The Future of Everything is Distributed – A Plea for Reliable, Responsible, and Resilient Systems Engineering

Berndt Müller

Department of Computer Science, Swansea University

Computer science has seen various trends during its history, but arguably there is one trend that is here to stay: distributed computation and storage. Now, why is this so remarkable to warrant a position paper? Human knowledge has always been distributed, and the generation of new knowledge has long been partial to distributed processes. We have come to take connectivity, and the distributed nature of the data and its processing, that together constitute our knowledge, for granted. We rely on it and most of the time it just works, but digging a bit deeper, one has to come to the conclusion that we are living a dangerous life, because we often use legacy systems and processes, not originally intended for the networked digital world we operate in.

Why does distribution call for a new methodology? For one, we don't have (direct) access to all resources, so traditional (formal) methods might not apply. Snapshots of our rapidly changing dynamic systems do not represent the reality soon after they have been taken. This means that we need to find ways of ascertaining properties that survive change. Ultimately, this has to go hand-in-hand with requirements of explainability and transparency that are becoming legally binding (e.g., the right to explanation in GDPR, and likely further requirements in the EU AI Act that is going to be implemented by the end of 2023). So, apart from a desirable code of conduct, regulation will – sooner or later – impose requirements on technology that traditional approaches cannot fulfil.

This leads to the question of whether technology regulation is good or evil? Those who think the latter might not appreciate the full picture. They often argue that regulation hinders (or at least slows down) innovation. However, they forget that the kind of technology we are seeing the most dramatic advances in, is also the most (ethically) critical. In addition, we have sector-specific strict regulation, e.g., for medicine, finance, and aerospace, which serve purposes hardly anyone would argue against, let alone call for these pieces of regulation to be abolished.

So, I hear you ask: "How can compliance with legal requirements be built into our legacy processes?", but as alleged above, this is the wrong question. We have, far too long, made do with tweaks, patches, and bolt-ons. Besides being extremely fragile, these solutions are typically very costly in the long run when further tweaks will be required. We need a new development methodology that can cope with the distributed nature of our resources, as well as the dynamics and evolving contexts that future (autonomous) systems will be exposed to. Due to the complexity, this will be a tool set that we can draw upon and facilitate

as appropriate. This concept is not new, e.g., UML provides a lot of tools and diagram types, but developers are under no obligation to use all of them.

1 What is Re^3?

In [2] we proposed a methodology for technology design following the **Re^3** principles of *reliability*, *responsibility*, and *resilience*. Despite some overlap in the ingredients of these aspects, the following groups of concepts have proven useful for defining the extent of each.

"Reliability of AI-driven systems can be achieved by facilitating
- use of **formal methods** leading to provably correct systems [...];
- extensive and **systematic testing** [...];
- **diversity in design** leading to fewer unexpected behaviours.

Responsible design of AI components should be based on
- legal and ethical **compliance by design**[, including privacy by design];
- **explainability** to enable understanding of decision-making processes;
- identification of a [...] person who is **accountable** for the system.

Resilient systems require
- **safe** and **secure** design of AI-driven components;
- **robust** systems design and **recovery** strategies."[3]

Combining some of the above ingredients and making sure these cover at least some aspects of each of the three 'Re' will contribute to trust in data-driven and AI-based systems, or – in fact – any other technology. Note again, that the groupings should not be seen as a partition. There are inter-dependencies or overlaps, such as: explainability/interpretability can be achieved by use of formal methods such as argumentation theory; formal methods can be instrumental for the safety of systems; diversity can help ensure legal and ethical compliance, as well as fairness. With growing concerns over misinformation and disinformation, as well as the unprecedented level of cybercrime and cyber-enabled crime, it can be seen as grossly negligent to ignore the above, since such a holistic view is required to identify new vulnerabilities and flag untruths.

2 Transparency and Explainability

Whether our distributed system just used distributed data sources or distributed processing, there is a growing need for certification of data and methods. With respect to data, a minimum requirement could be to declare its provenance. However, things become more complicated with distributed sourcing of data, where it is not sufficient to know that a data set A with certified provenance $\mathsf{prov}(A)$, when combined with a data set B of certified provenance $\mathsf{prov}(B)$ will lead to a combined data set $A \cup B$ with a combined provenance of $\mathsf{prov}(A) \cup \mathsf{prov}(B)$. Some more sophisticated tagging of data or logging of operations on (sets of) data is required, thereby addressing the basics of the transparency issue.

As data is processed, knowledge about potential inherent bias is necessary, but it is currently not fully understood what effect combination and processing of biased data might have on the output, in particular, we need to develop a theory of propagation of bias, akin to propagation of error in physics. This would use transparency in a move to explaining the potential (in)accuracy of an output.

To achieve some level of explainability of an output, there are a number of methods aimed at interpreting machine-learning results, e.g., explainable AI frameworks based on game theoretic optimisation, e.g., Shapley values [6], and local surrogate methods, such as LIME [4]. Of course, there is always the question of the target audience for an explanation. We cannot expect a layperson to understand the complex mathematical underpinnings of some AI models and their statistical analysis. It has recently been proposed to combine factual and counterfactual reasoning for explaining decisions [7].

Our own work in progress (PACE - Parameterised Automated Counterfactual Explanations) combines reasoning with ML and builds-in generation of counterfactuals as follows. (1) Experts define Δ steps for different input parameters, refinement threshold and iteration depth d. (2) Feature-based explanation informs about most relevant parameters. (3) Classification is repeated iteratively for the selected subset of input parameters. (4) The output is a set of explanations that is meaningful to the end user. It should be our aim to add such an approach into mainstream systems development as another tool in an Re^3-toolbox.

3 Extending the Scope: Towards $\mathsf{De}^3\mathsf{Re}^3$

With this position paper, we would like to argue for a shift towards systems design, development, and deployment for reliability, responsibility, and resilience, viewing all of these aspects as first-class citizens in a revamped and timely update to processes involving technology. We emphasise the main phases of technology creation: the conception phase, focusing on the abstract, logical, and tangible *design*; the implementation phase, in which a working prototype is *developed* following the design; and subsequent *deployment* and use of the technology.

These aspects are reflected in the following recommend minimal set of requirements for ethics-by-design $\mathsf{De}^3\mathsf{Re}^3$ systems engineering. Importantly, these requirements have to scale proportionately to the size of the workforce and to the scale of the project:

1. "**Embrace the diversity of diversity**: The emphasis should be the realisation that diversity does not only include the aspects of ethnicity, age, and gender but also socio-economic situation, geography, culture, religion, as well as skills, and methods. This is a fundamental requirement for any complex AI-based system to (a) avoid unconscious disadvantaging of subgroups, (b) explore suitability of solutions in different (cultural) contexts, and (c) exploit novel approaches an re-think existing processes.
2. "**Future-proof systems** by considering likely introduction of legislation around accountability and ethical AI systems design. Organisations will ben-

efit from demonstrable compliance with emerging industry standards (e.g., IEEE 7000 series [1]) and with local and global legislation.
3. "**Re-assess ethical compliance regularly:** Proactively and voluntarily engage in and document periodic or continual checking of ethical (and legal) compliance of evolving, learning, and autonomous systems. View this as an M.O.T.[1] for data-driven and AI-based systems [5].
4. "**Summarise in laypersons' terms** how the decision making process is supported by AI, how quality control is continually ensured, and where human oversight is employed, as well as what data is used and their provenance. This should demonstrate traceability and fairness where possible, and help enhance transparency, which [...] leads to increased user acceptance."[3]

Addressing each of these four requirements would meet a minimum standard as set out by the Re^3 principles. To achieve them, some of the tools in our envisioned De^3Re^3 toolbox can be used: E.g., documentation of the steps taken to comply with regulation (item 2); filling in a checklist of types of diversity (e.g., ethnicity, age, skills, etc.) and the level of engagement (e.g., diversity within the design or development team, diversity on a panel overseeing the development, retrospective ethical approval by a diverse panel within the organisation or by user surveys) for diversity (item 1); Specification of a strategy for re-assessing the deployed system, as well as a mitigation strategy addressing cases in which a re-assessment shows ethical misalignment (item 3). This is comparable to and should go hand-in-hand with a cyber-security risk assessment strategy.

References

1. IEEE Standards Association. IEEE Standard Model Process for Addressing Ethical Concerns during System Design. Website, 2021. https://standards.ieee.org/ieee/7000/6781/.
2. Müller, B. Keynote Presentation. Intelligent Automation Re^3 – An Approach to Reliable, Responsible, and Resilient Systems, 2019.
3. Rees, C., and Müller, B. All that glitters is not gold: trustworthy and ethical AI principles. *AI Ethics* (2022).
4. Ribeiro, M. T., Singh, S., and Guestrin, C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proc. of 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining* (New York, NY, USA, 2016), KDD '16, Association for Computing Machinery, p. 11351144.
5. Sekinah, T., and Müller, B. Should AI systems be given MOTs to put the brakes on hackers? https://www.dataiq.co.uk/articles/articles/should-ai-systems-be-given-mots-to-put-the-brakes-on-hackers, 2019.
6. Shapley, L. S. Notes on the n-Person Game – II: The Value of an n-Person Game, 1951.
7. Stepin, I., Alonso, J. M., Catala, A., and Pereira-Faria, M. Generation and evaluation of factual and counterfactual explanations for decision trees and fuzzy rule-based classifiers. In *2020 IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE)* (2020), pp. 1–8.

[1] M.O.T. is an annual safety, roadworthiness, and exhaust check for UK road vehicles.