Dr. Vincent Miller
BA MA PhD PGCHE vincent-miller-a18044221/
Reader in Sociology, University of Kent, Canterbury, UK

University of Kent

Artificial Intelligence Impact Assessment: the case for responsible AI adoption

Artificial Intelligence (AI) has made great strides in recent years, especially in the areas of machine learning, large language models, natural language processing, knowledge representation and generative AI. AI is already used extensively in industry, government, science and cultural production, but recent advances particularly in generative AI have led to an increased awareness of the potential for AI to be more widely adopted in industry, the public sector and work life more generally. This in turn has led to increasing concerns about the potential repercussions of more widespread AI use in terms of the robustness, safety and reliability of the technology itself, its lawful and ethical application, and its impact upon society and social wellbeing.

Numerous real-life instances of poor price prediction,[11] algorithmic discrimination in health care[12] and recruitment,[13] invasions of privacy,[14] copyright violation,[15] and automated payment systems[16] and the like have cost industry millions in corporate losses and legal redress, ruined corporate reputations, and raised the potential risks of implementing AI systems in public discourse. Academic research has demonstrated how the use of automated systems can lead to decisions that harm the poor, reinforce racism, and amplify inequality.[17,18,19] This not only has implications for corporate social and legal responsibilities, but for the wellbeing of society at large.

Such perceptions also create an overall lack of trust and acceptance of AI technologies among the public. A recent global survey by KPMG found 61% of respondents across 17 countries were wary about trusting AI systems, and a similar proportion were unwilling to accept the use of AI.[20] This varied drastically between countries, with Indian respondents most trusting at 75%, and Finns least trusting at 16%. Interestingly, willingness to accept AI was generally lower than trust, with less than 25% of Americans, Canadians, Australians, British, Dutch, Japanese and Finns willing to accept AI systems. This is clearly a barrier to innovation and the adoption of AI technologies moving forward.

One way to understand and mitigate the possible risks and problems associated with increasing AI adoption is to apply an impact assessment approach to AI, much in the same way that this has been applied to various other contexts such as the environment, human rights, social and data protection. Such an approach has been encouraged in academic literature.[21,22]

11 fortune.com/education/articles/what-zillows-failed-algorithm-means-for-the-future-of-data-science/.
12 www.nature.com/articles/d41586-019-03228-6.
13 www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine.
14 www.nytimes.com/2012/02/19/magazine/shopping-habits.html.
15 hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem.
16 www.bbc.co.uk/news/business-56718036.
17 O'Neil, C. (2016) *Weapons of math destruction: How big data increases inequality and threatens democracy*. London: Allen Lane.
18 Eubanks, V. (2017) *Automating inequality: How high-tech tools profile, police, and punish the poor*. New York: St. Martin's Press.
19 Noble, S.U. (2018) *Algorithms of oppression: How search engines reinforce racism*. New York University Press.
20 policy-futures.centre.uq.edu.au/files/16650/Trust%20in%20AI%20Global%20Report_2023_UQ.pdf.
21 Stahl, B.C. et al. (2023) 'A systematic review of artificial intelligence impact assessments'. *Artificial Intelligence Review, 56*, 12799–12831.
22 Calvo, R.A., Peters, D., & Cave, S. (2020) 'Advancing impact assessment for intelligent systems'. *Nature Machine Intelligence*, 2(2), 89-91.

Given these concerns, various governments and supra-national organisations such as the European Union,[23] European Commission,[24] the OECD,[25] the United Kingdom,[26,27] Canada,[28] and the United States[29] have adopted an impact assessment approach and begun to develop guidelines and principles for responsible and ethical AI and algorithmic application for the public sector and private industry to follow. Interest groups have begun to lay out various forms of AI impact assessment and governance frameworks, such as the Ada Lovelace Institute,[30] AI Now[31] and the Center for Long-Term Cybersecurity.[32] A small number of private organisations are also starting to advertise AI impact assessments as a service for clients considering adopting AI technologies into their businesses.

However, such guidelines are at an early stage of development, are only mandatory within public sector environments at best, and relatively unknown in private industry and the wider impact assessment community. At present, the AI impact assessment landscape is both fragmented[33] (in the sense that it lacks common standards), internally focussed (in the sense that the frameworks currently in place are based on self-evaluation), and narrow in scope (in the sense that they concentrate primarily on technological and legal risk management and play little attention to social risk management).

Surveying a wide range of AI impact assessment frameworks across government, private and non-profit sectors, it is found that most AI frameworks list as their primary concerns technological robustness and legal/ethical compliance. However, as wider adoption of AI starts to have a more substantial impact on our everyday lives, social and economic impacts stand to become more prominent, especially in terms of the transformation of employment and work and the risks involved with such a transformation. It is clear then that the developing field of AI impact assessment should not only address potential concerns around technological robustness, safety, legal and ethical concerns, but also the wellbeing of individuals, groups and larger society being affected by AI adoption. Corporate social responsibility in this context can include the economic, social and mental health impact on employees, greater polarisation of income and wealth, issues around equality and diversity, and impacts on customer, user or client relations.

AI impact assessment should not only address potential concerns around technological robustness, safety, legal and ethical concerns, but also the wellbeing of individuals, groups and larger society.

Repercussions of overlooking these responsibilities and potential risks can result in loss of trust, income and reputational damage for a company or organisation at a smaller scale, and widespread social disruption at a larger scale. Thus, this article argues that it is necessary to encourage and implement a widespread programme of AI impact assessment in the near future to encourage sharing the benefits of AI technology while mitigating the risks as much as possible. Such an approach will not only mitigate risk but clear the path for innovation and AI adoption in the long term.

23 www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf.
24 ec.europa.eu/futurium/en/european-ai-alliance/ai-impact-assessment-code-conduct-0.html.
25 oecd.ai/en/catalogue/tools/algorithmic-impact-assessment-tool.
26 publications.parliament.uk/pa/cm5803/cmselect/cmsctech/1769/report.html.
27 www.gov.uk/government/publications/data-ethics-framework.
28 canada-ca.github.io/aia-eia-js/.
29 www.cio.gov/aia-eia-js/#/.
30 www.adalovelaceinstitute.org/resource/aia-user-guide/.
31 ainowinstitute.org/publication/algorithmic-impact-assessments-report-2.
32 cltc.berkeley.edu/wp-content/uploads/2021/08/AI_Risk_Impact_Assessments_BRIEF.pdf.
33 ecnl.org/news/eu-ai-act-must-have-standardised-methodology-impact-assessments.