



Selected Papers of #AoIR2020:
The 21st Annual Conference of the
Association of Internet Researchers
Virtual Event / 27-31 October 2020

SOCIAL EXPECTATIONS OF AI & THE PERFORMATIVITY OF ETHICS

Prof. Aphra Kerr
Maynooth University

Dr. Marguerite Barry
University College Dublin

Prof. John D. Kelleher
Technological University of Dublin

This paper explores societal expectations of artificial intelligence (AI) in the EU, the UK and Ireland over the past decade. It combines analysis of public documents and a small-scale public survey to identify key actors, mechanisms and discourses shaping societal expectations of AI and driving the focus on ethics and AI.

LITERATURE REVIEW

This study is informed by the sociology of expectations within science and technology studies (STS). Expectations are statements that say something about the future (Pollock and Williams, 2016, van Lente, 2012, Borup et al., 2006). Formal expectation mechanisms include foresight and research prioritisation exercises deployed by governments, consultancy firms and companies to rationalise future innovation investment. Informal expectations are “images, statements and prophecies” (van Lente 2012:772) from experts and non-experts which circulate through social networks and the media. In the construction of expectations, different actors draw from, and add to, the repertoire of visions that shape the expectations around a technology and drive its innovation dynamic.

Previous studies identify three ‘forces’ of expectations: raising attention and legitimising investment; coordinating networks of companies and research institutions; and providing heuristic guidance and direction to research and innovation (Van Lente 2012: 773-774).

Kerr, A., Barry, M. and Kelleher, J. D. (2020, October). *Social Expectations of AI and the Performativity of Ethics*. Paper presented at AoIR 2020: The 21th Annual Conference of the Association of Internet Researchers. Virtual Event: AoIR. Retrieved from <http://spir.aoir.org>.

Expectations are 'performative' in the sense that they may prompt certain social actions. Mackenzie (2008:17) distinguishes between two types of performativity: 'generic' performativity is when theoretical models, language or approaches are adopted but do not change things in practice, and 'effective' performativity is when the models, language or approaches make a difference in the real world.

AI has had an undulating history of societal expectations and the current period can be viewed as a third period of growth in AI. As in each of the earlier cycles, it has a distinct technological trigger. This time the trigger is machine learning and big data.

METHODOLOGICAL FRAMEWORK

This study used a mixed methods approach: a thematic analysis of public documents to identify the key actors, formal and informal mechanisms shaping societal expectations of AI and a small-scale public survey to explore public expectations of AI and ethics.

Using theoretical sampling we identified 41 documents on AI research and policy published by the EC or other actors in the UK and Ireland since 2011. This sample included 7 types of actors: international consultancies (e.g. McKinsey); public agencies; academic and expert reports; public surveys; professional associations (e.g. IEEE) and media statements by workers and whistle-blowers. A full list of the documents is available [here](#).

A face to face administered survey of 164 individuals visiting a science gallery provided an opportunity to explore expectations of AI and ethical AI amongst a sub-section of the public. Respondents came from 25 different countries with diverse occupational backgrounds. More than half felt that they had some or good levels of familiarity with AI.

FINDINGS

Our analysis identified a range of positive and negative expectations of AI and the emergence of a discourse on ethics and AI since 2016.

The document analysis identified a narrow range of positive expectations including 'cost reductions' and increased 'efficiency' from AI innovations in services across formal consultancy, expert and government reports. A small number of international consultancies were influential in shaping core definitions and concepts. We also identified negative expectations of AI issued by whistle-blowers, the media and public surveys. These documents focussed on dataveillance, data misuse and the working conditions of content and community moderators. By 2016 formal public and professional reports directly addressed negative expectations of AI, and included efforts to develop ethical guidelines founded on European values and research on ethical AI. Most offered little operational detail.

Significantly, the survey respondents' expectations appeared to be shaped more by informal mechanisms including media stories and films rather than by formal mechanisms or everyday experience of using AI systems. They viewed automation and efficiency as both the top negative and positive aspects of AI. AI was viewed positively

when it could be used to help or serve humans, but negatively when control was ceded to it. They were concerned with large scale societal impacts (e.g. job losses, dehumanisation and inequality) rather than everyday questions. Further, respondents had difficulty conceptualising abstract ethical principles and values. They ranked privacy, transparency, safety and security as top ethical concerns but these issues varied by domain. They believed that it was possible to design ethical AI, but that both public and private actors should be responsible and accountable for negative AI impacts.

DISCUSSION

The sociology of expectations helps to identify the range of actors attempting to shape our expectations and the direction of AI innovation. There is currently a divergence between expectations of AI and current practices in its application. Formal public and private sector documents are used to justify significant investment in AI, but they say little about the practical technical, social and ethical challenges faced in different contexts. The emergence of a significant discourse on ethics and AI in 2016 appears to be a response to high-profile stories about the misapplication and misuse of AI.

Our findings point to shared societal expectations that we can design ethical AI (or a European approach to AI) and that developers (especially academic ones) will behave ethically. However, ethics and AI is challenging. Ethical issues are difficult to address in practice, they can vary from domain to domain, there is uncertainty about who should provide oversight, and it is unclear how to enforce accountability for negative impacts. If our scientifically literate or interested members of the public have difficulty grasping abstract ethical issues, we need to carefully consider how workers and the public will fare when tasked with designing, deploying and using these systems.

Societal expectations that we can create 'ethical AI' currently acts as a **generic** discourse for reassuring investors, governments and users. It is crucial that we move beyond this to **effective** solutions to specific challenges, including how data is used by AI systems, the conditions under which human workers develop and deploy AI, and the differential impacts of AI systems on users.

The gap between societal expectations of ethical AI in discourse and its practical development will remain until we accept the limitations of AI in complex social contexts and recognise that non-technological solutions, such as robust governance mechanisms and empowered human workers, are required to make AI work ethically. The latter is as much a social and regulatory question as an ethical one. A full paper is available in [Big Data and Society](#).

References

BORUP, M., BROWN, N., KONRAD, K. & VAN LENTE, H. 2006. The sociology of expectations in science and technology. *Technology Analysis & Strategic Management*, 18, 285-298.

MACKENZIE, D. 2008. *An Engine, Not a Camera. How Financial Models Shape Markets*, Cambridge, MA, MIT Press.

POLLOCK, N. and WILLIAMS, R. 2016. *How Industry Analysts Shape the Digital Future*, Oxford, Oxford University Press.

VAN LENTE, H. 2012. Navigating foresight in a sea of expectations: lessons from the sociology of expectations. *Technology Analysis & Strategic Management*, 24, 769-782.

VAN LENTE, H., SPITTERS, C. and PEINE, A. 2013. Comparing technological hype cycles: Towards a theory. *Technological Forecasting and Social Change*, 80, 1615-1628.