

Algorithmic Impact Assessment for an Ethical Use of AI in SMEs

Shefeh Prisilia Mbuy
School of Computing and Mathematics
Keele University, UK
s.p.mbuy@keele.ac.uk

Marco Ortolani
School of Computing and Mathematics
Keele University, UK
m.ortolani@keele.ac.uk

With the ever increasing reliance on artificial systems to automate the decision process of smart systems that have the potential to affect our daily lives, the question of how to attribute liability is becoming more and more relevant, especially when human control over technical systems is increasingly reduced. This study aims to provide an overview on algorithmic impact assessment for socio-technical systems, with a focus on the challenges for its adoption by small and medium enterprises.

Automated Decisions Systems, Impact Assessment, AI Ethics, Socio-technical systems

1. INTRODUCTION

Automated decision systems (ADSs) have attracted widespread interest thanks to their ability to effectively manage large amounts of information at a rate that surpasses human capabilities. Even though they are a key enabling factor for the implementation of autonomous artificial systems, the human factor cannot be eliminated altogether as decision making in real-life contexts is inherently a socio-technical process. This of course raises interesting liability challenges when human control over the system is increasingly reduced. Examples of use of ADSs are abundant both within the public as well as the private sector, where the race to push the boundaries further in order to achieve technological superiority is constantly accelerating, with industries making huge investments in areas ranging from healthcare, finance, media, and manufacturing to surveillance, education and procurement. For instance, ADSs have been used to implement risk assessment algorithms for the criminal justice system, to manage critical infrastructure through AI-driven resource allocation, and to enhance employment and educational procedures by way of automated evaluation tools and matching algorithms (Washington 2018; Guzman et al. 2016; Van Esch et al. 2019). A report by Gartner stated that the global AI software market is forecast to reach \$62 billion in 2022 (an increase of over 21% from 2021)¹. Despite the favourable reception, questions have been raised about the *impact* that such systems are having on people's lives, their *transparency* and their

trustworthiness when used in such complex social contexts (Whittaker et al. 2018).

As a consequence of these concerns, the need for an "ethics of AI" has emerged with the aim to reign in the unintended and potentially harmful effects of algorithms on vulnerable individuals or groups. The challenge for society in general and for businesses in particular lies in finding a way to promote the design and adoption of ethically aligned AI systems even when the commercial incentives for such an approach are yet to be translated into return on investments. In their pursuit to find a way to make ADSs more accountable to the public, researchers, developers, and regulators have been increasingly focused on *algorithmic impact assessment* (AIA) as a practical tool to mitigate the risks and potential harms created by such tools. Unfortunately, consensus has yet to be reached not only on how to implement such tools but also on the very definition of what constitutes impact and how this can be measured in the context of ADSs. It is worth noting that the General Data Protection Regulation (GDPR) has focused mainly on the aspect of *data protection* (with the provision for Data Protection Impact Assessments under its Article 35) and on the right of individuals to be provided with an *explanation* of an algorithmic decision, and only more recently the focus has shifted towards the need to provide systemic accountability tools (Kaminski and Malgieri 2020; Cobbe et al. 2021). UK has incorporated the directives of GDPR into its law, and has produced a guidance for public sector organisations on how to use data appropriately and responsibly

¹A summary is available online [here](#).

when using or designing new services (UK Data Ethics Framework²). Embedding those directives into everyday practice is not straightforward, as AIA tools are bound to be intrinsically highly context-specific. This includes the identifications of harms, and of the resulting impact, as well as the data used to make the assessment. Nonetheless, a common ground for any AIA process must be the focus on *documentation* as well as on *participation* (Ada Lovelace Institute 2022). In other words, AIA is a co-creation process where decision making must not overlook outside perspectives, including the experiences of those who will be impacted by the algorithmic deployment. This clearly suggests that a *human-in-the-loop* approach must guide the development of any meaningful AIA tool.

Many large tech companies, including Microsoft, Google, IBM, Salesforce, are increasingly focusing on these topics, often establishing internal divisions on AI ethics. On the other hand, the practical effects of those regulatory provisions in small and medium enterprises (SMEs) are still unclear. As detailed in a recent report (Ada Lovelace Institute 2021), “AIAs also have scope for adoption within private-sector institutions, under the condition of regulators and public institutions incentivising their adoption and compelling their use in certain private sector contexts. Conversely, AIAs also help provide a lens for regulators to view, understand and pass judgement on institutional cultures and practices”.

With this paper, we aim to provide an overview on the proposed definition of impact in the context of socio-technical systems, and on the current proposals for tools of algorithmic impact assessment. We will also analyse how current technology may be leveraged to effectively implement a human-in-the-loop for AIA systems, and finally we will analyse the effects that their adoption might have on SMEs.

2. WHAT CONSTITUTES IMPACT?

The concept of measuring the impact of technology on its users is not a new one. Assessment tools are used extensively in various domains to mitigate risks and as a form of ensuring accountability of the system under examination. Some examples include the Privacy Impact Assessment (PIA), Environmental Impact Assessment (EIA), National Environmental Protection Act (NEPA), Human Rights Impact Assessment (HRIA), Fiscal Impact Assessment (FIA). With the advancement of public discourse on ADSs, some proposals addressing *algorithmic impact assessment* specifically have also been put forward, such as the EU Commission’s

Assessment List for Trustworthy Artificial Intelligence (ALTAI), and the Data Protection Impact Assessment (DPIA). A conspicuous amount of work is reported in the literature showing that we are in fact awash with statements of ethical principles and guidelines. However, while there is general agreement about the need to build upon concepts like responsibility, transparency, fairness, privacy, sustainability and explicability (Yeung 2020; Hagendorff 2020; Jobin et al. 2019; Floridi 2019), their full adoption and embedding into the design and implementation of algorithmic systems is yet to be seen (Morley et al. 2019). What is worse, there appears to be a lack of consistency even in the use of basic terminology, starting from the very definition of what constitutes impact.

When assessing socio-technical systems, one needs to consider the various stages of the system’s lifecycle and the multiplicity of actors involved, ranging from decision makers, to developers, and users. Each of these key stakeholders needs to feel the obligation to explain and justify all decisions concerning the design or use of system and the subsequent effects of them (Wieringa 2020). *Impact* can thus be defined as a *proxy* for social or material *harms* deriving from the use of the system (Metcalf et al. 2021). Machine learning (ML) is the practical tool at the core of most ADSs, and provides an exemplary case study of unforeseen harms (Suresh and Guttag 2021). Modern ML algorithms typically operate by learning models from existing data, in a way that can be generalised to unseen data. All development stages, from data gathering, to model construction and deployment can give rise to potentially harmful downstream consequences. In the recent past, we have witnessed examples in diverse contexts including facial recognition – where publicly available algorithms performed significantly worse on dark-skinned women (Buolamwini and Gebru 2018) – and risk assessment for the criminal justice system – where an algorithm used in court was more likely to incorrectly predict a high risk of recidivism for black defendants (Angwin et al. 2016). In all those cases, the trust of the user in the decision system is severely undermined and the system produced a harm that clearly greatly impacted the users’ lives.

Although impact assessment cannot directly mitigate or address identified harms, a meaningful assessment process can provide information upon which other interventions or processes can build. (Selbst 2021) argues that impact assessments have two primary aims. First, to get system developers to consider the potential impact and mitigate risks and the second, to capture decisions made as part of system development for accountability and to inform

²<https://www.gov.uk/government/publications/data-ethics-framework>

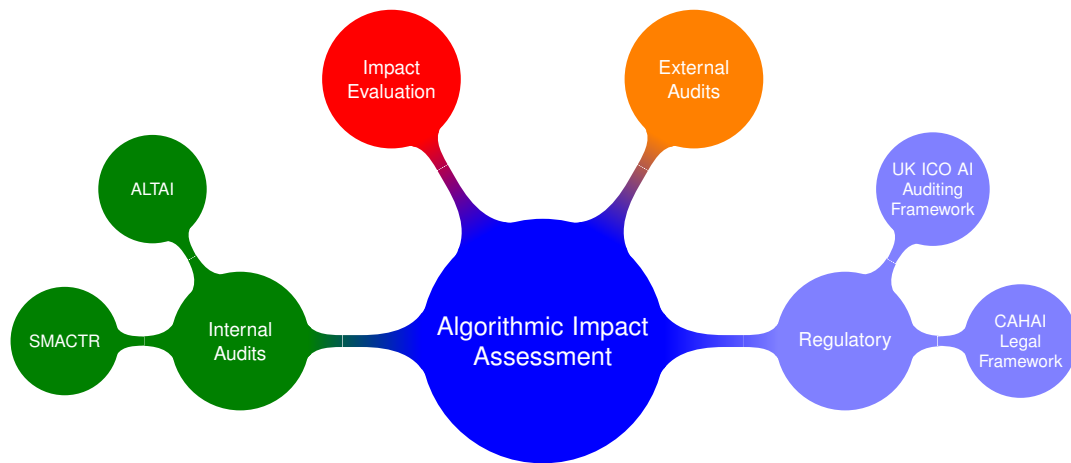


Figure 1: A taxonomy of current proposals for AIA.

future policy. Planning a response to potential harmful impacts helps to hold the responsible parties accountable for them. Metcalf et al. (2021) stated that assessing impact is “establishing an accountability relationship”. A huge contributory factor as to why there is still lack of consistency about what constitutes AIAs can be attributed to the complexity and “black box” nature of most current ADSs. Whittaker et al. (2018) pointed out that “opaque software tools work outside the scope of meaningful scrutiny and accountability. This is concerning, since an informed policy debate is impossible without the ability to understand which existing systems are being used, how they are employed, and whether these systems cause unintended consequences”.

Finally, a key consideration regards *who* defines what impact actually is in a specific context. It is crucial to ensure that those who will experience the impact are represented in the process of defining it. This process may well be complex as it spans social, technological, legal and political aspects. Furthermore, while impact might be linked back to well-recognised factors such as the choice of model or the training data, specific factors also need to be taken into account, such as the domain within which the ADS is deployed in.

3. THE LANDSCAPE OF AIA TOOLS

A review of publicly available AI ethics tools conducted in 2019 identified over 106 methods and tools available to software developers, engineers, and designers to assist them in applying AIA to system design (Morley et al. 2019). Many more tools have since been documented as part of a drive to move from highly abstract principles to practical implementation of tools to build systems aligned to AI ethical principles. Figure 1 shows the main

categories under which these current AIA proposals can be grouped into.

One of the core questions is whether impact assessment is performed *ex ante* (before development) or *ex post* (at deployment time). In the former case, the assessment is a prediction about the risks and consequences of a proposed system and might benefit from simulations in controlled testing environments, whereas in the latter case the assessment is a record of information obtained following the introduction of the system in the field, and may use field observations, interviews with stakeholders, or measurements of system outcomes in real environments (Watkins et al. 2021).

Another dichotomy is between *internal* versus *external* audits as a tool for AIA. Raji et al. (2020) define a framework called SMACSTR intended to support AI system development throughout the development lifecycle. The framework consists of 5 distinct stages: Scoping, Mapping, Artifact Collection, Testing and Reflection, and the documentation produced at each of these stages constitutes the final audit report. The authors argue that while traditionally external audits which take place post deployment are usually the preferred approach for conducting audits, an internal audit conducted as part of the system development lifecycle provides opportunities for identifying potential ethical concerns prior to deployment. Furthermore, internal auditors would have access to parts of the system and associated information which might typically not be made available to external auditors due to trade secrets; thereby potentially leading to some ethical concerns being identified which might not be apparent as part of an external audit process. The authors’ conclusion is that while there are valid concerns about the effectiveness of internal audits due to factors such as conflicts of interest, their use in other domains has proven to add value to a broader quality assurance process.

In terms of regulatory approaches, the Canadian Government has implemented an AI impact assessment tool which applies to Canadian agencies and departments (The Government of Canada 2018). The tool is based on a questionnaire and is used to understand and manage the risks associated with AI systems. A scoring system is used based on the responses to the questions to determine a Risk Impact Score which is mapped to an impact level ranging from level 1 (little to no impact) to level 4 (very high impact). The assessment scores are based on many factors including system design, algorithm model used, decision type, impact and data. Remedial actions must be taken to mitigate the anticipated harms if the scores fall within a given band. The momentum for AIAs has recently gained traction in the USA as well, where a revised version of the Algorithmic Accountability Act is under discussion in Congress. The proposal would require impact assessments from companies using automated systems to make critical decisions. The burden of releasing reports for public scrutiny would fall on the Federal Trade Commission. Within the European Union, the AI Act has been proposed by the European Parliament, as a complement to the existing GDPR regulation. The proposal does not contain explicit provision for algorithmic impact assessment, although it does require conformity assessments, but has been met with some criticism due to the “lack of clarity concerning its scope” and the fact that it may not provide sufficient protection against the actual and potential harms generated by ADSs, thereby leaving too much discretion to system providers (Smuha et al. 2021). The EU Artificial Intelligence High Level Expert Group (AI HLEG) have defined a self-assessment tool called the Assessment List for Trustworthy Artificial Intelligence (ALTAI Tool). This is intended to help business and organisations to self-assess the trustworthiness of their systems throughout development. The tool offers guidance based on 7 key requirements: Human Agency and Oversight, Technical Robustness and Safety, Privacy and Data Governance, Transparency, Diversity, Non-discrimination and Fairness, Societal and Environmental Wellbeing, and finally Accountability. While the ALTAI tool is similar to the SMACTR internal framework for algorithmic auditing, it does not use a scoring system and the questions are grounded on the concept of Fundamental Rights i.e. the protection of people’s fundamental human rights as referred to in the EU Treaties.

While the variety of proposals might appear to offer a wide choice to developers willing to implement AIA within their process, what still seems to be missing though from an examination of the current landscape of methodologies and frameworks is a thoroughly integrated approach. A lot of the proposals to date

have focused so far on the pre-deployment phase which understandably provides opportunity for early intervention in identifying and mitigating potential harms and unintended outcomes. However, an impact assessment approach involving monitoring of the ADS system over time might help detect unexpected outcomes and uncover actual harms which were not identified during the project design and implementation phase. An AIA-by-design approach would therefore provide opportunity to identify and mitigate potential harms and unintended impact resulting from future system maintenance and enhancements. Another key consideration regards *measuring* the impact. Several AIAs to date have adopted a quantitative approach using defined metrics, often with reference to individual aspects, such as algorithmic fairness, accountability, transparency (Beutel et al. 2019; Metcalf et al. 2021; Thomas and Uminsky 2022). However not all harms can be mapped to a metric and therefore there is a risk that solely measuring evaluative metrics which do not reflect actual harms does not offer a comprehensive approach for an AIA process. Using questionnaires provides a reasonably quick method (approximately 35 minutes for the Canadian Government tool) and structured approach and guarantees some level of consistency in the way the assessment is conducted across different systems. However, pre-set questions lack the flexibility to assess aspects of the system that cannot be pre-determined. Finally, when AIA relies on self-assessment, the people conducting the audit are part of the organisation being audited. As a result, there is a risk that an internal audit might not be seen as robust as one conducted by external parties. Organisations conducting self-assessments should acknowledge the risk of conflicts of interest and take steps to mitigate them. Failing to manage this risk can lead to perceptions of “ethics washing” or on a more serious basis could undermine the process and lead to reputational damage (Morley et al. 2019). In order to avoid this, (Raji et al. 2020) propose that auditors need to “be mindful of their own and the organizations’ biases and viewpoints”. Lessons of how self-assessments are organised and conducted from other well established internal audits in industries such as finance, automobile, and aviation can be adopted for the AIA process. Cross-disciplinary expertise can also help in identifying a wider range of impacts therefore ensuring the best possible chance to uncover impacts that may otherwise not be considered. (Watkins et al. 2021) argue that assessing impact through the lens and perspectives of various expertise is crucial in order to construct the relationship between impact and harm; however, assembling such diverse expertise is challenging, especially for smaller businesses.

Table 1: Publicly available tools for AIA.

Organisation	Framework
Microsoft	MS Responsible AI
IBM	AI Fairness 360
Google	Facets, MMD-critic
EU HLEG	ALTAI framework

4. A PARTICIPATORY APPROACH TO AIA

Despite the growing interest in AIA from academia and industry alike, a crucial point that most current proposals seem to overlook is the fact that accountability cannot be achieved by technical means alone. Rather, it requires a substantial shift in the company's mindset, and the introduction of a shared *culture of assessment*. This is eloquently highlighted in (Ada Lovelace Institute 2022) which points out that when AIA processes are primarily controlled by the very people who drive the decision-making process, and insufficient emphasis is given to outside perspectives, AIA risks adopting a partial or inconsistent view of potential impacts. On the contrary, what is needed is a *participatory* effort involving all interested stakeholders, from developers, to internal and independent assessors, to end users. Although the focus of the report by the Ada Lovelace Institute is on a case study regarding the healthcare sector, it contains some considerations of general relevance. First of all, AIA tools need to promote *reflexivity*, i.e. an ongoing process of examination of the company's own practices, motives and beliefs; they need to be open to *independent scrutiny* and finally they need to be *transparent*, not only by documenting the full AIA process but also by making sure that its results are publicly available. Similar points are also made by (Raji et al. 2020), where internal auditing is promoted as a practice to allow all stakeholders to hold the company accountable for the tools they use.

While the adoption of a fully participatory approach fosters a more independent and unbiased review of the impacts of ADS, it raises the additional issue of how to effectively embed the diverse sources of knowledge into the AIA process, and how to share its outcome with the interested stakeholders. Crucially, this requires keeping the human (whether domain expert, developer or end user) in the loop. Many organisations are already creating AIA tools under the "responsible AI" agenda, some of which have been made publicly available as open-source toolkits (see Table 1). The vast majority of these tools are aimed at software developers as well as assessors, auditors and ethics AI officers, but typically each of them separately addresses individual aspects of the assessment and lack

systematic instruments for feedback provision, thus inhibiting the reflective approach that should be part of AIA. In this context, conversational agents, such as chatbots (Adamopoulou and Moussiades 2020), could prove effective to fill this gap. While the use of chatbots is well documented in many fields, ranging from education, to entertainment, e-commerce or health (Shawar and Atwell 2007), to the best of our knowledge their adoption has not yet been proposed in the context of AIA. On one hand, the well-established use of AI techniques – such as Natural Language Processing (NLP) or sentiment analysis – in chatbots makes them excellent candidates to establish a natural form of communication even with non-expert users (Khanna et al. 2015; Ciechanowski et al. 2019), thus promoting the transparency that is advocated for AIA tools. On the other hand, chatbots can incorporate existing proprietary knowledge in the form of ontologies (Al-Zubaide and Issa 2011) in order to drive the assessment process. Chatbots might for instance be trained using existing documentation, such as public policies and regulations as well as internal policies of the specific organisation. The conversational process would facilitate the presentation of information to the different categories of users in order to solicit the most accurate response. Chatbots could also control the storage and analysis of user feedback, promoting the use of a standardised format, regulating who will have access to the information and finally, how the output will be presented in a transparent and interpretable manner. Finally, the adoption of chatbots might prove helpful in facilitating the translation of AIA tools from one domain to another, which is a scenario of particular interest at the onset of more widely adopted regulatory approaches to AI as will be discussed in the following section.

5. AIA AND THE PRIVATE SECTOR

As discussed in Section 3, most of the approaches to AIA have so far focused on the public sector in an attempt to introduce a verifiable degree of accountability to the use of automated systems in public service decision making. Even when the legislation currently under discussion in the US or EU will be in force, SMEs investing in creating or using ADSs will probably be excluded from the strictest form of compliance detailed in the regulations. However, voluntarily adoption of a responsible approach to implementing AI will be certainly encouraged and many companies may choose to follow suit. The question that we want to discuss here is whether AIAs designed for public sector accountability are also suitable for the private sector.

The challenges facing the private sector – and SMEs specifically – are different to public sector organisations. The main dilemma for small businesses and start-ups consists in how to approach AIA in a structured and cost-effective way as no established tool or methodology has yet emerged as a reference standard and is not likely to appear in the near future. Considerations related to costs in terms of time and effort will also be a key factor in the AIA process. For this to be wholeheartedly adopted, the process would have to take into account the impact on time to market. An added delay to normal business procedures would inevitably increase project costs and would be perceived as a potential threat to the company's competitiveness. This could happen for instance when the window of optimum product launch is missed due to perceived additional project tasks related to the AIA process.

Another non negligible factor affecting the adoption of AIAs in the private sector is the required expertise. As previously stated, engaging a cross disciplinary expertise and diverse expertise is crucial as this will ensure that different perspectives can contribute to identifying impacts. However, the private sector and in particular SMEs may not have the knowledge or indeed the capability to identify the cross disciplinary expertise that may be required for the AIA process. Even if they understand expertise that are required, the cost implication might be a hindrance. While for public sector projects this may not necessarily be an issue (as it can be argued that such costs would help to bring accountability to the use of AI in public services). However, this is an entirely different scenario for the private sector.

Finally, it is crucial to consider the timing aspect. Many of the current proposal fall into the *ex ante* category, which might sound appealing to SMEs as it is less likely to disrupt the development phase and on the overall costs. Moreover such approach holds the potential for unintended impact to be identified early during system design and therefore creates the opportunity to take steps to address these prior to system deployment. However, this leaves room for other relevant unintended impacts to go unnoticed and unreported, as they may be related to previously unidentified risks, system maintenance and enhancements, as well changes to intended use cases. In this case, rather than protecting the end user, the AIA would end up creating a false sense of security and might even obfuscate liabilities. Therefore, in order to ensure that the impact of ADSs can be monitored on a regular basis (including post deployment), we recommend that an ex-post approach is also required. This would allow for real impact to be identified once the system is deployed in the environment of its intended use. The challenge

is that once again, for smaller organisations, the requirement to factor in an AIA process as part of an on-going operational activity in order to monitor deviations from an ethics perspective over time and take remediation for an ADS means securing the expertise required, time and effort and budget; all of which could be deemed as 'unattractive'.

6. CONCLUSIONS

As awareness of the social and ethical risks of ADSs is increasingly being discussed and questions are being asked by the public on the impact, transparency and trustworthiness of these complex systems, regulation is struggling to keep the pace with the technological advances. Tools for algorithmic impact assessment are thus suggested as a means of promoting accountability.

We provided an overview of AIA and noted how they have so far focused mainly on the public sector. Nevertheless, we anticipate a need for them to be ported to the private sector as well. When SMEs in particular will find themselves in the position of having to, or choosing to, comply to the new regulatory frameworks, the AIA process will need to be adapted to recognise their specific and sometimes unique challenges. As we have outlined here, these will likely include time, resource, and expertise which all relate to cost.

Given that SMEs will likely adopt AIAs on a voluntary basis at least in the first stage, coupled with the fact that efficacy of AIAs is yet to be effectively demonstrated, we have discussed how the process will need to be less burdensome and streamlined as much as possible in order to promote the design and development of ethically aligned ADSs.

REFERENCES

- Ada Lovelace Institute (2021). Technical methods for the regulatory inspection of algorithmic systems in social media platforms. Technical report, Ada Lovelace Institute.
- Ada Lovelace Institute (2022). Algorithmic impact assessment: a case study in healthcare. Technical report, Ada Lovelace Institute.
- Adamopoulou, E. and L. Moussiades (2020). Chatbots: History, technology, and applications. *Machine Learning with Applications 2*, 100006.
- Al-Zubaide, H. and A. A. Issa (2011). Ontbot: Ontology based chatbot. In *International Symposium on Innovations in Information and Communications Technology*, pp. 7–12. IEEE.

- Angwin, J., J. Larson, S. Mattu, and L. Kirchner (2016). Machine bias. In *Ethics of Data and Analytics*, pp. 254–264. Auerbach Publications.
- Beutel, A., J. Chen, T. Doshi, H. Qian, A. Woodruff, C. Luu, P. Kreitmann, J. Bischof, and E. H. Chi (2019). Putting fairness principles into practice: Challenges, metrics, and improvements. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 453–459.
- Buolamwini, J. and T. Gebru (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*, pp. 77–91. PMLR.
- Ciechanowski, L., A. Przegalinska, M. Magnuski, and P. Gloor (2019). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems* 92, 539–548.
- Cobbe, J., M. S. A. Lee, and J. Singh (2021). Reviewable automated decision-making: A framework for accountable algorithmic systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 598–609.
- Floridi, L. (2019). Translating principles into practices of digital ethics: Five risks of being unethical. *Philosophy & technology* 32(2), 185–193.
- Guzman, A., S. Ishida, E. Choi, and A. Aoyama (2016). Artificial intelligence improving safety and risk analysis: A comparative analysis for critical infrastructure. In *2016 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pp. 471–475. IEEE.
- Hagendorff, T. (2020). The ethics of ai ethics: An evaluation of guidelines. *Minds and machines* 30(1), 99–120.
- Jobin, A., M. Ienca, and E. Vayena (2019). The global landscape of ai ethics guidelines. *Nature Machine Intelligence* 1(9), 389–399.
- Kaminski, M. E. and G. Malgieri (2020). Multi-layered explanations from algorithmic impact assessments in the gdpr. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 68–79.
- Khanna, A., B. Pandey, K. Vashishta, K. Kalia, B. Pradeepkumar, and T. Das (2015). A study of today's ai through chatbots and rediscovery of machine intelligence. *International Journal of u-and e-Service, Science and Technology* 8(7), 277–284.
- Metcalf, J., E. Moss, E. A. Watkins, R. Singh, and M. C. Elish (2021). Algorithmic impact assessments and accountability: The co-construction of impacts. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 735–746.
- Morley, J., L. Floridi, L. Kinsey, and A. Elhalal (2019). From what to how: An initial review of publicly available ai ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics* 26(4), 2141–2168.
- Raji, I. D., A. Smart, R. N. White, M. Mitchell, T. Gebru, B. Hutchinson, J. Smith-Loud, D. Theron, and P. Barnes (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 33–44.
- Selbst, A. D. (2021). An institutional view of algorithmic impact assessments. *Harvard Journal of Law & Technology* 117(35), 75.
- Shawar, B. A. and E. Atwell (2007). Chatbots: are they really useful? In *Ldv forum*, Volume 22, pp. 29–49.
- Smuha, N. A., E. Ahmed-Rengers, A. Harkens, W. Li, J. MacLaren, R. Piselli, and K. Yeung (2021). How the eu can achieve legally trustworthy ai: A response to the european commission's proposal for an artificial intelligence act. Available at SSRN 3899991.
- Suresh, H. and J. Gutttag (2021). A framework for understanding sources of harm throughout the machine learning life cycle. In *Equity and Access in Algorithms, Mechanisms, and Optimization*, pp. 1–9.
- The Government of Canada (2018). Algorithmic impact assessment tool.
- Thomas, R. L. and D. Uminsky (2022). Reliance on metrics is a fundamental challenge for ai. *Patterns* 3(5), 100476.
- Van Esch, P., J. S. Black, and J. Ferolie (2019). Marketing ai recruitment: The next phase in job application and selection. *Computers in Human Behavior* 90, 215–222.
- Washington, A. L. (2018). How to argue with an algorithm: Lessons from the compas-publica debate. *Colorado Technology Law Journal* 17, 137.

Watkins, E. A., E. Moss, J. Metcalf, R. Singh, and M. C. Elish (2021). Governing algorithmic systems with impact assessments: Six observations. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 1010–1022. Association for Computing Machinery.

Whittaker, M., K. Crawford, R. Dobbe, G. Fried, E. Kaziunas, V. Mathur, S. M. West, R. Richardson, J. Schultz, and O. Schwartz (2018). *Ai now report 2018*. Technical report, AI Now Institute at

New York University.

Wieringa, M. (2020). What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, pp. 1–18.

Yeung, K. (2020). Recommendation of the council on artificial intelligence (oecd). *International legal materials* 59(1), 27–34.