Build Back Better with Responsible AI

Lars Kunze¹

Accepted: 14 January 2021

© Gesellschaft für Informatik e.V. and Springer-Verlag GmbH Germany, part of Springer Nature 2021

Dear readers,

What a year 2020 was! The worldwide coronavirus pandemic (COVID-19) has drastically changed our work and social lives and fundamentally altered our understanding of how the world works. However, the pandemic has not only forced us to adapt to rapidly changing circumstances, but also to seek new opportunities and to rethink what the world after COVID-19 could look like. Here, I want to explore the role of artificial intelligence (AI) in *building back better*—a strategy which aims to mitigate risks posed by future threats. These risks are not limited to those related to COVID-19, but include a whole range of ethical, societal, and environmental issues. Hence, the question is how AI can help us to recover and to reshape our economy and society in a way that is more sustainable and resilient to future disruptions and challenges?

In the future, AI will, without a doubt, play a key role in many areas including manufacturing, logistics, transport, education, security, as well as health and social care. Many of the technological challenges that AI research and development will face have been recently described by Kristian

Kersting in his editorial “Rethinking Computer Science Through AI” (KI Journal, Issue 4, 2020).¹ However, in addition to solving these difficult technical challenges, we also need to ensure that future AI technologies are designed, developed, operated, and used in ways that are not only efficient, safe, and robust but also sustainable, fair, and inclusive to benefit both the society and the environment.

Responsible research and innovation (RRI) frameworks² and tools³ provide guidelines and good practices with the aim to foster the design of inclusive and sustainable research and innovation. Adopting such frameworks for *building back better* within AI research would help to pro-actively address challenges arising when deploying AI technologies in the wider society. For example, EPSRC’s AREA Framework⁴ encourages all stakeholders to *anticipate* and *reflect* on the consequences of scientific and technological innovations. Furthermore, the framework *engages* stakeholders through formal and informal processes of dialogue and debate. And finally, the framework uses these processes to *act* and to influence the direction and trajectory of the research and innovation process itself. Hence, by integrating responsible innovation approaches in the design, manufacturing, operation, repair, and end-of-life recycling of AI systems we can anticipate and mitigate societal, commercial and financial, and environmental risks.

Let us consider an example of what RRI could mean when developing AI methods for autonomous driving. What are potential risks and implications that we need to anticipate, reflect on, engage with, and act upon? As a hypothetical example, we could *anticipate* that the performance of a safety-critical (deep) learning-based model will depend on the characteristics and the type of the environment in which an autonomous vehicle is deployed. If the trained model has a significant lower performance in rural areas compared to urban areas, we could *reflect* that the deployment of the

✉ Lars Kunze
lars@robots.ox.ac.uk

¹ Department of Engineering Science, Oxford Robotics Institute (ORI), University of Oxford, Oxford, UK

¹ <https://doi.org/10.1007/s13218-020-00692-5>.

² <https://www.orbit-rri.org>.

³ <https://www.rri-tools.eu>.

⁴ <https://epsrc.ukri.org/research/framework>.

AI technology could have dramatic (although unintended) implications for the safety of its users in rural areas. Reflecting on the underlying assumptions and the framing of the problem one might discover imbalanced training data in favour of urban over rural areas. Moreover, when *engaging* with government bodies and looking at the national statistics of road accidents, one could also see that the majority of fatalities occurs on rural roads. While this can be mostly attributed to higher average speeds which often result in more serious collisions, other factors include the lack of road markings and street lighting which are also underrepresented in the training data. Hence, to assure that the performance and safety provided by AI methods does not disadvantage the population in rural areas we could *act* by setting guidelines and requirements for the creation of balanced data sets for training and testing of AI-based components for autonomous driving.

However, the adoption of RRI processes in AI (or other disciplines) poses also new challenges. While RRI processes might be easily integrated into new design processes, it is not straight forward to apply them to those which already exist. Furthermore, we need to answer some of the following questions: how can we balance an agile development against a very thorough consideration of future implications? How can we ensure meaningful stakeholder involvement? And how can we encourage international responsibility and sustainability? Finally, it is not clear how much RRI is needed or appropriate, in particular when encountering major tensions between profit and responsibility. And who should make the decision?

It seems very clear, overall, that responsible research in AI must play a major role if we want to tackle future societal and environmental challenges head-on. To this end, current RRI frameworks and tools provide a valuable starting point to steer the research and innovation process in AI. However, so far, only a small number of research projects are carried out under these principles. Hence, to *build back better* RRI approaches need to be embraced and adopted by more AI researchers, projects, and the wider community. So, why not consider it for your next project?

In any case, I hope you will enjoy reading the current issue of the KI journal. As ever, it comprises a range of interesting and diverse articles and contributions on a special topic. This time the issue is dedicated to the fascinating research area of Developmental Robotics.

Lars Kunze

Lars Kunze

1 Forthcoming Special Issues

1.1 Education in Artificial Intelligence K-12

Guest Editors: Gerald Steinbauer (Graz University of Technology), Martin Kandlhofer (Austrian Computer Society), Tara Chklovski (Founder and CEO of Technovation), Fredrik Heintz (Linköping University), and Sven Koenig (University of Southern California).

The upcoming special issue of the KI Magazin addresses the emerging topic of education in artificial intelligence (AI) at the K-12 level. In recent years, artificial intelligence (AI) has attracted a lot of attention from the public, and become a major topic of economic and societal discussion. AI already has a significant influence on various areas of life and across different sectors and fields. The speed and force with which AI is impacting our work and everyday life poses a tremendous challenge for our society and educational system. Teaching fundamental AI concepts and techniques has traditionally been done at the university level. However, in recent years several initiatives and projects pursuing the mission of K-12 AI education have emerged. In this context we also see education organizations and AI experts as well as governments developing and deploying AI-curricula and programs for a K-12 audience. The aim of this special issue is to provide a compact overview over this growing field. We invite contributions from researchers, practitioners, and educators interested in education in AI at K-12 level.

1.2 NLP and Semantics

Guest Editors: Daniel Hershcovich (University of Copenhagen), Lucia Donatelli (Saarland University), and Stephan Oepen (University of Oslo).

Making computers as intelligent as humans has been argued to be as difficult as making them understand human language, which is one of the focus points of natural language processing (NLP). The field has been changing over the past decades, generally moving from rule-based methods to statistical ones. Machine learning (ML) methods, in particular deep learning, are today omnipresent, challenging methods based on linguistic theories by fully end-to-end data-driven modeling. However, combining powerful ML models with flexible pipelines and frameworks based on human and linguistic insight is an exciting development promising the best of both worlds.

Natural language processing applications are abundant, and are already changing people's lives, enabling effortless translation, learning and interaction with human-centric systems in robotics and virtual assistants. While many classical NLP problems deal with modeling the surface form of

linguistic utterances, general natural language understanding and generation depend on explicit or implicit modeling of semantics, including meaning, communicative intent, and the complex mapping to the linguistic form. Computational semantics is the study of how to automate the process of constructing and reasoning with meaning representations of natural language expressions, which can take many forms, such as continuous vectors or discrete graphs.

For this special issue, we welcome contributions including, but not limited to the following topics: lexical semantics, compositional semantics, cross-lingual semantics, semantic parsing, syntax-semantics interface, semantic role labeling, textual inference, formal semantics, coreference, discourse, reading comprehension, knowledge acquisition, common sense reasoning, summarization, multimodal semantics, semantic annotation, ethical aspects in semantic representations, under specification, ontologies, sentiment analysis, stylistic analysis, argument mining, and human-robot interaction.

1.3 Learning Computational Thinking

Guest Editors: Nina Bonderup Dohn (University of Southern Denmark), Yasmin Kafai (University of Pennsylvania), Anders Mørch (University of Oslo), and Marco Ragni (University of Southern Denmark).

Our call is informed by the ever-increasing digitalization and integration of AI into everyday life. Computational thinking (CT) is the use of problem-solving processes to enable humans to design creative and computational solutions to many of the wicked problems which the world faces today. Utilizing computational thinking is an integral part of developing AI systems—and the AI developers of tomorrow are the learners of today. CT is not restricted to algorithmic thinking, computer visualizations, and programming, but includes as well analogue representations and even aspects

of embodied cognition. We call for papers that develop cognitive, educational, and computational models, report psychological studies, in situ case studies of design activities, children's engagement in computational participation and critical reflection on IT and AI, and investigate pedagogical designs that support the learning of CT and AI using a range of different tools and methods, e.g., computer visualizations, simulation programs, analogue algorithmic thinking, bodily interactions, or physical things that can be programmed, or using all methods from AI. We aim to integrate researchers from all disciplines to present novel research to advance the understanding of CT. All submissions will be peer-reviewed.

The topics of interest for the special issue include, but are not limited to:

- Theoretical investigations of CT compared with other types of thinking.
- Any form of AI support for AI thinking and developing AI thinking.
- Analysis of AI models for taxonomy of AI thinking abilities.
- Comparison of analogue and digital forms of CT.
- Multimodal interactions of users and systems.
- Comparison of successful educational methods for AI and CT (Fachdidaktik).
- In situ experiments with learning designs supporting the learning of CT and AI.
- Practice- or design-based research on CT in different disciplines.
- Case studies of learners engaged in CT activities, computational participation, and critical reflection on AI in formal and informal learning.
- AI interfaces or Apps for enabling learners to engage in AI and CT.