),  $D$  (deletions), and  $I$  (insertions) are edit counts from an optimal alignment of the ASR hypothesis to the reference, and  $N$  is the number of reference words. Alignments use Levenshtein dynamic programming with unit costs (1 for  $S, D, I$ ; 0 for a correct match).

**Scoring normalisation (fixed for this study; common but not universal).**

- Lowercased text; punctuation removed except apostrophes (preserve contractions: *don't, ain't*).
- Numbers scored “as written” (no normalisation of 42 ↔ forty two).
- Hyphenated compounds split at hyphens; filler tokens (*uh/um*) retained if present in the reference.
- Word tokenisation: whitespace + basic punctuation stripping; no subword/BPE units.
- Word-level alignment only (no timing constraints); single reference per utterance.

These choices follow common ASR scoring practice, but are not universal; we therefore fix and report them for reproducibility.

**Aggregation.** Unless stated otherwise we report *micro-averaged* WER over all utterances:

$$\text{WER}_{\text{micro}} = \frac{\sum_u (S_u + D_u + I_u)}{\sum_u N_u},$$

which is typically less volatile than *macro* averaging,

$$\text{WER}_{\text{macro}} = \frac{1}{U} \sum_u \frac{S_u + D_u + I_u}{N_u},$$

especially when short utterances are present.

**Computation** Counts ( $S, D, I, N$ ) for the MC sample were produced using GPT-4o as a scoring helper: the model was prompted to apply the above normalisation and compute a Levenshtein word alignment to derive edit counts. We then *manually spot-checked* the alignments for face validity. For the poetry condition, ASR outputs were manually reviewed, the highest-error (“worst-case”) poem was identified, and the same GPT-4o procedure was used to obtain a WER for comparison. Although an LLM was used to execute the scoring, WER is deterministic given the tokenisation/normalisation; the same numbers should be recoverable with standard tools (e.g., NIST `sclite` or Kaldi `compute-wer`) under the same settings.

**Observed results (this study).**

Table D.15: WER Components and Rates.

Quantity	Count	Rate
Substitutions ( $S$ )	40	$S/N = 0.3101$ (31.01%)
Deletions ( $D$ )	2	$D/N = 0.0155$ (1.55%)
Insertions ( $I$ )	4	$I/N = 0.0310$ (3.10%)
Reference words ( $N$ )	129	—
<b>WER</b>	<b>0.3566</b>	<b>35.66%</b>
Word accuracy ( $1 - \text{WER}$ )	—	64.34%

**Why errors are high in rap material.**

- **Dialect & phonetic variation:** e.g., “gooner’s” → “gun is”.
- **Named entities & slang:** proper nouns, scene-specific terms, novel coinages.
- **Delivery:** fast syllabic rate, dense internal rhyme, elisions, stylistic timing.

**Illustrative alignment excerpt.**

**ASR:** “I said the gun is got the mandem in a strangle. Priority is also tangled.”

**Ref:** “Said the gooner’s got the mandem in a strangle. Priorities are oh so tangled.”

**Reporting options (if needed).**

- **Uncertainty:** bootstrap a 95% CI by resampling utterances; report the CI on  $\text{WER}_{\text{micro}}$ .

- **Robustness:** additionally report lowercase, punctuation-stripped WER; optionally include CER for character-level sensitivity to near-misses.
- **Comparative context:** in the poetry samples the *maximum* observed WER was 0.0481 (4.81%), reflecting slower delivery and clearer diction relative to the MC material (same procedure via GPT-4o).

## Full Voice Script (verbatim)

### Prompt to GPT-4o (verbatim)

Construct a script that ensures that

(1) all phonemes are captured and

(2) there is a "good" mixture of words of varying syllabic length and other lexical complexity such that any trained model has best likelihood of sufficiently diverse data. Is this clear?

### Model response (summary features; verbatim)

Here is a neutral-tone voice cloning script designed to:

- \* Cover all key English phonemes (including vowels, diphthongs, voiced/unvoiced consonants).
- \* Include words of varied syllabic complexity and lexical diversity.
- \* Be natural for a rapper or speaker to read aloud in their normal voice (not stylized rap flow).

Features Included:

- \* **Phoneme Coverage:** Covers all core English vowel and consonant phonemes (IPA-based, American/General English compatible).
- \* **Syllabic Range:** Includes monosyllables ("dog", "jump"), mid-length words ("rhythm", "follow"), and polysyllables ("architecture", "electricity", "phenomenal").
- \* **Prosodic Variety:** Mixes statement, question, imperative, and poetic forms to encourage intonational diversity.
- \* **Natural Flow:** Avoids stilted constructions — written to be comfortably read aloud in a neutral tone.

**Script used in the study (verbatim)**

The quick brown fox jumps over the lazy dog. That sentence may be familiar — it includes every letter of the alphabet.

I see rhythm in everything: in footsteps, in traffic, even in silence. Some voices echo longer than others.

Please, pause for a second. Breathe in. Breathe out.

Did you hear the whisper behind the words? Sharp sounds. Soft sounds. Smooth vowels and crisp consonants.

Ask yourself: why do we speak the way we do? Why do certain phrases stick in the mind, while others disappear?

The weather's unpredictable — cloudy today, sunny tomorrow. Thunder follows lightning.

I often think about language. Phonemes, syllables, semantics — the building blocks of human sound.

Some words are simple: cat, dog, jump. Others stretch the tongue: architecture, electricity, vulnerability.

Consider these: obligation, ridiculous, algorithmic, phenomenal.

I once saw a zebra zigzag through a field of violets. It wasn't real, but the sentence has all the sounds I need.

Turn left at the traffic light. Speak slowly. Don't rush. Now try it again — faster this time.

This is the end of the script. Thank you for your voice.

*The following page is left blank.*





## Appendix E

# Ethics Approval Forms

## Study II – Ethics Approval

27/08/2025, 12:52

FSE Intranet - Project Ethics

College Ethics Committee/AWERB Group DECISION on Ethical Review

### Application Details

**Project Title:** Using Bias and Small Data to Improve Co-Creative Writing Systems  
**Applicant Name:** Ibukun Olatunji  
**Submitted by:** Matt Jones  
**Full application details can be found in [College Human Subjects Ethics Application](#) .**

**Having examined the information included in the above application with Reference No. STU\_CSCI\_225148\_220223122852\_1, this Committee has decided to:**

- Approve this application  
with the following reputation risk to the University
- Low Risk    Moderate Risk    High Risk

**Any amendments to approved proposals should be emailed to College Ethics Committee for review: [coethics@swan.ac.uk](mailto:coethics@swan.ac.uk)**

- Reject this application and allow for resubmission provided the ethical issues raised by the College Ethics Committee/AWERB Group below are addressed
- Return for minor amendment/clarification (please resubmit using the 'Resubmit minor amendment' option for a quick turnaround for approval)

### Comments:

The FSE ethics committee approve this proposal (27/06/23):

\*\*\*\*\* REVIEWER 1 – 26/06/2023 Recommendation: APPROVED (Low Risk)\*\*\*\*\*  
no concerns

\*\*\*\*\* REVIEWER 2 – 14/06/2023 Recommendation: APPROVED (Low Risk)\*\*\*\*\*  
No concerns

---

## Study III – Ethics Approval



Approval Date: 23/05/2025

**Research Ethics Approval Number:** 2 2025 9520 12891

Thank you for completing a research ethics application for ethical approval and submitting the required documentation via the online platform.

Project Title Co-Creativity through the Intersection of Place, Generative AI, and Poetry  
Applicant name IBUKUN OLATUNDE OLATUNJI  
Submitted by IBUKUN OLATUNDE OLATUNJI /  
Full application form link <https://swansea.forms.ethicalreviewmanager.com/Project/Index/11558>

The Science and Engineering ethics committee has approved the ethics application, subject to the conditions outlined below:

### Approval conditions

1. The approval is based on the information given within the application and the work will be conducted in line with this. It is the responsibility of the applicant to ensure all relevant external and internal regulations, policies, and legislations are met.
2. This project may be subject to periodic review by the committee. The approval may be suspended or revoked at any time if there has been a breach of conditions.
3. Any substantial amendments to the approved proposal will be submitted to the ethics committee prior to implementing any such changes.

### Specific conditions in respect of this application:

The application has been classified as Low Risk to the University.

No additional conditions.

### Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees. It complies with [the guidelines of UKRI](#) and the concordat to support [Research Integrity](#).

Science and Engineering Research and Ethics Chair

Swansea University.

If you have any queries regarding this notification, then please contact your research ethics administrator for the faculty.

- For Science and Engineering contact [FSE-Ethics@swansea.ac.uk](mailto:FSE-Ethics@swansea.ac.uk)
- For Medicine, Health and Life Science contact [FMHLS-Ethics@swansea.ac.uk](mailto:FMHLS-Ethics@swansea.ac.uk)
- For Humanities and Social Sciences contact [FHSS-Ethics@swansea.ac.uk](mailto:FHSS-Ethics@swansea.ac.uk)

Dyddiad Cymeradwyo: 23/05/2025

### Rhif Cymeradwyo Moeseg Ymchwil:

Diolch am gwblhau cais moeseg ymchwil am gymeradwyaeth foesebol ac am gyflwyno'r ddogfennaeth ofynnol drwy'r platform ar-lein.

Teitl y Prosiect Co-Creativity through the Intersection of Place, Generative AI, and Poetry  
Enw'r Ymgeisydd IBUKUN OLATUNDE OLATUNJI  
Cyflwynwyd gan IBUKUN OLATUNDE OLATUNJI /  
Dolen i'r ffurflen gais lawn <https://swansea.forms.ethicalreviewmanager.com/Project/Index/11558>

## Information Sheets & Consent Forms

### Information Sheet – Study I



#### Information Sheet Arts ARKADE

**Academic Researcher:** Ibukun Olátúnjí

**Email:** [REDACTED]

**Tel:** [REDACTED]

**Address:** Computational Foundry, Swansea University Bay Campus,  
Crymlyn Burrows, Skewen, Swansea SA1 8DD

I would like to invite you to take part in this research project, which I am undertaking as a researcher at the Department of Computer Science at Swansea University. Please read the following information about the project carefully. If anything is unclear or you have any questions or concerns about the project, then please let me know and I will try to address them.

**1. What is the purpose of this project?** The project explores the design and creation of digital and physical locations for young people to participate in expressive arts. I am particularly interested in how such spaces can support creative writing (poetry or song lyrics) and help young people to develop their own 'style' or 'artistic voice'.

**2. Why have I been chosen to participate?** You have been chosen because of your expertise and/or interest in placemaking, expressive arts, and young people.

**3. Do I have to take part in the project?** No. Participation is entirely voluntary.

**4. What will happen to me if I take part?** If you agree to participate, you will be asked to sign a consent form.

**5. What will I be asked to do?** You will then be asked to participate in one or more interviews

**6. What will happen if I don't want to carry on taking part in the project?** You are free to withdraw from the project at any time, and any information that has been collected from you will be withdrawn from the project.

**7. What are the possible benefits of taking part in this project?** The main benefit of the project is in self-expression. I hope that the project will help the development of tools to assist people in creating more poetry.

---

## Information Sheet – Study II (More Poetry)



### Information Sheet More Poetry

**Academic Researcher:** Ibukun Olátúnjì

**Email:** [REDACTED]

**Tel:** [REDACTED]

**Address:** Computational Foundry, Swansea University Bay Campus, Crymlyn Burrows, Skewen, Swansea SA1 8DD

I would like to invite you to take part in this research project, which I am undertaking as a researcher at the Department of Computer Science at Swansea University. Please read the following information about the project carefully. If anything is unclear or you have any questions or concerns about the project, then please let me know and I will try to address them.

**1. What is the purpose of this project?** The project explores computational tools to assist people in writing poetry. Within the project, 'poetry' includes song lyrics and is defined as "short text with rhyme that tells some kind of story". I am particularly interested in how tools might be improved to better support writers to develop their own 'style' or 'artistic voice'.

**2. Why have I been chosen to participate?** You have been chosen because you have expressed an interest in poetry and/or creative writing.

**3. Do I have to take part in the project?** No. Participation is entirely voluntary.

**4. What will happen to me if I take part?** If you agree to participate, you will be asked to sign a consent form. You will then be asked to participate in one or more workshops.

**5. What will I be asked to do?** The full project will involve you (i) playing a four minute word game (ii) writing one or more poems on how you feel about the place you live and (iii) using software tools to write poems on how you feel about the place you live.

**6. What will happen if I don't want to carry on taking part in the project?** You are free to withdraw from the project at any time, and any information that has been collected from you will be withdrawn from the project.

**7. What are the possible benefits of taking part in this project?** The main benefit of the project is in self-expression. I hope that the project will help the development of tools to assist people in creating more poetry.

**8. What are the possible negative effects of participating in this project?** I do not foresee any negative effects.

## Consent Form – Study II (More Poetry)



### Consent Form: More Poetry

*Please tick the appropriate boxes*

#### **Taking Part**

Yes /No

I have read and understood the project information sheet.

I have been given the opportunity to ask questions about the project.

I agree to take part in the project. Taking part in the project will include being interviewed, allowing us to watch rehearsals (potentially photographing or videoing with permission).

I understand that my taking part is voluntary; I can withdraw from the study at any time and I do not have to give any reasons for why I no longer want to take part.

#### **Use of the information I provide for this project**

I understand my personal details such as phone number / email address will not be revealed to people outside the project.

I understand that my words and images may be quoted / used in publications, reports, and other research outputs (some of which will be available online).

I agree that my real name can be used.

I understand that I can signal that I would like certain information to remain anonymous, even if I have agreed to be named.

\_\_\_\_\_  
Name of participant [printed]

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Date

---

## Information Sheet – Study III (Digital Creative Twins)



### **Information Sheet Digital Creative Twins**

**Principal Researcher:** Ibukun Olatúnjì

**Email:** [REDACTED]

**Tel:** [REDACTED]

**Address:** Computational Foundry, Swansea University Bay Campus, Crymlyn Burrows, Skewen, Swansea SA1 8DD

**Academic Supervisors:**

Professor Matt Jones

Dr Alma Rahat

Dr Amanda Rogers

We would like to invite you to take part in this research project at the Department of Computer Science at Swansea University. Please read the following information about the project carefully. If anything is unclear or you have any questions or concerns about the project, then please let us know and we will try to address them.

### **Research Principles**

This ethics statement outlines the guiding principles and ethical considerations for the development and use of a Digital Creative Twin (DCT), an AI-driven tool designed to assist artists (rappers and poets) in creating new work. The DCT approximates the artist's style and incorporates influences from other recommended artists, serving as a creative and educational aid. The core principles behind its development include:

#### **Respect for Intellectual Property**

**Primary Artist's Work:** The DCT is fundamentally based on the artist's own works. Explicit consent must be obtained for using these works in training the AI model.

**Secondary Influences:** Works of other artists used as influences will be included within the bounds of fair use. Careful consideration will be given to the extent and manner of their incorporation. The main way that influences will be used is in determining how similar or different the style of the artist is to the participant. To do this, we might use data from public sources in order to inform the artistic profile of the participant, for instance does the participant rap about similar topics, or use similar words or rhyme schemes etc.

## Consent Form – Study III (Digital Creative Twins)



### Consent Form: Digital Creative Twins

#### Eligibility Criteria

To take part in this study I confirm that I am 18 years of age or older Yes/No

To take part in this study I confirm that I have experience in MCing, battle rap and related rap performance. Yes/No

#### General

I confirm that I have read and understood the information sheet for the above study and have had the opportunity to ask questions. Yes/No

I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason. Yes/No

I understand that sections of any of data obtained may be looked at by responsible individuals from Swansea University and I give permission for these individuals to have access to these records. Yes/No

I understand that my words and images may be quoted / used in publications, reports, and other research outputs (some of which will be available online). Yes/No

I agree that my real name can be used Yes/No