AI is multidisciplinary

Virginia Dignum (Umeå University and WASP-HS; virginia@cs.umu.se)
DOI: 10.1145/3375637.3375644

Introduction

The last few years have seen a huge growth in the capabilities and applications of Artificial Intelligence (AI) and autonomous systems. Hardly a day goes by without news about technological advances and the societal impact of the use of AI. AI is changing the way we work, live and solve challenges. For example, it can improve healthcare, protect elephants from poachers, and work out how broadband should be distributed.

There are large expectations about the potential of AI to help to solve many current problems and to support human well-being, but we are also witnessing an increasing concern about the potential of AI to do wrong. In particular issues of biased decisions, increased inequality, breaches of privacy, but also how the future of work will look like, and how can we keep AI under control, just to mention a few.

Many questions are arising from the development and use of AI. These are fundamental questions about what is intelligence, what is a fair and just decision, how to balance between individual and collective interests, how to solve ethical dilemmas or how automation will impact the labor market, are questions technology alone cannot answer. Answers to these questions require interdisciplinary approaches.

AI is not some kind of magic over which we have no control. We people are the ones that make AI happen. At the same time, it is increasingly more difficult to define AI (Monett & Lewis, 2017). Depending on the focus and the context, AI can refer to a computational technology (e.g. deep learning), a field of scientific research, or an autonomous entity capable of sensing and acting (Theodorou & Dignum, 2019). However, it is important to realise that eventhough AI is in fact a piece of software that we people design, understanding and guiding the impact of AI in society requires more than understanding its technical design.

All technological change has traditionally been accompanied by fundamental societal changes. We are now at the brink of yet another one. It is up to us to make this a change for the better, for all of humankind and for the environment. This challenge is too large and too important to be left to engineers alone. All fields of knowledge from humanities and social science to art and design are needed to better build, understand and use AI.

A multidisciplinary perspective

Artificial Intelligence (AI) offers the potential for augmentation and potential replacement of human tasks and activities within a wide range of applications. The current pace of change for AI innovation is high, requiring societal, institutional and technological adjustments, and new opportunities for continued innovation across different domains including business and management, government, public sector, and science and technology. In order to navigate this potential, explore opportunities and mediate challenges, it is essential to integrate humanities and social science into the conversation about law, economics, ethics and impact of AI and digital technology. Only together we can chart a way forward into a beneficial and trustworthy future with our increasingly algorithmic societal systems.

The rapidly growing capabilities and increasing presence of AI-based systems in our lives raise pressing questions about the impact, governance, ethics, and accountability of these technologies around the world (Dwivedi et al., 2019). How can decisions about when, what for and how should AI be applied and how to integrate the variety of views and requirements from people who use, interact with, and are impacted by these technologies? How do we harness the potential of AI systems while ensuring that they do not exacerbate existing inequalities and biases, or even create new ones? These questions cannot be answered from a computer science or engi-

neering perspective alone. In fact, we can say that Artificial intelligence is not any longer an engineering discipline but requires a broad involvement of different disciplines and participants.

Education plays here an important role. However, most current AI and robotics curricula worldwide deliver engineers with a too- narrow task view. The wide impact of AI on society requires a broadening of engineering education to include (Dignum, 2020):

1. analysis of the distributed nature of AI applications as these integrate socio-sociotechnical systems, and the complexity of human-agent interaction;
2. reflection on the meaning and global effect of the autonomous, emergent, decentralized, self-organizsing character of distributed learning entities and how they operate;
3. incremental design and development frameworks, and the unforeseen positive and negative influence of individual decisions at a system level, and as well as how these impact human rights, democracy, and education;
4. the consequences of inclusion and diversity in design, and how these inform processes and results;
5. understanding of governance and normative issues, not only in terms of competences and responsibilities, but also in the case of views on health, safety, risks, explanations, and accountability; and,
6. the underlying societal, legal, and economic models of socio-sociotechnical systems.

Broadening AI curricula is possibly also a way to attract a more diverse student population. When AI curricula are known to be transdisciplinary, it can be expected that female students, who traditionally choose humanities and social subjects over engineering ones, may be motivated to choose AI. In parallel, humanities and social sciences curricula also need to include subjects on the theory and practice of AI. For example, law curricula need to prepare law experts on how to address legal and regulatory issues around AI.

A Diversity Perspective

Impact of AI is now a global topic of discussion in academic and policy circles. At least 84 public–private initiatives have produced statements describing high-level principles, values and other tenets to guide the ethical development, deployment and governance of AI (Mittelstadt, 2019; Jobin, Ienca, & Vayena, 2019). Governance is necessary for the reduction of incidents, to ensure trust, and for society's long-term stability through the use of well-established tools and design practices. Well-designed regulations do not eliminate innovation but instead enhance it through the development and promotion of both socio-legal and technical means to enforce compliance [1]. Moreover, policy is needed to determine human responsibility in the development and deployment of intelligent systems, filling the gap that emerges from the increased automation of decision. Further, the ultimate aim of regulation is to ensure well-being for all in a sustainable world so it should can guide responsible research, development and use of AI (Theodorou & Dignum, 2019).

Besides disciplinary diversity, it is also important to consider cultural diversity, which includes factors such as education, religion, language. Artificial intelligence is increasingly pervasive and applied across cultures and geographic regions. Failure to understand cultural diversity impacts negatively the universal right to access to the advantages that the technology brings about. In an increasingly connected AI world, incentives and regulations can support awareness and commitment to a diverse perspective ensuring that AI applications are truly adaptable to a diverse cultural space, and thus enabling access to all. Or, as described in (Floridi et al., 2018) "Debates about technological challenges may lag behind the actual technical progress, but if they are strategically informed by a diverse, multistakeholder group, they may steer and support technological innovation in the right direction. Ethics should help seize opportunities and cope with challenges, not only describe them. It is essential in this respect that diversity infuses the design and development of AI, in terms of gender, class, ethnicity, discipline and other pertinent dimensions, in order to increase inclusivity, toleration, and the richness of ideas and perspectives."

Accountability is an important dimension of decision-making and therefore an essencial element for responsible and trustworthy AI. Inclusion, diversity and fairness are crucial to ensure that the impact of AI on individual and society is aligned with human rights and social values, and to analyse the nature and the role of biases that emerge from theoretical or empirical models that underpin AI algorithms and the interventions driven by such algorithms. While the biases emerging from the theoretical and empirical models also affect human-controlled educational systems and interventions (e.g. hindsight and unconscious biases), the key mitigating difference between AI and human decision-making is that human decisions involve individual flexibility, context-relevant judgements, empathy, as well as complex moral judgements, missing from AI (Porayska-Pomsta & Rajendran, 2019). Again here, a multidisciplinary approach is essential to mitigate risks and increase benefits of AI applications.

AI, Humanities and Society: The Swedish approach

In the past years, the Swedish Wallenberg Foundations (https://www.wallenberg.org/en) have made significant investments in research on AI, autonomous systems and software through the program WASP (http://wasp-sweden.org/). There is no doubt these technologies will have major effects on our society. Realising that fundamental research in AI must go hand in hand with a deep understanding of its societal, ethical, legal, economical and cultural impact, the Wallenberg Foundations are now investing over SEK 660 million in humanities and social sciences research, through the WASP-HS program (Wallenberg AI, Autonomous Systems and Software Program - Humanities and Society: wasp-hs.org). The WASP-HS program will analyze the impact of AI and autonomous systems in society, study the consequences of technology transition, and how these insights can contribute to the design and development of trustworthy and responsible systems. The WASP-HS program therefore extends and complements technological research and advances on AI with a strong investment on research in social science and humanities.

The WASP-HS program is planned to run 2019 – 2028 and will form an independent and parallel program to WASP, while maintaining a close dialogue with the WASP program. WASP-HS includes the following components:

- A research program aiming at forming an interdisciplinary community across the funded projects
- A national graduate school
- Recruitment of junior and visiting faculty across disciplines
- International partnerships and activities.

This multidisciplinary approach will advance our understanding of the challenges and impact of intelligent and autonomous technology, as well as contributing to the development of theory and practice of human and societal aspects of AI and autonomous systems. WASP-HS research tackles the challenges and impact of upcoming technology shifts as well as contributing to the development of theory and practice of human and societal aspects of AI and autonomous systems, and in particular, focus on potential ethical, economic, labor market, social, cultural and legal aspects of technological transition.

Conclusions

Many questions are arising from the development and use of AI. These are fundamental questions about what is intelligence, what is a fair and just decision, how to balance between individual and collective interests, how to solve ethical dilemmas or how automation will impact the labor market, are questions technology alone cannot answer. Answers to these questions require interdisciplinary approaches.

That is why the research program WASP-HS is of crucial importance for the future of AI, and in fact for the future of all of us. The aim of the WASP-HS program is to strengthen research and competence on the challenges and impact of AI, autonomous systems and software in the humanities and social sciences and to build a ground to utilise this expertise to inform society and industry on the design of socially aligned systems and on consequences for industry, society and humanity. We look forward

to collaboration with similar programs across the world.

References

Dignum, V. (2020). Responsibility and Artificial Intelligence. In M. Dubber, F. Pasquale, & S. Das (Eds.), *The Oxford Handbook of Ethics of AI.* OUP.

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., ... Williams, M. D. (2019). Artificial intelligence (ai): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*. Retrieved from http://www.sciencedirect.com/science/article/pii/S026840121930917X doi: https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... others (2018). Ai4people—an ethical framework for a good ai society: opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689–707.

Jobin, A., Ienca, M., & Vayena, E. (2019). Artificial intelligence: the global landscape of ethics guidelines. *arXiv preprint arXiv:1906.11668*.

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical ai. *Nature Machine Intelligence*, 1–7.

Monett, D., & Lewis, C. W. (2017). Getting clarity by defining artificial intelligence—a survey. In *3rd conference on" philosophy and theory of artificial intelligence* (pp. 212–214).

Porayska-Pomsta, K., & Rajendran, G. (2019). Accountability in human and artificial decision-making as the basis for diversity and educational inclusion. Springer Nature.

Theodorou, A., & Dignum, V. (2019). Towards ethical and socio-legal governance in AI. *Nature Machine Intelligence*, *to appear*.

Virginia Dignum is Professor of Social and Ethical Artificial Intelligence at Umeå University, Sweden and associated with the TU Delft in the Netherlands. She has a PHD in Artificial Intelligence from Utrecht University. Currently she is the scientific director of WASP-HS, the Wallenberg Artificial Intelligence, Autonomous Systems and Software Program - Humanities and Society. She is a Fellow of the European Artificial Intelligence Association (EURAI), a member of the European Commission High Level Expert Group on Artificial Intelligence, of the World Economic Forum's Global Artificial Intelligence Council, of the Executive Committee of the IEEE Initiative on Ethically Alligned Design, and a founding member of ALLAI-NL, the Dutch AI Alliance. She is the author of "Responsible Artificial Intelligence: developing and using AI in a responsible way" published by Springer in 2019.