Trustworthy AI: Towards the Golden Age of RE?

Matthieu Vergne
Consultant Engineer
matthieu.vergne@meritis.fr

Meritis PACA
Les Algorithmes Aristote B
2000 Route des Lucioles
06901 Sophia Antipolis Cedex
FRANCE

Abstract

In April, 2018, the European Commission established its vision of Artificial Intelligence (AI), leading to the production of guidelines to achieve trustworthy AI one year later. These guidelines, although not mentioning it explicitly, overflow with issues well known in Requirements Engineering (RE). By relating recent RE works to these guidelines, this position paper attempts to show that RE is one of the core components for achieving trustworthy AI, and thus can have a critical impact on the evolution of AI systems and the AI field as a whole for the next few years in Europe.

1 Introduction

In April, 2018, the European Commission established its vision of Artificial Intelligence (AI) [Eur18]. Similarly to "the steam engine or electricity in the past", AI is considered as "one of the most strategic technologies of the 21st century". It may help to "solve some of the world's biggest challenges" and transform "our world, our society and our industry". Since the way we approach AI "will define the world we live in", the Commission considers that "a solid European framework is needed". Pushed by this strong incentive, the Commission formed a high-level expert group in AI (AI HLEG) which, in April, 2019, has produced guidelines to foster trustworthy AI [AI 19].

In this position paper, we claim that these guidelines offer a tremendous amount of opportunities for the field of Requirements Engineering (RE). Far to be a mere field of application of RE techniques, we try to show that RE is a core element in achieving these guidelines. To show that, we describe in Section 2 the core content of the AI HLEG guidelines, especially the seven key requirements they deem as imperative for achieving trustworthy AI. In Section 3, we then relate recent RE works to these guidelines to put in light the deep bonds between the two, before to highlight remaining challenges in Section 4.

In: M. Sabetzadeh, A. Vogelsang, S. Abualhaija, M. Borg, F. Dalpiaz, M. Daneva, N. Fernández, X. Franch, D. Fucci, V. Gervasi, E. Groen, R. Guizzardi, A. Herrmann, J. Horkoff, L. Mich, A. Perini, A. Susi (eds.): Joint Proceedings of REFSQ-2020 Workshops, Doctoral Symposium, Live Studies Track, and Poster Track, Pisa, Italy, 24-03-2020, published at http://ceur-ws.org

2 Trustworthy AI

The AI HLEG guidelines [AI 19] highlight three components which should be met throughout the *entire life cycle* of an AI system, from its conception to its disposal. It must be *lawful*, thus complying with applicable laws and regulations, *ethical*, thus ensuring adherence to ethical principles and values, and *robust* (both from a technical and social perspective), thus avoiding unintentional harm. While the lawful aspect, despite being complex to achieve, is a rather straightforward objective (i.e. comply to national, European, and international laws), the ethical and robust aspects bring more interpretations, and thus are further investigated in these guidelines. This investigation has lead the AI HLEG to establish seven key requirements to be met, through technical and non-technical means, for the development, deployment, and use of AI systems.

- *Human agency and oversight*, thus allowing humans to assess or challenge the system, as well as governance mechanisms of humans over the system.
- *Technical robustness and safety*, including resilience to attack and security, fall back plan and general safety, accuracy, reliability and reproducibility.
- *Privacy and data governance*, especially privacy of acquired and generated user data, the quality and integrity of the data used in learning processes, and a controlled access to user data.
- *Transparency* by supporting the traceability and explainability of the processes and decisions of the AI system, as well as communicating about its abilities and limitations.
- *Diversity, non-discrimination and fairness*, especially the avoidance of unfair bias, allowing all people to access and use the AI system through accessibility and universal design, and the participation of stakeholders who might be directly or indirectly affected.
- *Environmental and societal well-being* by monitoring the impact on society and democracy and considering the broader society, other sentient beings and the environment, as stakeholders for sustainability and environmental friendliness.
- *Accountability* through auditability, without impairing business models and intellectual property, minimisation and reporting of negative impacts, including whistle-blowers and NGOs, ensuring trade-offs when conflicts arise between these requirements, and supporting redress when adverse impacts occur despite the care provided to meet these requirements.

To implement the above requirements, both technical and non-technical methods must be considered. By technical methods, we could mention specific AI system architectures, "by design" conceptions (e.g. privacy-by-design, security-by-design), explanation methods (e.g. field of Explainable AI (XAI)), testing and validating, and quality of service (QoS) indicators. Non-technical methods could be regulations (e.g. product safety laws), codes of conduct and key performance indicators (KPIs), standardisation (e.g. ISO, IEEE P7000), certification, accountability via governance frameworks (e.g. person in charge of AI ethics), education, and social dialogue. These methods, and more globally the requirements to satisfy, should encompass *all stages* of an AI systems life cycle, as shown in Figure 1, reproduced from [AI 19].

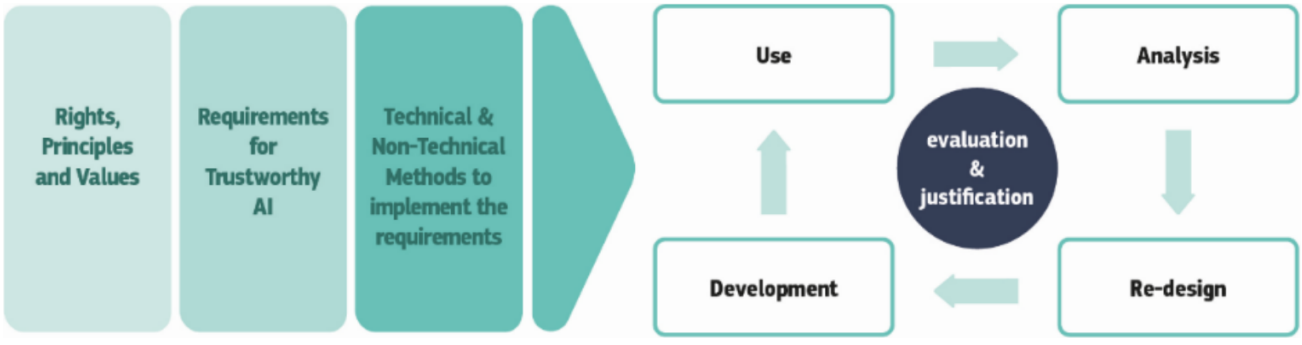

Figure 1: Realising Trustworthy AI throughout the system's entire life cycle

3 Opportunities for RE

If the reader is familiar with RE literature, we might dare to claim that reading the previous section was enough to remind him or her of several RE works. Indeed, since the guidelines should be operationalized throughout the whole life time of AI systems, it also covers all the phases of RE, from eliciting the requirements of the AI system to verifying their satisfaction. Actually, each requirement can be related to recent works in the field:

- *Human agency and oversight* can be supported at development time, by supporting the decision making about quality requirements at strategic and operational levels [OW18], as well as run time, by monitoring the satifaction of the requirements and diagnosing their violations [VRCH17].
- *Technical robustness and safety* is also of interest, whether we speak about methods to follow official standards or the impacts of these methods on the people implementing them [PH17]. The AI HLEG also includes in this requirement the security aspect, which can be covered by various sources of security requirements, thus helping to reach a comprehensive set of security goals [GBO18].
- *Privacy and data governance* can be subject to interpretation misalignments, which can be mitigated through shared ontologies [BHBN18], as well as incomplete privacy policies, which can increase the perceived privacy risks of the users [BEB19].
- *Transparency* regarding how the AI system works can be achieved by showing how the components of the produced AI system relate to its requirements, in other words having a performant traceability [HP18], but transparency can also bring issues like information overload, information starvation, and transparency leading to biased decisions [HSPA18].
- *Diversity, non-discrimination and fairness* can be achieved by ensuring that the relevant stakeholder profiles are explicitly considered, for example through a comprehensive list of personas and persona-based modelling [NRJTW+18] or by analysing the massive feedback of actual users of an AI system [LZW18].
- *Environmental and societal well-being* is also an aspect that some RE researchers investigate, like integrating green strategies in quality requirements for optimizing energy and other resources consumption [CFL18].
- *Accountability* is also of interest in RE, whether by supporting auditors of AI systems by ensuring that traceability links reach a satisfying level of quality [HP18] or by supporting the audited companies in evolving their acceptance tests based on the evolution of the requirements they should meet [HBCG18].

We can also go further by considering RE frameworks, like OpenReq [PFF18], which help requirements analysts through automated recommendations of various kinds, like new requirements, quality tips, stakeholders to consider, or requirements prioritization. Such kind of framework could be strengthened and extended to help requirements analysts to cope with the various dimensions covered by the AI HLEG requirements list. Although not always related to AI, and without assuming they are the most relevant, these works deal with these requirements in some ways, and thus provide illustrative examples of how RE is deeply linked to these guidelines.

At this point, we only related existing RE works to the high level requirements of the guidelines. But we can further look in the details, especially by taking the dichotomy operated by the AI HLEG with technical and non-technical methods to satisfy these requirements.

The technical methods mainly relate to what the AI scientific community may produce, but works in RE can also help on some regards. For instance, we work on translating human-written mind map diagrams into machine-readable graphs to produce knowledge bases to be queried [BGOK18]. The structured graphs and the models used for translating the mind maps into the graphs offer a support to explain the translation. An AI system based on such a knowledge base would be then at least partially explainable, as opposed to a knowledge base directly built from unstructured data. We can also mention the writing of acceptance tests: automatically translating requirements into acceptance tests is not a recent idea [GH99] and we are also working on how to evolve tests with their requirements [HBCG18]. Quality of service (QoS) indicators can also build on the satisfaction of quality requirements if we structure and formalize them well enough to build concrete quality measures [OW18].

At the opposite, non-technical methods can have a rather comprehensive coverage of RE works. Regulations are far to be ignored in RE, with for example the Nòmos framework to deal with them [Sie10, IJS+14]. Of course, they are only one source of norms, and other approaches consider more sources of requirements to be

more comprehensive [GBO18], including standards and certifications. [BOJRG18] also highlights the lack of standards to ensure the compatibility of the Internet of Things (IoT) ecosystems, an aspect also applicable to ethical concerns since each component of an IoT ecosystem might fulfil them to different degrees and in different ways, thus making hard to grasp the fulfilment of the whole. RE can also help in the elicitation of user preferences from massive amounts of users [LZW18] and thus help in establishing lists of concrete requirements, based on social dialogue, that governance frameworks could build on.

In brief, it is easy to find recent RE works to illustrate the multiple focuses of the AI HLEG guidelines. In fact, to do so, we were prepared to look at papers published in several occurrences of RE-related venues like REFSQ, EmpiRE, CAiSE, or iStar. In the end, the attentive reader may have noticed that, among the 18 RE references we cite in this section, 11 are from REFSQ 2018 (almost half of the 23 papers of the venue), with 9 of them used to illustrate all the key requirements presented. In other words, a single occurrence of a RE-related conference can suffice to broadly illustrate the key requirements of the AI HLEG guidelines, an evidence that further strenghten the correlation between these guidelines and the field of RE.

4 Challenges for RE

Of course, relating existing works to the AI HLEG requirements does not mean that we have all we need to satisfy them. Rather, there is still a lot of work, starting from communicating better what we do. Indeed, despite the 59 occurrences of "requirements" (without counting pictures) in the 41 pages of the AI HLEG document, the total absence of "requirements engineering" should ring a bell.

Beside making people aware of the RE methods and tools, there is also space for improvements to support the AI HLEG key requirements. For instance, although we have a growing interest in the challenges of AI-related topics, like the Internet of Things (IoT) [BOJRG18], it is still young on the specificities of Big Data issues [AM18], a core aspect of AI. There is also non AI-specific aspects which are still open issues in RE, like how to deal with ambiguous and incomplete requirements [DvdSL18]. [CHD18] also shows that human analysts can have various degrees of reliability in recovering traces, and thus further support is required. Requirements analysts may exploit tools to help them in various RE tasks, but some are still far from a reliable use in production [HP18]. Dealing with requirements in complex organizations is by itself challenging, with the difficulty to detect infeasible requirements, the variance in assumptions and definitions between teams, or the overlook of sources of requirements [ADW18, WPKG18].

We can also take a step further: after considering how to use RE for AI, we can also consider how to use AI for RE. Like other fields, RE is gradually relying on more AI techniques to help the requirements analysts doing their job [WV18, DvdSL18, PFF18, LZW18]. But if we claim that RE can help in designing ethical AI systems, can we still say so if the AI systems RE relies on are not ethical? Can we ensure that the requirements we help to discover allow to reach ethical AI if the AI used to identify those requirements is not ethical? In other words, RE4AI is not the end of the loop: it is but a mere step towards RE4RE.

5 Conclusion

In this position paper, we attempted to show that RE has a tremendous opportunity in the current European trends in the field of AI. We first explained how AI is considered as a key component for the future by the European Comission, with the production of guidelines to produce trustworthy AI systems (i.e. lawful, ethical, and robust). We listed the seven key requirements of these guidelines and tried to show how recent works in RE relate to them, from both technical and non-technical perspectives. We hope that these relations show how the field of RE, despite not being explicitly mentioned, is at the core of these guidelines, and thus the great amount of opportunities that RE researchers may found in working on this topic. We also relied on recent works in RE to remind some current challenges that may increase the difficulty of this task. Nevertheless, we have no doubt that the RE community is strong and dynamic enough to tackle them. We are aware that the guidelines are already two years old, and that our presentation is far from the high standards of the scientific rigour, but we hope that this position paper brings a relevant and interesting point to the workshop discussions, maybe helping at drawing a future research agenda.

6 Acknowledgements

This publication reflects the view only of the author, and Meritis PACA cannot be held responsible for any use which may be made of the information contained therein.

References

[ADW18] Wasim Alsaqaf, Maya Daneva, and Roel Wieringa. Quality Requirements Challenges in the Context of Large-Scale Distributed Agile: An Empirical Study. In Erik Kamsties, Jennifer Horkoff, and Fabiano Dalpiaz, editors, *REFSQ*, volume 10753, pages 139–154. Springer International Publishing, Cham, 2018.

[AI 19] AI HLEG. Ethics Guidelines for Trustworthy AI. Report, European Commission, April 2019.

[AM18] Darlan Arruda and Nazim H. Madhavji. State of Requirements Engineering Research in the Context of Big Data Applications. In Erik Kamsties, Jennifer Horkoff, and Fabiano Dalpiaz, editors, *REFSQ*, volume 10753, pages 307–323. Springer International Publishing, Cham, 2018.

[BEB19] Jaspreet Bhatia, Morgan C. Evans, and Travis D. Breaux. Identifying incompleteness in privacy policy goals using semantic frames. *Requirements Engineering*, 24(3):291–313, September 2019.

[BGOK18] Robert Andrei Buchmann, Ana-Maria Ghiran, Cristina-Claudia Osman, and Dimitris Karagiannis. Streamlining Semantics from Requirements to Implementation Through Agile Mind Mapping Methods. In Erik Kamsties, Jennifer Horkoff, and Fabiano Dalpiaz, editors, *REFSQ*, volume 10753, pages 335–351. Springer International Publishing, Cham, 2018.

[BHBN18] Mitra Bokaei Hosseini, Travis D. Breaux, and Jianwei Niu. Inferring Ontology Fragments from Semantic Role Typing of Lexical Variants. In Erik Kamsties, Jennifer Horkoff, and Fabiano Dalpiaz, editors, *REFSQ*, volume 10753, pages 39–56. Springer International Publishing, Cham, 2018.

[BOJRG18] Johanna Bergman, Thomas Olsson, Isabelle Johansson, and Kirsten Rassmus-Grhn. An Exploratory Study on How Internet of Things Developing Companies Handle User Experience Requirements. In Erik Kamsties, Jennifer Horkoff, and Fabiano Dalpiaz, editors, *REFSQ*, volume 10753, pages 20–36. Springer International Publishing, Cham, 2018.

[CFL18] Nelly Condori Fernandez and Patricia Lago. The Influence of Green Strategies Design onto Quality Requirements Prioritization. In Erik Kamsties, Jennifer Horkoff, and Fabiano Dalpiaz, editors, *REFSQ*, volume 10753, pages 189–205. Springer International Publishing, Cham, 2018.

[CHD18] Bhushan Chitre, Jane Huffman Hayes, and Alexander Dekhtyar. Second-Guessing in Tracing Tasks Considered Harmful? In Erik Kamsties, Jennifer Horkoff, and Fabiano Dalpiaz, editors, *REFSQ*, volume 10753, pages 92–98. Springer International Publishing, Cham, 2018.

[DvdSL18] Fabiano Dalpiaz, Ivor van der Schalk, and Garm Lucassen. Pinpointing Ambiguity and Incompleteness in Requirements Engineering via Information Visualization and NLP. In Erik Kamsties, Jennifer Horkoff, and Fabiano Dalpiaz, editors, *REFSQ*, volume 10753, pages 119–135. Springer International Publishing, Cham, 2018.

[Eur18] European Comission. Artificial Intelligence for Europe. Communication from the Comission, European Commission, April 2018. COM (2018) 237 final.

[GBO18] Ana-Maria Ghiran, Robert Andrei Buchmann, and Cristina-Claudia Osman. Security Requirements Elicitation from Engineering Governance, Risk Management and Compliance. In Erik Kamsties, Jennifer Horkoff, and Fabiano Dalpiaz, editors, *REFSQ*, volume 10753, pages 283–289. Springer International Publishing, Cham, 2018.

[GH99] Angelo Gargantini and Constance Heitmeyer. Using model checking to generate tests from requirements specifications. *ACM SIGSOFT Software Engineering Notes*, 24(6):146–162, November 1999.

[HBCG18] Sofija Hotomski, Eya Ben Charrada, and Martin Glinz. Keeping Evolving Requirements and Acceptance Tests Aligned with Automatically Generated Guidance. In Erik Kamsties, Jennifer Horkoff, and Fabiano Dalpiaz, editors, *REFSQ*, volume 10753, pages 247–264. Springer International Publishing, Cham, 2018.

[HP18] Paul Hbner and Barbara Paech. Evaluation of Techniques to Detect Wrong Interaction Based Trace Links. In Erik Kamsties, Jennifer Horkoff, and Fabiano Dalpiaz, editors, *REFSQ*, volume 10753, pages 75–91. Springer International Publishing, Cham, 2018.

[HSPA18] Mahmood Hosseini, Alimohammad Shahri, Keith Phalp, and Raian Ali. Four reference models for transparency requirements in information systems. *Requirements Engineering*, 23(2):251–275, June 2018.

[IJS+14] Silvia Ingolfo, Ivan Jureta, Alberto Siena, Anna Perini, and Angelo Susi. Nmos 3: Legal Compliance of Roles and Requirements. In Eric Yu, Gillian Dobbie, Matthias Jarke, and Sandeep Purao, editors, *Conceptual Modeling*, volume 8824, pages 275–288. Springer International Publishing, Cham, 2014.

[LZW18] Tong Li, Fan Zhang, and Dan Wang. Automatic User Preferences Elicitation: A Data-Driven Approach. In Erik Kamsties, Jennifer Horkoff, and Fabiano Dalpiaz, editors, *REFSQ*, volume 10753, pages 324–331. Springer International Publishing, Cham, 2018.

[NRJTW+18] Genana Nunes Rodrigues, Carlos Joel Tavares, Naiara Watanabe, Carina Alves, and Raian Ali. A Persona-Based Modelling for Contextual Requirements. In Erik Kamsties, Jennifer Horkoff, and Fabiano Dalpiaz, editors, *REFSQ*, volume 10753, pages 352–368. Springer International Publishing, Cham, 2018.

[OW18] Thomas Olsson and Krzysztof Wnuk. QREME Quality Requirements Management Model for Supporting Decision-Making. In Erik Kamsties, Jennifer Horkoff, and Fabiano Dalpiaz, editors, *REFSQ*, volume 10753, pages 173–188. Springer International Publishing, Cham, 2018.

[PFF18] Cristina Palomares, Xavier Franch, and Davide Fucci. Personal Recommendations in Requirements Engineering: The OpenReq Approach. In Erik Kamsties, Jennifer Horkoff, and Fabiano Dalpiaz, editors, *REFSQ*, volume 10753, pages 297–304. Springer International Publishing, Cham, 2018.

[PH17] Luciana Provenzano and Kaj Hnninen. Specifying Software Requirements for Safety-Critical Railway Systems: An Experience Report. In Paul Grnbacher and Anna Perini, editors, *REFSQ*, volume 10153, pages 363–369. Springer International Publishing, Cham, 2017.

[Sie10] Alberto Siena. *Engineering Law-Compliant Requirements: the Nomos Framework*. phd, University of Trento, January 2010.

[VRCH17] Michael Vierhauser, Rick Rabiser, and Jane Cleland-Huang. From Requirements Monitoring to Diagnosis Support in System of Systems. In Paul Grnbacher and Anna Perini, editors, *REFSQ*, volume 10153, pages 181–187. Springer International Publishing, Cham, 2017.

[WPKG18] Rebekka Wohlrab, Patrizio Pelliccione, Eric Knauss, and Sarah C. Gregory. The Problem of Consolidating RE Practices at Scale: An Ethnographic Study. In Erik Kamsties, Jennifer Horkoff, and Fabiano Dalpiaz, editors, *REFSQ*, volume 10753, pages 155–170. Springer International Publishing, Cham, 2018.

[WV18] Jonas Paul Winkler and Andreas Vogelsang. Using Tools to Assist Identification of Non-requirements in Requirements Specifications A Controlled Experiment. In Erik Kamsties, Jennifer Horkoff, and Fabiano Dalpiaz, editors, *REFSQ*, volume 10753, pages 57–71. Springer International Publishing, Cham, 2018.