Foreword

Sebastian Hallensleben¹

Accepted: 24 August 2021 / Published online: 29 September 2021
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2021

A lot of progress has been made in AI governance over the past year. Several new tools and frameworks have emerged, and the institutions and working groups contributing to a better handling of AI governance are getting almost too numerous to count. A particularly strong role is being played by the OECD network of AI experts (ONE.AI). This network has recently presented one of the first comprehensive frameworks for classifying AI. The framework strikes the delicate balance between the undesirable extremes of oversimplification on one hand, and an overabundance of technical detail on the other hand. It achieves the goal of structuring AI technology and applications at the right level of granularity to be useful for policymakers as well as decision makers elsewhere. Included with this framework is a proposal for an initial risk assessment of AI applications. This addresses concerns expressed by businesses about being “overloaded” by AI governance, in particular in industrial settings and similar contexts where ethical concerns are less likely to play a significant role. The proposed risk assessment aims to leverage the overall classification to determine almost without further effort whether an application is in the lowest risk category or not. Meanwhile, a working group on trustworthy AI is also addressing issues such as quality and safety at the OECD level.

There has been broad agreement for quite some time about the key principles for ethical and trustworthy AI, including transparency, fairness, privacy, robustness, accountability and sustainability. Effective AI governance needs to ensure that AI systems and applications support rather than endanger democratic values. In its recent draft regulation, the European Union has taken a risk-based approach as a guiding principle to focus the effort required for good AI governance where it is most needed and to avoid interfering with innovation as far as possible. The

draft regulation also recognises that effective AI governance relies on the multi-cultural and multi-institutional cooperation of stakeholders. In this spirit, regulatory activities in the European Union will benefit from close links to the activities within the OECD, to CAHAI within the Council of Europe as well as other efforts. A shared objective of such international cooperation has to be to create technical and regulatory as well as standardisation tools to make good AI governance testable and enforceable. Good AI governance is too important to be left at risk of being turned into a mere PR exercise without substance.

The year 2020 has shown that, for all its promise, AI is not an automatic magic bullet for the challenges of our time. Despite numerous projects, AI has not yielded any breakthroughs in handling the COVID-19 pandemic. AI also has not brought us significantly closer to a solution for the climate crisis and in particular for addressing its societal and transformational demands and implications. Nevertheless, we believe that AI as a new type of technology that is different in several ways from anything that came before it, is bound to eventually yield useful insights and solutions for some of the hardest challenges we are facing today. Even in places where we do not yet understand the underlying causalities and mechanisms, AI will help us recognise correlations, exploit them, and use them as an inspiration and springboard towards furthering our understanding of the world.

AI forces us to rethink one of the most fundamental concepts—the concept of identity in the digital space. The availability of AI toolkits such as Deepfakes and GPT3 makes it possible for the first time in human history to convincingly mimic human beings, at least in the digital space. Thus, humans will be joined online by millions of bots that are mass-generated with AI techniques and can, e.g. express pre-determined political views, promote specific products, act as influencers (even in Youtube or Instagram), and garner followers. This has far-reaching consequences for the functioning of democratic discourse as well as the functioning of markets. As a result, we will need new ways of identifying ourselves in the digital space—ways that establish our

✉ Sebastian Hallensleben
Sebastian.hallensleben@vde.com

¹ VDE Association for Electrical, Electronic and Information Technologies, Stresemannallee 15, 60596 Frankfurt am Main, Germany

humanness but do not give away our real world identity. More broadly, AI forces us to reconsider business and social models and to innovate models that are responsible and sustainable in the context of new technological possibilities.

AI does not exist in a vacuum. It has multiple interfaces with the existing regulatory and standard setting worlds. As an example, there is already a well-developed framework and community for testing and enforcing functional safety of products. In fact, it is a field that has grown and matured over several decades, both in a horizontal, cross-cutting sense and in industry-specific concretizations. AI products, like all other products, have to conform to functional safety regulations and standards but pose fundamental new challenges that go beyond established functional safety approaches. Therefore, a well-managed multi-stakeholder effort will be needed, not least to bring together the disjoint communities of experts.

Finally, as the Chair of the European AI standardisation at CEN-CENELEC, I want to emphasise the crucial role of standardisation. In most countries, it is a well-established principle that regulation sets out broad objectives and the boundaries of what is permissible, whereas standardisation

fills in the crucial technical details that would be difficult for policymakers or regulatory bodies to define. Within the European Union, this synergy is even codified as the New Legislative Framework. AI governance relies to a substantial extent on standardisation. This is not only true in Europe where governance issues are being discussed in a formal setting in standardisation organisations since 2019 but also at the international level, for example at the IEC as well as in joint ISO/IEC activities. So far, the links between standardisation and regulation have been close, with a shared desire to come up with the best possible outcome for citizens, businesses, workers, and consumers alike. I hope that this special issue on AI governance contributes to the debate and to a constructive collaboration of all stakeholders.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.