SMES AND EXPLAINABLE AI: AUSTRALIAN CASE STUDIES

EVANA WRIGHT,[*] JIANLONG ZHOU,[†] DAVID LINDSAY,[‡]
LINDA PRZHEDETSKY,[§] FANG CHEN[**] AND ALAN DAVISON [††]

ABSTRACT

There is little understanding of the difficulties small to medium enterprises (SMEs) encounter in ensuring that the AI systems they develop, or use, are ethical. SMEs' are less likely than larger businesses to have the resources or time to familiarise themselves with ethical AI principles or how those principles should be applied in practical contexts. This paper reports the results of qualitative research conducted with Australian SMEs and start-ups that design and/or utilise AI technologies as part of their core business practices with a focus on the principle of explainability. The study identified a high level of inconsistency in both attitudes to ethical AI and to practices for implementing ethical AI within businesses in the interviewed SMEs. The paper identifies initiatives that may be implemented to promote greater understanding by SMEs of ethical AI principles, in particular, the principle of explainability.

CONTENTS

[*] Senior Lecturer, Faculty of Law, University of Technology Sydney. The authors acknowledge the valuable contribution of Dr Althea Gibson to the research project in her capacity as Research Assistant. The research outlined in this paper was conducted according to UTS ethics approval UTS HREC ETH21-6172.

[†] Associate Professor, Data Science Institute, Faculty of Engineering and IT, University of Technology Sydney.

[‡] Professor, Faculty of Law, University of Technology Sydney.

[§] PhD Candidate, University of Technology Sydney.

[**] Distinguished Professor, Data Science Institute, Faculty of Engineering and IT, University of Technology Sydney.

[††] Professor, Faculty of Arts and Social Sciences, University of Technology Sydney.

I INTRODUCTION

Micro, small and medium-sized enterprises ('SMEs') are the 'economic backbone' of many countries and economies.[1] For instance, SMEs represent 99 per cent of all business in the European Union[2] and create nearly two-thirds of new private sector jobs in the USA.[3] However, due to fewer financial resources and the prevalence of economies of scale, SMEs are typically slower to adapt to information and communication technologies than larger companies.[4] Consequently, in the context of significant recent advances in Artificial Intelligence (AI), and its wide-scale deployment, governments and international bodies have recognised the need for special measures to support the adoption and use of AI systems by SMEs.[5]

For example, in its 2020 *Recommendation on Artificial Intelligence,* the OECD emphasised the importance of national AI policies and international cooperation paying 'special attention' to SMEs,[6] including support to facilitate the ethical and trustworthy development, implementation and use of AI.[7] The European Commission's 2021 proposal for an *AI Act* points to the importance of removing barriers to the adoption of AI by SMEs and the need for national governments to develop initiatives targeted at small-scale providers and users of AI systems, including awareness-raising initiatives.[8] Moreover, in Europe, several states have established 'Digital Innovation Hubs' where SMEs can access technical expertise and experiment with AI technologies, and the EU and member states have committed to investing € 1.5 billion to roll out the hubs further.[9] Similarly, the Australian *AI Action Plan,* released in June 2021, incorporates the establishment of a National AI Centre,

---

1 Emil Blixt Hansen and Simon Bøgh, 'Artificial intelligence and internet of things in small and medium-sized enterprises: A survey' (2021) 58 *Journal of Manufacturing Systems* 362, 362.

2 European Commission, 'Internal Market, Industry, Entrepreneurship and SMEs: SME definition', (Web Page) <https://ec.europa.eu/growth/smes/sme-definition_en>.

3 Office of the United States Trade Representative, 'Small and Medium-Sized Enterprises' (Web Page) <https://ustr.gov/trade-agreements/free-trade-agreements/trans-pacific-partnership/tpp-chapter-chapter-negotiating-8>.

4 Organisation for Economic Cooperation and Development, *ICT, E-Business and Small and Medium Enterprises* (OECD Digital Economy Papers No 86, 2004).

5 See, e.g, House of Lords, Select Committee on Artificial Intelligence, *AI in the UK: Ready, Willing and able?* (Technical Report, 2018).

6 Organisation for Economic Cooperation and Development, *Recommendation of the Council on Artificial Intelligence* (2019) OECD/LEGAL/0449, rec V.

7 Ibid.

8 European Commission*, Proposal for a Regulation of the European Parliament and of the Council: Laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts*, Doc No COM(2021) 206 final, 21 April 2021 *('Proposed AI Act')*, Recitals (72) – (73).

9 European Commission, *Fostering a European approach to Artificial Intelligence,* COM(2021) 201 5 final, 21 April 2021, 22.

which will specifically address barriers facing SMEs in adopting and developing AI, as well as four AI and Digital Capability Centres, which are intended to assist SMEs in accessing AI technologies and expertise.[10]

Despite an understandable focus on the problems facing SMEs in adopting and using AI systems, this has yet to be matched by a similar level of attention to the particular difficulties SMEs encounter in ensuring that the AI systems they develop or use are ethical. Given SMEs' constraints, they are less likely than larger businesses to have the resources or time to familiarise themselves with ethical AI principles or how those principles should be applied in practical contexts. If, however, SMEs are to deploy AI systems successfully, they need to be adequately equipped to address the challenges of ensuring that the systems, and the way they are used, comply with ethical principles.

A pre-requisite for promoting ethical AI for SMEs is building an understanding of the current level of knowledge, and engagement with, ethical AI by SMEs. This paper reports the results of qualitative research conducted with Australian SMEs and start-ups that design and/or utilise AI technologies as part of their core business practices. The research was specifically directed at investigating the understanding among SMEs of ethical issues relating to the explainability of AI systems, which is one of the foundation issues in ethical AI. The overall objective of the research was to provide a baseline of information relating to the approaches and attitudes of SMEs to ethical AI to be used as part of the overall project of translating ethical AI principles into practice. The study identified a high level of inconsistency in both attitudes to ethical AI and to practices for implementing ethical AI within businesses in the interviewed SMEs. Ethical considerations are often viewed as secondary to other business priorities, and additional resources are required to support the adoption of ethical AI principles by SMEs in Australia.

II THE ETHICAL PRINCIPLE OF 'EXPLAINABILITY'

One of the fundamental problems posed by the significant advances in non-symbolic or statistical AI systems is that – due to reliance on complex 'black box' functions - it can be difficult or impossible to determine how an output is produced.[11] It is, therefore, unsurprising that the majority of the many statements of principles of ethical AI incorporate a version of the principle that, in certain contexts, it must be possible for AI systems to be satisfactorily explained to humans. For example, in one of the most commonly cited surveys of AI ethical principles, Jobin et al. concluded that transparency was the 'most prevalent principle'.[12]

There are, however, many variations in how this principle is expressed; and seemingly intractable terminological confusion. For example, the European Commission's High Level Expert Group (HLEG)'s *Ethics Guidelines for Trustworthy AI* incorporates the principle of 'explicability' as one of four fundamental ethical AI principles, but when operationalising the principle, it effectively translates it into the practical requirement of 'transparency'.[13] According to the HLEG, the principle of 'explicability' means 'that processes need to be transparent, the capabilities and purpose of AI systems openly communicated, and decisions

---

[10] Australian Government, *Australia's AI Action Plan* (June 2021) 12.

[11] See, e.g., Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez and Javier Del Ser et al., 'Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI' (2020) 58 *Information Fusion* 82.

[12] Anna Jobin, Marcello Ienca and Effy Vayena, 'The global landscape of AI ethics guidelines' (2019) *Nature Machine Intelligence* 389, 391.

[13] European Commission High-Level Expert Group on Artificial Intelligence, *Ethical Guidelines for Trustworthy AI* (2019).

– to the extent possible – explainable to those directly and indirectly affected'.[14] The HLEG Guidelines further divide the requirement of transparency into the following three elements:

(i) traceability—the ability to 'trace back' the data, model, rules and recommendations of an AI system;

(ii) explainability—the ability to 'explain both the technical processes of the AI system and the reasoning behind the decisions or predictions that the AI system makes'; and

(iii) open communication about the limitations of the AI system—this includes advising users that they are interacting with an AI system and informing them of the purpose, criteria and limitations of the decisions generated by the system.[15]

Australia's *Artificial Intelligence Ethics Framework,* on the other hand, treats the principles of transparency and explainability as effectively co-extensive, providing that:

> There should be transparency and responsible disclosure so people can understand when they are being significantly impacted by AI, and can find out when an AI system is engaging with them.[16]

In effect, the Australian framework reduces the principles to a requirement of 'responsible disclosures', which 'should be provided in a timely manner, and provide reasonable justifications for AI system outcomes'.[17]

The term 'explainability' (or sometimes 'explicability') is part of a cluster of related terms, including 'transparency', 'interpretability', 'understandability' and 'intelligibility', which are sometimes used interchangeably and sometimes distinguished. The terms are commonly organised hierarchically, with 'transparency' often appearing as an umbrella term. For example, commenting on the approach taken in statements of ethical AI principles, Jobin et al. observed that '[r]eferences to transparency comprise efforts to increase explainability, interpretability or other acts of communication and disclosure'.[18] This usage is reflected in the European Commission's proposed *AI Act,* which uses 'transparency' as an umbrella term, requiring that '[h]igh risk AI systems shall be designed and developed in such a way to ensure that their operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately'.[19]

However, terms other than transparency also appear as umbrella terms. After noting the inconsistent use of terminology, a 2019 UK House of Lords committee report adopted the general term 'intelligibility' to apply to both 'technical transparency' and 'explainability'.[20] According to the report, 'technical transparency' means ensuring that experts understand an AI system, including how and why outputs are produced. On the other hand, the report confined 'explainability' to ensuring that AI systems 'are developed in such a way that they can explain the information and logic used to arrive at their decisions'.[21] In a 2018 paper, however, Floridi et al. had a different take on 'intelligibility': the paper applies the catch-all term 'explicability' in both the epistemological sense of 'intelligibility', relating to how an AI

---

[14] Ibid 13.

[15] Ibid 14–15.

[16] Australian Government, *Australia's AI Ethics Principles* (7 November 2019).

[17] Ibid.

[18] Jobin Ienca and Vayena (n 12).

[19] *Proposed AI Act* (n 8) art 13(1).

[20] House of Lords, Select Committee on Artificial Intelligence, *AI in the UK: Ready, Willing and able?* (Technical Report, 2018) 36.

[21] Ibid 39.

system works, and the ethical sense of 'accountability', relating to responsibility for how an AI system works.[22]

This distinction between epistemological and ethical usages partially explains the terminological confusion: it is one thing to focus on the technical features of an AI system and quite another to focus on human understanding of, or responsibility for, technical systems. Another source of the confusion is the extent to which the problem is approached from different disciplinary perspectives, including computer science, human-computer interaction, psychology and law.[23] This paper returns to the problem of how to more precisely formulate the principle of explainability after introducing three specific issues arising from ethical obligations to provide an explanation: the circumstances in which an explanation may be required; the context-dependent nature of explanations; and the potential for trade-offs between explainability and accuracy. In these sections, we use the term 'explainability' indiscriminately to encompass other related terms, such as 'interpretability'.

A *When is an Explanation Required?*

An important threshold question when considering the explainability principle is: when should an explanation be required? Different approaches may be taken to specify the circumstances in which an explanation is needed. Most approaches apply a framework based on the assumption that the higher the risk posed by an AI system, the greater the need for an explanation. There are, however, differences in conceptualising risks. It has, for example, been argued that the obligation to provide an explanation should depend on the domain in which the AI system is used. For instance, the UNESCO Ad Hoc Expert Group on the Ethics of AI states that the principle of explainability is more important when an AI system is used in a high-risk 'domain' such as law enforcement, security, education, recruitment or health care.[24]

Other approaches define risks by reference to the impacts on affected persons, and especially impacts that affect an individual's rights or interests. For example, the European Commission's proposed *AI Act* defines 'high risk' AI systems as 'systems that pose significant risks to the health and safety or fundamental rights of persons'.[25] Similarly, the Australian Human Rights Commission (AHRC) has made the point that 'it is good practice to provide reasons for decisions that affect a person's legal or similarly significant rights, regardless of the status of the decision maker and even where there is no legal requirement to provide reasons'.[26]

The considerable difficulties in specifying the circumstances in which an explanation may be required, which extends to difficulties in predicting risks, emphasises the need for most AI systems to be potentially explainable. That said, counter-veiling considerations must be considered in both imposing obligations to provide an explanation and determining the

---

[22] Luciano Floridi, Josh Cowls, Monica Beltrametti et al, 'AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations' (2018) 28(4) *Minds and Machines* 689, 700.

[23] Amina Adadi and Mohammed Berrada, 'Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)' (2018) 6 *IEEE access* 52138; Tim Miller, 'Explanation in Artificial Intelligence: Insights from the Social Sciences' (2019) 267 *Artificial Intelligence* 1.

[24] UNESCO Ad Hoc Expert Group (AHEG) for the preparation of a draft text of a recommendation on the ethics of artificial intelligence, *First Draft of the Recommendation on the Ethics of Artificial Intelligence* SHS/BIO/AHEG-AI/2020/4 REV.2 (7 September 2020) [71], 14.

[25] *Proposed AI Act* (n 8) 1.1.

[26] Australian Human Rights Commission, *Human Rights and Technology* (Final Report 2021) ('*Human Rights and Technology*') 81.

form an explanation should take. First, as explained further below, there may be trade-offs in design decisions in developing AI systems between explainability and other values, such as efficiency or accuracy. Secondly, both designing explainable systems and explaining AI systems imposes costs.[27] Thirdly, complete transparency may be undesirable to the extent to which it may result in the release of trade secrets or personal data.[28]

This indicates that considerable care is needed in determining the circumstances in which an explanation should be provided and the form an explanation should take.

B *Explanations Necessarily Depend upon Context*

It is widely accepted that explanations of AI systems are highly contextual. For example, in their guidance on explaining decisions made with AI, the UK Information Commissioner's Office (ICO) and the Alan Turing Institute note that several factors, such as the application of the AI system, the type of data involved, the setting, and the individual recipient of the explanation all 'affect what information an individual expects or finds useful'.[29] The guidance further points out that organisations should tailor explanations to their audience so that they 'avoid creating explanation fatigue … (by saying too much) and at the same time allow … [organisations] to protect … [their] intellectual property and safeguard [their] system from being gamed'.[30]

Similarly, Preece et al. argue that the question of whether AI is explainable cannot be answered before answering the question 'explainable to whom?' Accordingly, they point out that explainability means different things to system creators, system operators, those making decisions based on AI systems, those affected by AI decisions, and those whose data has been used in AI systems and system regulators.[31] Dawson et al. take this further by pointing to the different purposes of different audiences. Accordingly, they observe that while explanations for users of AI systems may focus on what the system is doing and why, explanations for creators may aim to explain how the system is working for validation or certification, and explanations for the general public may aim to build user trust and confidence in AI systems.[32]

Similarly, in a detailed discussion, Zhou and Danks distinguish the goals of different groups, which they label 'engineers', 'users' and 'affectees'.[33] For example, they argue that an 'affectee' (such as a person whose face is recognised by a facial recognition system) may simply require a non-technical 'difference-based intelligibility', namely 'an input-output characterization of the decision processes embodied by the algorithm, thereby reducing uncertainty and demonstrating the reliability of the system'.[34] On the other hand, they claim that 'users' of an AI system require 'function-based intelligibility', which entails more

---

[27] Adadi and Berrada (n 23).

[28] *Human Rights and Technology* (n 21) 66.

[29] Information Commissioner's Office and The Alan Turing Institute, *Explaining decisions made with AI* (20 May 2020), 21.

[30] Ibid 49.

[31] Alun Preece, Dan Harborne, Dave Braines et al, 'Stakeholders in Explainable AI' (Conference Paper, AAAI FSS-18: Artificial Intelligence in Government and Public Sector Conference, 2018) <https://arxiv.org/abs/1810.00184v1>.

[32] D Dawson, E Schleiger, J Horton et al, *Artificial Intelligence: Australia's Ethics Frameworks* (Data61 CSIRO, 2019).

[33] Yishan Zhou and David Danks, 'Different "Intelligibility" for Different Folks' (Conference Paper, AAAI/ACM Conference on AI, Ethics and Society, 7-8 February 2020).

[34] Ibid 196.

information about the operation of the AI system, including information about the inputs it requires to operate and 'appropriate conditions for its use and adaptation'.[35] This information can be supplied by AI developers in design documents. Finally, AI engineers require 'causal-process intelligibility', which is the type of intelligibility that has been the focus of much of the scholarly literature.[36] This involves information about 'computational architecture, specific models, parameter values, internal states and their relationships … [and] hardware or user interface constraints'.[37]

Apart from distinctions based on the purposes of the recipients of an explanation, distinctions have been drawn between different types of explanation. For example, the ICO and the Alan Turing Institute's guidance on explainability notes that explanations can be either 'process-based' - that is, they can provide information demonstrating responsible design and deployment of an AI system - or 'outcome based'. The guidance identifies six main types of explanation: rationale explanations; responsibility explanations; data explanations; fairness explanations; safety and performance explanations; and impact explanations.[38] According to the guidance, the particular type of explanation that may be required depends on the use of the AI system. For instance, if an AI system was used to process applications for a job, an unsuccessful applicant may wish to know that they have not been discriminated against (i.e., they may require a 'fairness explanation'). Alternatively, a patient who has received a medical diagnosis generated by an AI system will wish to know that the diagnosis is accurate (a 'safety and accuracy explanation').[39] In yet another taxonomy, computer science scholars Vilone and Longo note that the existing research on the explainability of AI systems has identified the following different types of explanations intended to fulfil different purposes:[40]

- Traced-based explanations—for system designers
- Reconstructive explanations—for end users
- Mechanistic explanations—how does it work
- Operational explanations—how do I use it
- Ontological explanations—describe the structural properties of the system
- Teaching explanations
- Introspective tracing explanations
- Introspective informative explanations
- Post hoc explanations
- Execution explanations.

Regardless of the 'type' of explanation or the recipient's purpose, the quality of the decision will dictate the extent to which a decision is 'explainable'. Meske et al. observe that consideration of the quality of an explanation involves consideration of multiple factors, such as plausibility, comprehensibility, interpretability, fairness, and privacy.[41] Gilpin et al. have

---

[35] Ibid 197.

[36] Ibid.

[37] Ibid.

[38] Information Commissioner's Office and The Alan Turing Institute, *Explaining decisions made with AI* (20 May 2020) 20.

[39] Ibid 52-54.

[40] Guilia Vilone and Luca Longo, 'Explainable Artificial Intelligence: A Systematic Review' (2020) <https://arxiv.org/abs/2006.00093>.

[41] Christian Meske, Enrico Bunde, Johannes Schneider and Martin Gersch, 'Explainable Artificial Intelligence: Objectives, Stakeholders, and Future Research Opportunities' (2022) 39(1) *Information Systems Management* 53, 57-58.

linked the quality of an explanation to the level of understanding of the recipient of such an explanation, claiming that a good explanation has been provided when the recipient 'can no longer keep asking why'.[42]

Using examples and making relationships or causal links explicit will also improve the quality of an explanation.[43] Vilone and Longo observe that 'it is part of human nature to assign causal attribution of events' and, as such, explanations of AI systems 'must make the causal relationships between the inputs and the model's predictions explicit'.[44] Similarly, Graaf and Malle argue that, as people often regard AI systems as operating with human-like intention, it is important that explanations fall within 'the bounds of the conceptual and linguistic framework' used to explain human behaviours.[45] After reviewing over 250 social science publications on explanations, Miller concluded that explainable AI researchers should consider that effective explanations are generally selective, 'contrastive' (that is, they should explain why one event happened instead of another) and focus more on causal links than on probabilities.[46]

In their guidance, the ICO and the Alan Turing Institute advise that organisations can 'layer' their explanations by first providing individuals with priority explanations, and then making additional explanations available in further layers.[47] The guidance also advises that explanations should be conceptualised as a two-way conversation and use visual aids such as 'visualisation media, graphical representations, [or] summary tables' where appropriate.[48]

C *The Trade-off Between Explanation and Accuracy*

It is commonly argued that there is a correlation between the accuracy and complexity of AI systems and that this has a negative impact on the 'explainability' of a decision. As Burrell observes, '[m]achine learning models that prove useful (specifically, in terms of the 'accuracy' of classification) possess a degree of unavoidable complexity'.[49] For this reason, many commentators observe that there is an unavoidable trade-off between performance and explainability - '[o]ften, the highest performing methods (e.g., DL [deep learning]) are the least explainable, and the most explainable (e.g., decision trees) are the least accurate'.[50] The design of an AI system will dictate the extent to which a system is explainable, and 'trade-offs

---

[42] Leilani H. Gilpin, David Bau, Ben Z. Yuan et al 'Explaining Explanations: An Overview of Interpretability of Machine Learning' in *2018 IEEE 5th International Conference on Data Science and Advanced Analytics* (IEEE, 2018) 80–89.

[43] Katie Atkinson, Trevor Bench-Capon and Danushka Bollegala, 'Explanation in AI and law: Past, present and future' (2020) 289 *Artificial Intelligence* 103387.

[44] Vilone and Longo (n 40) 8.

[45] Maartje M. A. de Graaf and Bertram F Malle 'How People Explain Action (and Autonomous Intelligent Systems Should Too)' in *2017 AAAI Fall Symposia* (AAAI Press, 2017) 19–26.

[46] Tim Miller, 'Explanation in artificial intelligence: Insights from the social sciences' (2019) 267 *Artificial Intelligence* 1.

[47] Information Commissioner's Office and The Alan Turing Institute, *Explaining decisions made with AI* <https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/explaining-decisions-made-with-artificial-intelligence/>.

[48] Ibid.

[49] Jenna Burrell, 'How the machine 'thinks': Understanding opacity in machine learning algorithms' (2016) 3(1) *Big Data & Society* 1, 5.

[50] David Gunning, Mark Stefik, Jaesik Choi et al 'XAI – Explainable artificial intelligence' (2019) 4(37) *Science Robotics* <doi: 10.1126/scirobotics.aay7120>; See also, Philipp Hacker, Ralf Krestel, Stefan Grundmann and Felix Naumann, 'Explainable AI under contract and tort law: legal incentives and technical challenges' (2020) 28(4) *Artificial Intelligence and Law* 415.

might have to be made between enhancing a system's explainability (which may reduce its accuracy) or increasing its accuracy (at the cost of explainability).'[51]

However, not all scholars agree that such a trade-off is inevitable. Increased explainability may, in fact, lead to increased accuracy to the extent that it helps lead to the correction of deficiencies in AI systems.[52] Cynthia Rudin goes as far as to state that the existence of any trade-off between explainability and accuracy is a myth.[53] Furthermore, she argues that the reliance on the post-facto explanation of high-stakes decisions by complex systems may be inadequate and, in fact, 'perpetuate bad practice and … potentially cause great harm to society.'[54] Instead, the focus should shift to the design of interpretable systems that will 'provide their own explanations, which are faithful to what the model actually computes.'[55] Using the surprising outcome of the 2018 Explainable Machine Learning Challenge as an example, Rubin and Radin argue that simple, interpretable models may be as accurate as more complex models while also avoiding some of the other issues that arise in relation to black box systems.[56]

III SMEs and Explainability

The literature on how SMEs approach the issue of AI ethics and explainability is limited. A recent study by Ayling found that surveyed SMEs did not view explainability as necessary beyond 'communicating with their customers about their products' as part of their sales process.[57] In contrast, Bessen et al., following a survey of 225 AI start-ups, found that 58% of surveyed companies had established a set of codified firm-level ethical AI principles but that many of these companies 'have never invoked their ethical AI principles in a costly way, such as firing an employee, dropping training data, or turning down a sale.'[58] The study established that resources are critical to the adoption and implementation of AI ethics principles. It also found that larger start-ups, companies that had collaborated with high-technology firms and companies with prior experience in implementing GDPR obligations, were more likely to have established ethical AI frameworks.[59]

In the absence of specific literature considering how SMEs approach the issue of explainability, research on general business responses to AI may provide further insights into SMEs and the adoption of AI ethics principles. In a small study of nine executive managers of businesses (of varying sizes and in a range of sectors) across Germany, Austria and Scandinavia, interviews revealed that 77.77% of interviewed managers believed that AI ethics

---

[51] European Commission High-Level Expert Group on Artificial Intelligence, *Ethical Guidelines for Trustworthy AI* (2019) 18.

[52] Arrieta et al (n 11) 83.

[53] Cynthia Rudin, 'Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead' (2019) 1(5) *Nature Machine Intelligence* 206, 207.

[54] Ibid 206.

[55] Ibid.

[56] Cynthia Rudin and Joanne Radin, 'Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition' (22 November 2019) 1(2) *Harvard Data Science Review* <https://hdsr.mitpress.mit.edu/pub/f9kuryi8/release/8>.

[57] Jacqueline Ayling, 'Putting AI ethics to work: Are the tools fit for purpose for SMEs?' (PhD Thesis, University of Southampton, 2021) 78, 101.

[58] James Bessen, Stephen M. Impink, Lydia Reichensperger and Robert Seamans, *Ethical AI Development: Evidence from AI Startups* (Working Paper, March 2022) <https://scholarship.law.bu.edu/faculty_scholarship/1188> 4, 5.

[59] Ibid 14, 17.

should be a high priority in their business.[60] However, while interviewees indicated that transparency was an important value in their business practices, some of these managers were concerned that revealing information could compromise intellectual property rights and competitive advantage.[61]

Literature suggests that the successful adoption of ethical AI by businesses requires organisational transformation. For example, an international survey of 1580 executives in 510 large companies conducted by Capgemini Research Institute showed that 77% of executives are uncertain about the ethics and transparency of their AI systems.[62] The same survey showed that concern about using AI systems influenced strategic business decisions: '41% of senior executives report that they have abandoned an AI system altogether when ethics concerns were raised; 55% implemented a "watered-down" version of the system'.[63] The Capgemini report is important for this paper for two reasons: one, it is likely that executives in SMEs are also uncertain about AI ethics and explainability, and two, the adoption of AI systems and thus purchasing decisions by potential customers, may be influenced by the extent to which AI systems are explainable.

Explainability is especially important to SMEs for many reasons. As discussed above, at a macro level, explainability helps ensure that AI systems remain accountable – that is, they can be audited and their accuracy assessed. As argued by Bauer et al., explainability also assists humans to 'widen their horizons of reasoning and understanding'.[64] From a business perspective, explainability helps to ensure customer trust and satisfaction.[65] KPMG notes that 'organisations must think about the governance of algorithms to build trust in outcomes and achieve the full potential of artificial intelligence'.[66] Failure to ensure AI systems are explainable may expose SMEs to 'financial, reputational, and regulatory risks'.[67]

Customer satisfaction is particularly vital to SMEs, which depend on repeat business far more than multinational enterprises, which tend to have large customers.[68] Further, SMEs have fewer resources to invest in technology and governance systems. By ensuring that AI systems are explainable from the outset, SMEs can minimise the risk of having to update or upgrade systems to provide explainability. The lack of financial resources also makes it imperative that SMEs do not unnecessarily expose themselves to reputational loss or legal liability. For example, the OECD has flagged 'reputational and legal risks' as one of several

---

[60] Joseph Baker-Brunnbauer, 'Management perspective of ethics in artificial intelligence' (2021) 1 *AI and Ethics* 173, 177.

[61] Ibid 180.

[62] Capgemini Research Institute, 'Why addressing ethical questions in AI will benefit organisations' (4 July 2019) 9.

[63] Ibid 2.

[64] Kevin Bauer, Oliver Hinz, Wil van der Aalst and Christof Weinhardt, 'Expl(AI)n It To Me – Explainable AI and Information Systems Research' (2021) 63(2) *Business and Information Systems Engineering* 79, 80.

[65] See, e.g., IBM, 'Transparency and trust in the cognitive era' *IBM THINK* (Blog Post, 17 January 2017) <https://www.ibm.com/blogs/think/2017/01/ibm-cognitive-principles/>; Mary T. Dzindolet, Scott A. Peterson, Regina A. Pomranky et al, 'The role of trust in automation reliance' (2003) 58(6) *International Journal of Human Computer Studies* 697.

[66] KPMG, *Uncover the full potential of artificial intelligence* (2019) <https://assets.kpmg/content/dam/kpmg/xx/pdf/2019/02/uncover-the-full-potential-of-artificial-intelligence.pdf> 2.

[67] Ilana Golbin, Anand S. Rao, Ali Hadjaran et al, 'Responsible AI: A Primer for the Legal Community' in *2020 IEEE International Conference on Big Data* (IEEE, 2020) 2121, 2123.

[68] Dóra_Horvath and Roland Zs. Szabo, 'Driving forces and barriers of Industry 4.0: Do multinational and small and medium-sized companies have equal opportunities?' (2019) 146 *Technological Forecasting and Social Change* 119.

barriers to SMEs' use of data analytics and implementation of data solutions.[69] Accountability of AI systems, including explainability, may also be an important consideration in any potential future investment by venture capital or other external investors with ethical AI 'seen as one of the important drivers of portfolio risk and return.'[70] Potential collaborators may also require compliance with AI ethics principles, particularly with established '[h]igh technology firms, which often share their data sources with startups'.[71]

As discussed above, the OECD acknowledges that SMEs require special support to facilitate the ethical and trustworthy development, implementation and use of AI.[72] The question remains as to how SMEs may implement ethical principles in practice and how SMEs should be supported to do so. The remainder of this paper outlines the methodology and findings of qualitative research conducted with Australian SMEs and start-ups investigating the understanding among SMEs of ethical issues relating to the explainability of AI systems.

IV METHODOLOGY

The study was designed to elicit in depth qualitative information about the views, challenges and expectations of Australian SMEs[73] concerning AI ethics, with a focus on the explainability of AI systems. It was conducted in two main stages: semi-structured interviews with selected participants and a follow-up survey questionnaire.

The participants in the study were chosen based on pre-selection criteria, which ensured that they were well-placed to answer questions on the use of AI systems in their respective businesses. Most participants interviewed were involved in key business decisions relating to the selection, design, implementation, use and/or evaluation of AI technologies in their businesses. Although every participant either designed or used AI in their business, not every business intended to use AI when the business began; in several cases, businesses started using AI after they had already commenced providing a product or service.

Due to the small size of the businesses, which included start-ups, interviewees often noted that team roles and responsibilities overlapped prior to the business expanding sufficiently to engage specialists responsible for technology-related decisions. Business founders and key decision-makers came from a range of backgrounds, including business, finance, law, computer science and academia; however, every participant interviewed had, at a minimum, a baseline understanding of how AI worked in their business. The selection of a

---

[69] Organization for Economic Cooperation and Development, *The Digital Transformation of SMEs* (OECD Studies on SMEs and Entrepreneurship, 2021) ch 5.

[70] AI Asia Pacific Institute, *Transforming Ethics in AI Through Investment* (Web Page, 7 February 2021) <https://aiasiapacific.org/2020/08/20/transforming-ethics-in-ai-through-investment/>.

[71] Bessen et al (n58) 4.

[72] Organisation for Economic Cooperation and Development, *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449, 22 May 2019, s 2.

[73] There is no universal agreement as to what constitutes an SME, with definitions varying between countries and even industries. The European Commission defines an SME as an organisation with less than 250 employees and either an annual turnover of less than €50 million or a balance sheet of less than €43 million. See *Commission recommendation of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises* [2003] OJ L 124/36, art 2(1). By contrast, in the US the definition of an SME varies according to the industry in which the enterprise operates. See US Small Business Administration, 'Table of Small Business Size Standards Matched to North American Industry Classification Codes' (Web Page, 14 July 2022) <https://www.sba.gov/sites/default/files/2022-07/Table%20of%20Size%20Standards_Effective%20July%2014%202022_Final-508.pdf>. In Australia, federal legislation contains a number of different definitions of 'small businesses'. See e.g., Australian Government, *Australian Securities and Investments Commission Act 2001* (Cth), ss 12BC, 12BF; *Corporations Act 2001* (Cth) s 761G.

range of professional backgrounds was designed to ensure the broadest possible understanding of AI systems, including the benefits and risks of the technologies.

Eight participants took part in the open-ended interviews, while five completed the follow-up survey. Although this was admittedly a small sample, the study resulted in a rich collection of qualitative material concerning attitudes to ethical AI and the understanding of explainability among Australian SMEs.

A *Semi-structured Interviews*

The semi-structured interviews, based on a series of open-ended questions, consisted of Zoom sessions of approximately one hour. The interviews were subsequently transcribed using a transcription service.

The open-ended questions, which were designed to form part of a free-flowing discussion, investigated the following issues:

- the professional background of the interviewee, including their role in the business
- the nature of the business, including its clients or customers
- the use of AI in the business, including how AI is developed and/or used in products or services
- the participant's understanding of 'explainability' in the context of the use of AI in the business
- whether the 'explainability' of AI was regarded as important by the business
- whether, in what circumstances, and how, the business explained its AI systems
- what was entailed by an explanation
- the use of data in AI systems and the safeguards, such as documentation, applied by the business to the use of data
- the implications of the use of data for explainability
- the use of the AI model (or algorithm) by the business and the safeguards, such as documentation, applied by the business to the model
- understanding of any potential trade-offs between explainability and accuracy
- the business risks entailed in providing an explanation, including potential disclosure of trade secrets
- responsibility within the organisation for AI governance, including responsibility for ensuring ethical AI and addressing potential problems
- organisational understanding of ethical AI, especially as applied to SMEs
- particular areas of uncertainty in complying with ethical AI
- whether there is a specific need for guidance about ethical AI for SMEs.

B *Follow-up Survey*

Analysis of the responses to the semi-structured interviews revealed areas that required clarification or further exploration. Participants were therefore asked to complete a follow-up online survey approximately four weeks following the initial interviews. The follow-up survey canvassed the following issues:

- the risks associated with providing explanations of AI systems, including the potential for revealing trade secrets
- views about the potential role of government in ensuring ethical AI
- the most effective tools or methodologies for promoting ethical AI
- how the participant/SME interpreted 'explainability'
- where the business would be likely to seek guidance about ethical AI

- if an ethical AI framework were to be mandated, who should be responsible for creating the framework?
- which of the existing ethical AI frameworks would be preferred?
- the techniques or methodologies preferred by the business for ensuring explainability and ethical AI
- tolerance for risk associated with practices that may fail to ensure explainable AI

The responses to the follow-up survey added valuable detail to the information collected from the qualitative interviews.

V FINDINGS

Following analysis of the qualitative interviews and the follow-up survey, the responses were grouped into the following overarching themes:

- organisational knowledge and awareness of AI ethics
- problems with data quality
- approaches to selecting and implementing AI models
- attitudes to AI assurance processes, including documentation and auditing
- organisational AI governance mechanisms
- approaches to understanding 'explainability' and implementing explainable systems
- understanding of potential trade-offs between 'explainability' and accuracy
- concerns about risks of explaining AI systems
- approaches to 'explainability for SMEs'
- approaches to the role of government in ensuring ethical AI

Each of these themes is expanded upon immediately below.

A *Knowledge and Awareness of AI Ethics*

The interviews revealed considerable variation in the level of knowledge and understanding the SMEs expected employees to have concerning ethical AI. While some businesses suggested that their staff inherently knew about AI ethics ('the machine learning guys all know the consequences of getting some of the stuff wrong'), some expected staff who join the business to have learnt about AI ethics through formal study ('people who are involved on the AI part have at least a master's of data science and AI from one of the good Australian universities, so it's part of their curriculum in any case. So we are not training them; we expect them to have that insight and training before they come and work with us').

There were, however, notable differences in the level of in-house training provided on ethical AI: while some proactively trained staff in AI ethics ('so as part of the onboarding we have a section on talking about data and we … [consider that] … the team we have should treat that data as our own data'), some were completely unaware of ethics training ('I've haven't had or seen any ethics training'), while others failed to see the value of AI ethics training ('obviously, no amount of training can teach someone to be ethical').

B *Data Quality*

Statistical AI systems, including machine learning systems, are highly dependent on the quality of the data sets used to train the systems. Yet leveraging data was critical to most of the businesses that participated in this study. For example, the business model adopted by a number of the SMEs is based on deriving value from previously underutilised data. As one

interviewee put it, 'data is a pure margin product', adding that much data is effectively costless to collect.

Nevertheless, a strong theme emerging from the interviews is that SMEs implementing AI systems face considerable challenges with data quality. The difficulties encountered are illustrated by the following comments:

- 'the data is notoriously bad quality'.
- 'we had quite a lot of issues with just dealing with the varying quality of imagery that we had'.
- '[the data was] poor quality so we had to kind of get rid of the whole data set'.
- 'it's the only dataset we've got, so it's either going to work or it isn't'.

Overall, the responses indicated more focus on the practical difficulties of 'wrangling' data than on ethical dilemmas relating to low quality data. That said, there was a general awareness of the need for exercising caution in working with data, with one interviewee observing, in relation to the selection of data for a machine learning system, that 'you just got to be careful how you put that into a model because it re-weights everything in its own way.'

C *Selecting and Implementing AI Models*

The SMEs that formed part of this study applied a variety of approaches to selecting and implementing AI models, with some building their own models, others modifying existing models, and yet others using 'off the shelf' solutions. The choice taken depended upon the accessibility of a suitable model for the task at hand. For example, as one interviewee explained the choice of models:

- 'We had [Microsoft] Azure Cognitive Services available to us [for this project], [so] we just let Microsoft build the detailed models. So in this case, we have not built the AI models, but in previous cases where we had done other work around deep learning algorithms, that's where we had a decision, and primarily we used an ensemble learning methodology [which combines multiple models]. So we just ran multiple algorithms and then checked out what results were most commensurate with what we wanted'.

The extent to which a business uses an 'off the shelf' model or build their own has important implications for the explainability of the AI system. For those businesses with the in-house technical skills to build their own models, model development was strongly influenced by literature reviews. The comments on model development included the following:

- 'We go and do a scan of I guess, literature - a bit like … a research project - do a mini literature review and come back and say, "Well, look. This is what from a research perspective the latest practice looks like or the latest research tells us is relevant."'
- 'We didn't build off others. We did look at literature reviews…'.
- 'I guess, [our model is] based on research about what sort of literature review of what models are being used in this general area trying to pull together combinations of natural language and existing building research'.

D *AI Assurance Processes*

The explainability of AI systems, not to mention the safety and reliability of systems, depends upon the implementation of assurance processes, such as appropriate documentation and system auditing. The study was therefore interested in the attitudes of SMEs to implementing

internal assurance processes. There was a high degree of inconsistency in the approaches taken to documentation of AI systems, with considerable differences in the interpretation of what might be required and how documentation was implemented:

- 'We document everything internally. We use Confluence, so just like a Wiki page type of tool'.
- 'There's just one document, generally. There's a project write-up at the end of each project'.
- 'We have more like system descriptions and architecture design documentation that says these are the steps that we went through, the data cleaning steps'.

Apart from the inconsistent approaches to documentation, some SMEs did not separately document processes, with one expressing scepticism about the value of documentation. These attitudes were reflected in comments such as the following:

- 'All of the documentation as far as the model itself is concerned would be contained in the code'.
- 'We didn't document stuff, because who are we documenting it for? It's kind of a backend system'.

Among the interviewees, there was general scepticism about the value of auditing AI systems, at least in the context of current AI auditing practices:

- 'I've not met an AI auditor. I've met lots of people who make up these checklists, that haven't got a clue what's going on inside the technology'.
- 'I think the AI audit fraternity needs to upskill before it can start actually … providing a valid service around what they're looking into, because the guys I've spoken to so far just don't have a clue what's under the hood. They're just really making questions off feature lists'.
- 'At the end of the day, the audit is only as good as … the person doing it and the day that they did it on'.

These comments appear to reflect concerns that the practice of AI auditing is not well-established and that third party auditors do not have sufficient expertise to adequately assess a business's AI system. That said, some interviewees indicated that they incorporated independent expert evaluation in their system development, with one reporting that 'we have an industry expert that looks at the results and works way back'. Another interviewee reported that they had engaged one of the big four consulting firms to audit their system, which included providing the firm with access to the code.

E *AI Governance Mechanisms*

The interviews explored a range of issues relating to the internal governance mechanisms of SMEs. Understandably, the SMEs forming part of the study had flat structures, which conditioned their governance mechanisms. The majority of the interviews were conducted with the founders of the SMEs, with some being conducted with data scientists. Typically, the founders had considerable technical expertise and were initially either closely involved with, or intimately aware of, the governance of AI systems used in their business.

In most cases, given the flat structures, the founders or CEOs were regarded as ultimately responsible for the AI systems deployed by the business. The following comment was typical: '(t)he buck kind of stops with me in terms of the ownership of whether or not it's ethical'. However, where data scientists were employed in-house, they commonly shared responsibility, with one data scientist commenting, 'I'm responsible ultimately for any changes or code that gets deployed into a production environment'. Referring to the shared

responsibility, another data scientist observed, '(t)hey [ie. management] shouldn't be asking for something that's unethical and I shouldn't be giving it to them'.

Regardless of perceptions of ultimate responsibility, the scale of the businesses meant that there was considerable overlap in roles and responsibilities, with a range of people assuming roles in using or modifying algorithms.

F *Approaches to Explanations*

There were significant variations among interviewees in the approaches taken to the 'explainability' of AI systems. A number of interviewees accepted that, in the context of their business, it was important to provide users with an explanation. The following are examples of comments that emphasised the importance of explanations:

- '… if you make a decision based on an algorithm's recommendation, that decision … has to be explained'.
- '… from our perspective, we have to make sure that … the way those models are working, if a human were to look at them, that it would make sense and is interpretable'.
- '…we have to really think about what we are giving the users and [about] the explanation. So we add an additional explanation…'.
- 'I think everyone should be given an explanation: everyone who is part of that journey or interacts with [the AI system] or is affected by it'.

On the other hand, some interviewees questioned whether it was always possible or desirable to provide an explanation. Comments to this effect included the following:

- '… black box algorithms like neural networks are very difficult because … [from] … the internal architecture … it is really difficult to explain how we came to that conclusion'.
- 'But does it matter? Does it matter that you can say why it came up with a certain result? I don't think it does that much'.
- 'We actually didn't do explainability. We had a black box opaque AI system …'.

The interviewees generally regarded explainability as subservient to overarching business goals. Therefore, explainability tended to be valued if it could build trust in the business. As one interviewee put it, 'explaining… why they have been recommended what they have been recommended, what the logic behind that is… [can make users] feel more assured'. Similarly, another interviewee commented:

- '… we have the human element in there, which is actually our competitive advantage. So our algorithm may not be superior but the way we present information is far superior [to] anyone else.'

Against this, as illustrated by the following comments, several interviewees considered that users or customers do not care about explainability:

- 'Query how much a client cares once a [practitioner says], 'This is good to go''.
- 'You've just got to tell them what it does. No one really cares about what's underneath it all, they care about what's the output and why that output came to be'.

Two strong themes to emerge were the perceptions that, first, providing an explanation would increase the information burdens placed on users or customers and, secondly, that users or customers are more interested in the accuracy of systems than in understanding precisely how they work. The following comments illustrate these themes:

- 'I mean, I do have the sense, more and more actually over time, that people are just experiencing tremendous information overload. And an AI system being explicable, I mean, it's just more data to grok'
- 'I think maybe in the B2C market, you can just say that you have AI and people just generally accept … [that] … because it's all about value'
- 'So I don't think just dumping them with a lot of technical jargon is going to help'.

Concerns about how users might respond to explanations also influenced the approaches taken by interviewees to the form in which explanations should be given. The various approaches taken by interviewees to how explanations should be given appear from the following comments:

- '… it would be just a human explaining the things that are clearly marked as automated. We are quite careful to mark clearly what is and isn't automated in our document selection. We have a few different categories of decision, default decisions, automated decisions, user decisions, system decisions. And we make very clear distinctions … between those'.
- 'We give a popup saying very specifically this is what it does. It … [asks]… do you still want to use it? So [we] give the users the choice of using the algorithm and so it's [an] instant thing on an app that they can yes say, "Yes, I know it's very clear." [In addition] … we give explanation videos so that gives them [some] context'.
- 'It needs to be short and sweet'. If it is pictorial, just one scan and … [they] … get it. Brilliant'.
- 'What we did instead was we showed … what … was recommended. We didn't explain why. And we also showed them how they performed on those recommendations'.
- 'They kind of just want a human being who knows the area to give them a simple explanation that's germane to them specifically. I think that's a lot of what people are paying for'.
- '… I don't know whether we need to explain the whole AI algorithm, how it works. I think it just needs to be … because you said this, we are recommending this. … Explaining the frequency and why we [are] saying it, but at a high level, rather than saying here is our algorithm and this is how it works, because not everyone needs to know and not everyone will follow. And they don't care, to be honest. They just want to know that there is some science in … this'.
- 'Look, what I try to do is I use a lot of analogies [that are used] at the moment because it's difficult for people to understand what the impact [is]'.
- '… typically what we try to do is we actually try to verbally explain [the system] and if there's any publicly available … [information] … we try to give that away and then in summarised format …'.
- 'Generally we'd like to … [provide an explanation] … face to face and talk people through, say, a presentation which then we would be happy to leave with them'.

G *Trade-offs Between 'Explainability' and Accuracy*

As the study was concerned with the value SMEs place on explainability compared with potentially competing objectives, it questioned interviewees about their approaches in the event of a trade-off between explainability and accuracy. As explained immediately above, one of the themes to emerge from the study was that interviewees generally tended to

consider that their customers valued accuracy over explainability. The following comment captures this general attitude:

- '… as long as the model is proven accurate and is working the way it wants, then generally you don't care how it came to that decision. It's similar to … [how] … a lot of people driving a car don't really care how the car works, as long as it works'.

The overarching concern among interviewees, as reflected in the following comment, seemed to be the need to assure customers of the accuracy of a system rather than trouble them with the details of how it works:

- 'The reality of it is, I think what the market wants from us is real accuracy. I have been surprised that it's like once people are paying any kind of money, the expectation of accuracy is very, very high'.

That said, some interviewees considered that providing an explanation can assist in assuring customers that the system is reliable and accurate. As one interviewee put it:

- 'And … [what] … I always just ask clients is, "is it better to have a model that you know predicts extremely well but I can't explain to you or is it better to have a model I can explain to you but doesn't predict as well?" And … they always get torn on that. ... The way I see it is that we want to try and get the most predictive model that we can, … [while] … understanding that sometimes to get people comfortable with the idea that … [the system] … is doing what is expected, we need to give them some level of explainability'.

H *Risks of Explaining AI Systems*

As reported above, most of the interviewees regarded explainability as subsidiary to the wider commercial goals of their business. This was particularly evident in general concerns that trade secrets would be compromised if too much information were to be provided about the algorithm. Comments reflecting this concern included the following:

- 'I would have concerns about sharing trade secrets, especially if it's a competitive advantage. If it's a means to an end and everyone is doing it and I'm borrowing from the world and giving it back to the world, I will have no problem. But if it's something that's proprietary, it's something that we have done and it gives us a competitive advantage, then that's something we don't want to erode'.
- 'It's very tough from an industry competition standpoint to make many of these decision-making processes comprehensible to your competitors by publicly disclosing them, or even your customers who may well be a mystery shopper from your competitor. I have real doubts, just from a purely commercial standpoint, about explaining too much how we do stuff, because I'm not really interested in competitors taking off with the ideas'.
- 'I think actually, the transparency required for people to understand it, you would have to disclose commercially sensitive trade secrets. It's a very tough problem'.

At the extreme end, one interviewee went so far as to say, that disclosing too much information to a client may mean they have 'designed a business for doomsday'.

While a minority of businesses dealt with the risks by not revealing any information about the AI system, the majority adopted strategies aimed at minimising the risks. One strategy, for example, was to provide a verbal summary ('we actually try to verbally explain it and if there's any publicly available things we try to give that away and then in summarized format…'). Another strategy was to limit the amount of information provided by the customer interface ('… we are seriously considering taking it … [explainability information]

… off the interface, because it gives away a lot to potential competitors without providing a great deal of value to people who just don't seem to want to really know'). The majority of SMEs interviewed therefore tended to regard the disclosure of information about their AI systems as a balance between revealing enough to assure customers while not revealing information that could advantage competitors.

I *Explainability for SMEs*

The interviews aimed to identify the specific issues faced by SMEs. To begin, most interviewees agreed that SMEs should comply with ethical principles. A good example of this was the observation that:

> '… there should be a requirement to comply, because you can't just say, "Well, it applies to large enterprises and then small businesses can do without it," because then that's not appropriate'.

That said, there was general agreement that SMEs face particular challenges in complying with ethical AI principles, arising principally from resource constraints. This led to suggestions that a degree of flexibility is required in how SMEs comply with ethical principles, including the principle of explainability. The following comments were typical of these attitudes:

- '… there needs to be some freedom for smaller companies, because they would die otherwise. They would not be able to innovate, because we don't have the resources to do what big companies can do'.
- 'It's just, how do you make … compliance workable for a small business?'
- '… I think the reality of startups is that startups are there to scale and make money and have a massive valuation. So they'll try to get to the quickest spot to POC [proof of concept] or MVP [minimum viable product], all right? So …[this] … means that ethics and all of those things may not be the priority…'
- '… being an SME, short of resources all the time … there was a million competing priorities. Being transparent wasn't the high on the list'.
- 'I think the small business don't have the support and the finance and the team and the experience to deal with it and to get around it. So large companies can get around it'.

While there was agreement both that SMEs should comply with ethical principles and that they face particular challenges in doing so, there was considerable uncertainty about the flexibilities that could assist SMEs in complying. As one interviewee put it, 'it's … hard to know what the compliance would look like'. One suggestion was that compliance thresholds might be flexible ('I am a big fan of thresholds for complicated compliance'). There was, however, general agreement among interviewees about the potential benefits of greater guidance on how SMEs can comply with ethical AI principles. As illustrated by the following comments, those supporting guidance emphasised the importance of a sufficient level of detail:

- 'I'd love to see … guidance because below the concepts of AI is all the detail, and that's where it all goes wrong'.
- '… to the extent that guidance is provided to small businesses, it should be super, super, super example heavy. And I can say that … the best thing you can show a client from any kind of regulator are just case study after case study after case study'.

Although guidance was supported, it was also generally regarded as less important than other measures ('I think … guidance [is] definitely helpful but it's more [important for other]

support'). The most common response to the challenges facing SMEs was to suggest mechanisms for supporting them to comply, especially by providing financial incentives or support, such as tax breaks. In this respect, the following comments were typical:

- 'Take, for example, tax. Everyone has the same tax rules, whether you are a small company or big companies. But then the government makes exceptions for small companies'.
- 'There's a lot of information but other than saying that you could go to jail for data handling, data privacy and data breach - the big sledgehammer - but there's nothing else to support. Where is the program? I don't even know if there's a program for a startup to say, "Oh, how can I be AI ethics compliant?"'
- 'There's so many people giving advice. There're so many people giving information sessions but what are you actually doing? Are you providing funding? Is the federal government giving a tax break…?'
- '…through the incubators and startup [subsidies] have things in there saying, "Here's the program, we will provide you an independent assessment to do that for free of charge or a tax break or we'll give you and help you to get down and look at your tools and your practices and help you to develop a data governance strategy and implement it for you free of charge or a tax break."'

J *Role of Government*

The interviewees held some common attitudes to the potential role of government in promoting or ensuring ethical AI. As explained above, there was a view that government could provide more assistance to SMEs in ensuring ethical AI. As one interviewee put it:

> 'If the ethics is important to the governments around AI, because it's the future, … and someone cares deeply enough that it needs to be done in the right way, then I think [there is] some policy benefit [in greater] … access to resources'.

The follow-up survey indicated general support for government involvement in developing ethical AI principles. There was, however, far less agreement about government involvement in more proactive regulation. From the interviews, one interviewee raised the possibility of government providing a form of 'certification that could be used in marketing', while another raised the prospect of voluntary regulation. However, none of the businesses that participated in the follow-up survey believed that government should develop AI-specific laws or regulation.

VI Analysis of Findings

SMEs are increasingly using AI systems, promising socially beneficial innovation. However, this raises significant questions about how SMEs can comply with ethical AI principles, especially in the face of resource constraints. This study examined the approaches and attitudes of selected Australian SMEs to ethical AI, focusing on the ethical principle of explainability. In general, there was a high level of inconsistency in both attitudes to ethical AI and to practices for implementing ethical AI within businesses. The overall impression is that SMEs and start-ups are largely charting their own path with very limited assistance; understandably, they regard ethical considerations as secondary to other business priorities.

While the interviewees displayed a general grasp of ethical AI issues, such as problems relating to data quality, there was a lack of detailed knowledge of ethical principles. This extended to limited familiarity with specific statements of ethical principles, such as the principles promoted by the Australian government. This appears to be associated with the considerable variations in the approaches to implementing ethical AI within the businesses.

For example, there were no consistent approaches to in-house training or where to seek guidance on ethical AI, whether within the business or externally.

The inconsistencies were particularly apparent in the approaches to ethical AI safeguards, such as documentation and third party auditing. As a generalisation, documentation practices appeared to vary widely, while scepticism was expressed about the value of third party auditing. Given the flat structures of SMEs, internal accountability mechanisms tended to be informal, with ultimate responsibility for systems residing with founders, CEOs and/or data scientists. In general, there was no formal allocation of responsibility for ensuring ethical AI to a particular officer or organisational unit.

Even though most interviewees saw a need for some degree of explainability, there was considerable variation in the approaches taken to providing explanations: some businesses invested in providing additional information about the AI, while others avoided providing explanations. Generally, an explanation was considered important solely to the extent it was believed to deliver value to the business by, for example, building trust. On the other hand, where a business believed that customers are more interested in results than in how a system operates, explanations were not provided. Overall, the interviewees were cautious about burdening customers with too much information and believed their customers were more interested in accuracy than explainability. Therefore, where explanations were provided, some care was taken with the form of explanation, with simple or concise summaries being preferred and often provided face-to-face. This general approach also reflected concerns that providing too much information could reveal trade secrets or otherwise benefit competitors.

The interviewees generally considered that SMEs should comply with ethical AI principles, but were acutely aware of the particular challenges facing SMEs in implementing ethical AI. This led to calls for some leeway in how SMEs might be required to comply with ethical principles, but also for government assistance to aid SMEs with compliance. While the interviewees supported greater guidance on ethical AI principles, such as using case studies, they generally considered that financial assistance, such as tax relief or financial support for assessing AI systems, would likely be more critical. In general, the SMEs considered that there was a legitimate role for government in developing ethical AI principles and providing guidance. On the other hand, responses indicated a concern that other forms of government involvement, such as through regulation, could impose unnecessary costs on SMEs.

While some of the inconsistencies in the approaches and attitudes of the SMEs may reflect the different industries and AI systems the interviewees are involved with, it seems likely that there are more generally applicable explanations. As it remains early in the adoption of AI systems by SMEs, there is still a lot of trial and error in how AI is being implemented, with considerable variations between businesses. For many start-ups and SMEs, the priorities of the business will largely reflect those of the founder or founders. More importantly, start-ups and SMEs are overwhelmingly concerned with building and maintaining their business, often in highly competitive markets, with other concerns being subsidiary. While larger enterprises have the resources to address broader concerns, SMEs must direct their resources to ensuring financial viability. These considerations strongly suggest that, unaided, it is likely that the attitudes and practices of SMEs will continue to exhibit an undesirably high degree of inconsistency.

VII Conclusion

Acknowledging the limited nature of this study, we consider it reasonable to conclude that there is an unarguable case for governments to pay greater attention to promoting ethical AI

among SMEs, including in promoting greater understanding of the principle of explainability. In particular, we believe there is a case for initiatives such as the following:

- Greater efforts are required to educate SMEs about ethical AI, including proactively disseminating information about ethical AI principles through SME networks, as well as developing detailed case studies illustrating how SMEs can apply the principles in practice;
- In particular, detailed guidance is needed to assist SMEs in understanding what is required for explaining AI systems, and how this can be done without exposing trade secrets or alienating customers;
- While there is a general need for more resources to be allocated for training businesses in ethical AI, there is a specific (and we think pressing) need for training to be targeted to the particular challenges faced by SMEs. Given the resource constraints faced by SMEs, there may be a case for financial aid to support SME participation in training initiatives;
- More guidance is needed in relation to the standard business safeguards reasonably expected from SMEs using AI systems, especially those systems that may significantly impact people. These safeguards should include standardised expectations relating to documentation, as well as standardised processes for assessing AI systems;
- As a step towards greater consistency in the practices adopted by SMEs, some consideration could be given to adopting a government-backed certification scheme for ethical AI. Such a scheme, while not a magic bullet, could assist with transparency and standardisation of practices, including practices relating to explaining AI systems; and
- Greater standardisation also seems to be required in relation to the qualifications and practices of third party auditors of ethical AI. This may require some form of certification.

In drawing these conclusions, we stress that the sorts of measures contemplated will require significant investment by government including funding to develop and deliver resources and support to SMEs. The findings of this research should inform the program of activities of the National AI Centre and the associated AI and Digital Capability Centres to be established by the Australian Government under the Australian AI Action Plan.