Building Trustworthy AI Solutions: A Case for Practical Solutions for Small Businesses

Keeley Crockett, *Senior Member, IEEE*, Edwin Colyer, *Senior Member, IEEE*, Luciano Gerber, and Annabel Latham, *Senior Member, IEEE*

***Abstract*—Building trustworthy artificial intelligence (AI) solutions, whether in academia or industry, must take into consideration a number of dimensions including legal, social, ethical, public opinion, and environmental aspects. A plethora of guidelines, principles, and toolkits have been published globally, but have seen limited grassroots implementation, especially among small- and medium-sized enterprises (SMEs), mainly due to the lack of knowledge, skills, and resources. In this article, we report on qualitative SME consultations over two events to establish their understanding of both data and AI ethical principles and to identify the key barriers SMEs face in their adoption of ethical AI approaches. We then use independent experts to review and code 77 published toolkits designed to build and support ethical and responsible AI practices, based on 33 evaluation criteria. The toolkits were evaluated considering their scope to address the identified SME barriers to adoption, human-centric AI principles, AI life cycle stages, and key themes around responsible AI and practical usability. Toolkits were ranked on the basis of criteria coverage and expert intercoder agreement. Results show that there is not a one-size-fits-all toolkit that addresses all criteria suitable for SMEs. Our findings show few exemplars of practical application, little guidance on how to use/apply the toolkits, and very low uptake by SMEs. Our analysis provides a mechanism for SMEs to select their own toolkits based on their current capacity, resources, and ethical awareness levels – focusing initially at the conceptualization stage of the AI life cycle and then extending throughout.**

***Impact Statement*—In parallel to the recent acceleration in development and adoption of artificial intelligence, there has been intense and worldwide discourse around the ethics of such systems. This debate has highlighted that without good governance, transparency and monitoring, indiscriminate use of AI could lead to significant harms, discrimination, and injustice. Consensus has settled on a broad set of overarching principles for ethical AI; now myriad resources and toolkits exist to assist with embedding ethical practices along the research-development-deployment value chain. Our evaluation of 77 toolkits reveals the breadth and depth of the themes they cover and barriers to their use, including a lack of adoption case studies. We provide organizations, especially SMEs, with an easy-to-use lookup table (Table V) to help them select a set of toolkits to ensure that as well as addressing all key ethical themes, they can also match their resources, skills and priority areas for implementing ethical best practice.**

***Index Terms*—Artificial intelligence (AI), business, ethics, responsible, toolkits, trustworthy.**

I. Introduction

THE ethical, social, and legal landscape of artificial intelligence (AI) driven systems is rapidly changing. Since the General Data Protection Regulation 2018 [1], stakeholders developing AI systems have faced numerous challenges in the interpretation and implementation of Article 22, specifically concerning an individual's rights in the context of automated decision-making, the ability to explain AI decisions, explanation of the logic involved, and the development of models using only "correct" data. This has caused major challenges because of the lack of legal guidance, case law, and ethical principles about the use of AI in different contexts. For small- and medium-sized enterprises (SMEs), these challenges are even greater due to a lack of specific skills, budget, and human resource. The international policy and impact landscape of AI is still fragmented in approaches to regulation, frameworks, guidelines, and standards (i.e., P7000), with numerous ethical principles being circulated which all convey broadly similar messages [2]–[15].

These "guidelines" often focus on the AI technology or service rather than organizational processes and human behaviors, providing little to no mechanisms for accountability and compliance (audit), and ignore the benefits of coproduction and public scrutiny [16]. From an SME perspective, practical implementation is difficult if not impossible. There has been significant "bad press" around poor design, poor rationale, and unethical applications of AI, which has fueled public mistrust. Pownall [17] provides an excellent, regularly updated repository of news stories that challenge whether the use of AI is ethical, for example, the use of face tracking tablets which profile customers and deliver relevant advertisements in UBER. As the public gains knowledge and understanding of issues around the use and application of AI (including bias, fairness, accountability, responsibility, etc.) coupled with an increased awareness of data privacy, both public services and the private sector will have to become more accountable if they win public trust and secure the vital public "license to operate." Reputational damage as a result of insufficient or ineffective data and AI governance can cause significant harm to a business, with greater impact on SMEs [17]. There is still a significant gap between top–down theory

Manuscript received 30 July 2021; revised 6 October 2021; accepted 4 December 2021. Date of publication 21 December 2021; date of current version 21 July 2023. This article was recommended for publication by Associate Editor F. Chowdhury upon evaluation of the reviewers' comments. *(Corresponding author: Keeley Crockett.)*

Keeley Crockett, Luciano Gerber, and Annabel Latham are with the Department of Computing and Mathematics, Manchester Metropolitan University, M1 5GD Manchester, U.K. (e-mail: k.crockett@mmu.ac.uk; L.Gerber@mmu.ac.uk; A.Latham@mmu.ac.uk).

Edwin Colyer is with Research and Knowledge Exchange, Manchester Metropolitan University, M1 5GD Manchester, U.K. (e-mail: E.Colyer@mmu.ac.uk).

This article has supplementary downloadable material available at https://doi.org/10.1109/TAI.2021.3137091, provided by the authors.

Digital Object Identifier 10.1109/TAI.2021.3137091

and practical adoption of robust ethical practices across the entire AI value chain [15], [18], [19], but our research suggests that this is more prevalent in SMEs.

In this article, we adopt the European Commission's definition of an SME which is an enterprise with fewer than 250 employees, a turnover below €50 million or a balanced sheet total below €43 million [20]. A small business has fewer than 50 employees and a micro business fewer than ten employees [20]. Global business will have different definitions on the size of SMEs, for example, in USA, an SME may have up to 500 employees dependent on the sector [21]. The World Bank states that globally, SMEs represent 90% of businesses and account for over 50% of employment, and in emerging markets, seven out of ten jobs are created by SMEs [22]. In many countries, SMEs are able to access competitive public funding to support growth acceleration and drive innovation in the AI space, but to date there has been little to no focus on responsible innovation. These programs have generally ignored the need for strong AI and data governance, and not provided training and upskilling in the domains. Fortunately, over the past few years numerous organizations and academics have published "ethical toolkits" to help organizations adopt and embed processes and practices that mitigate risks and "do AI ethically." These toolkits help organizations ensure their innovative systems adhere to the key pillars of "ethical tech" around beneficence, nonmaleficence, autonomy, justice, and explicability [19].

The overall aim of this article is to evaluate the thematic and AI life cycle coverage of these toolkits. We also assess the usability of the toolkits from an SME perspective and identify which toolkits are least onerous to adopt and address the barriers to adoption highlighted by SMEs. By categorizing the toolkits against ethical AI themes and adoption/usability, we provide organizations of all sizes, but especially SMEs, with an easy way to identify the most suitable tools, methods, and processes to implement. Our study is divided into two parts. First, we conducted qualitative SME consultations over two events to establish their understanding of both data and AI ethical principles and to identify the key barriers SMEs face in their adoption. As the collaboration between business and universities is a highly important mechanism for R&D activities and for stimulating innovation, it is important that academics make the good ethical research practices from within their institutions integral to contract research and knowledge exchange activities. Second, we conducted a review of available toolkits (published in academic, organizational, government, and gray literature) that support ethical and responsible AI practices. We evaluated these toolkits using criteria partly informed by our SME consultations across four aspects of ethical AI: 1) human-centric ethical principles; 2) applicability across the AI development life cycle; 3) barriers to adoption; and 4) key ethics themes covered.

In this article, we define a toolkit as a document or resource including guidelines (provided the described methods, techniques, or instructions for implementation), checklists, methodologies, activities, processes, frameworks, workflows, or approaches where the content focus is on responsible or ethical data (data ethics) or AI (ethical/responsible/trustworthy/trusted AI). We expand the definition of toolkit defined by Morley *et al.* [23] which focuses only on technical toolkits designed for data scientists and developers up to 2018.

This research aims to address the following research questions.

1) What are the barriers to ethical AI adoption by SMEs?
2) What is the current state of the market in practical toolkits for embedding AI ethical frameworks and governance into an SME culture?

The main contributions of this article are as follows.

1) An analysis of the viewpoints of SMEs on ethical data and AI practices established through two engagement events which are useful to those organizations which are developing toolkits.
2) Identification of barriers to adoption of ethical principles, practices, and toolkits for SMEs.
3) A review and evaluation of recent toolkits against four groups of criteria (common ethical principles, stages of the AI product life cycle, responsible AI aspects and practical application aspects) designed to facilitate practical application of data and AI ethical practices.
4) An easy-to-use lookup table of ranked toolkits based on expert intercoder agreements of criteria coverage – suitable for SMEs to use.
5) Recommendations to the research community on the role of data and AI ethics in business knowledge exchange.

The rest of this article is organized as follows. Section II presents a summary of the core risk factors associated with AI and an overview of the latest legal frameworks and current ethical guidelines and principles. In Section III, we present our two-part methodology; first, describing two SME events leading to the identification of barriers to adoption of ethical toolkits and second, our method for conducting a review and coding of the state-of-the-art toolkits against a range of criteria. We perform an analysis of these toolkits and SME events in Section IV, which leads to a series of recommendations, conclusions, and the wider implications of findings in Section V.

II. Background

A. Risk Factors in AI

When conceptualizing, creating, and implementing an AI system, it is important to consider the risk factors associated with the data used, the model(s) built, and the life span of the model [18], [19]. Furthermore, the societal outcomes and impacts (negative or positive; helpful or harmful) arising during the life span of application should also be considered. From a business perspective, there is a clear relationship between perceived risk in an AI system in a given context and how much trust users have in the decisions it makes [24], [25]. The majority of risk factors are well documented. *Bias* is one of the most complex factors as consideration must be given to bias that is embedded into organizational or industrial cultures, personal, unconscious, and human bias and data representation bias [26], [27]. For example, data that have been labeled by humans for training a model may be subjective, even among experts. Different models may need to be developed for different genders, cultures, etc., as it is rarely possible to generalize models to an entire human population

based on limited training data. *Fairness* is about treating people equally through developing models that encapsulate moral standards in the decision-making process. *Explainability* is required, so all stakeholders, including people impacted by the decisions of automated systems, can understand how a decision is made and the user knows why a system has made a decision [28], [29]. *Societal impacts* (potential benefits and harms) must be considered by a business, not only just to mitigate reputational damage in case of legal complaints but also to meet or exceed minimum standards of business ethics. Businesses must question where *responsibility* (tasks and obligations) lies within their AI governance framework and define *accountability* (oversight and liability) to roles across the design/development/ deployment life cycle. With AI legislation changes on the horizon, deep thinking and consensus surrounding these risk factors is required by both academics and industry regardless of size to assess the risk of an AI solution to both individuals and society. The problem is now bridging the gap between principles and practice, so there is some assurance that AI systems comply with the agreed principles.

B. Principles and Guidelines

Over the past five years, governments, corporations, and international bodies have produced a significant amount of guidance on the ethical dimensions of AI and data driven technologies. To understand how crowded this space is and the difficultly of choice for SMEs with regard to which guidelines to follow, this section provides a brief overview. In 2019, Jobin *et al.* [4] conducted a survey of global ethical guidelines comprised of 84 documents and analyzed their thematic coverage over 11 ethical principles identified by keywords. This work provides a good understanding of the coverage of ethical AI principles and guidelines between 2011 and April 2019. However, the landscape is very dynamic. In 2019, the Beijing Academy of Artificial Intelligence published the Beijing AI Principles advocating ethical AI [5], OECD proposed five value-based principles for the responsible stewardship of trustworthy AI [7], and the European Commission issued ethical guidelines for Trustworthy AI [2]. In 2020, the U.S. Office of Management and Budget issued Guidance for Regulation of Artificial Intelligence Applications [11]. In June 2021, The General Conference of the United Nations Educational, Scientific and Cultural Organization (UNESCO) presented the Draft Text of the Recommendation on the Ethics of Artificial Intelligence, which focuses on a human-centered approach to AI, recommending that "AI must be for the greater interest of the people, not the other way around" [8]. The U.K. government provided an updated summary of data and AI ethical principles developed by both the public sector and the government in 2020 [9], which included a joint publication on AI procurement guidelines developed with the World Economic Forum [30], and specific guidelines and a checklist for using AI in health care [31]. In 2021, the U.K. AI Council published an AI road map [32], further "guidance" on procurement [33] and its national data strategy [34]. A brief analysis of the commonality of ethical principles can be found as shown by Crockett [35], from which a subset of our toolkit evaluation criteria is derived.

C. Legal Frameworks

Legal frameworks in the space of AI and data driven technologies are relatively new and rapidly emerging. The GDPR 2018 [1] first introduced Article 22, a series of safeguards and information obligations in relation to automated decision-making. These included empowering the data subject as stated in Recital 71 "*not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her*" [1], the right to ask for human intervention, explanation of how the automated decision was made "*the logic involved.*" Recital 71 states that the data controller should use appropriate mathematical and statistical procedures for profiling and that data should be accurate in order to minimize the risk of errors [1]. In 2018, the EU also published its AI strategy which promoted a human-centric approach, which focused on respecting European values and human rights. Recently, the EU has published the proposed Regulatory Framework on AI [36], which contains a framework to assess the risk of any AI product, service, or system. Four risk levels are defined as follows.

1) *Unacceptable risk*: AI systems considered a clear threat to the safety, livelihoods, and rights of people will be banned.
2) *High risk*: AI systems identified as high risk (including law enforcement, credit scoring, and border control management) are subject to a deep risk assessment, mitigation strategy, high quality datasets, traceability, documentation, clear explainability protocols to the user, and a high level of robustness, security, and accuracy.
3) *Limited risk*: This includes chatbots where human–machine transparency is a requirement.
4) *Minimal risk*: This includes applications such as AI-enabled video games or spam filters [36].

An excellent primer on the principles and priorities required for a legal framework can be found in [37], produced by the Council of Europe's Ad Hoc Committee on Artificial Intelligence. Leslie *et al.* [37] also provide suggestions on options for a legal framework and a mapping between substantive human and legal rights and key obligations of AI developers when building AI systems and services.

III. Methodology

This article comprises a two-part methodology. The first part is an analysis of a series of practical SME engagement events. These events took place between July 2020 and June 2021 and were designed to capture the "SME voice" on their understanding of ethical AI, its practical implementation, awareness of ethical toolkits, and the barriers to adopting good ethical practices. The aim of the analysis was to establish which themes associated with ethical AI that SMEs are most aware of, and the perceived barriers to ethical AI adoption. Part two is a review of a range of practical toolkits designed to support the implementing into practice of ethical AI principles. These toolkits were evaluated and coded against the common themes and barriers from the SME events and against a range of criteria relating to coverage of the AI life cycle, and general ethics themes.

A. Part 1: SME Engagement and Consultation Study

This section outlines the methodologies for two distinct SME engagement events which explored the need for and barriers to ethical AI.

1) Event 1: Our Place Our Data: To understand the landscape for local businesses and local authorities in ethical AI understanding and practice, a qualitative research study took place in June and July 2020, comprising two roundtables and follow-up interviews. The study was initiated by Manchester Metropolitan University (MMU), designed in collaboration with an independent think tank and with the support of the U.K.'s All-Party Parliamentary Group on Data Analytics (APPGDA). During the roundtables, participants were provided with an overview of a proposed model for place-based support for ethical AI to build a local ecosystem in which ethical and responsible AI development could be nurtured and thrive. The theme for the first roundtable ($n = 20$) was "Data and Public: Creating a data-driven future for Greater Manchester" and sought to capture responses to a series of key questions, which included the following.

1) How can the public be better engaged with policies around ethical data use?
2) What are the current challenges and shortcomings associated with ethical guidelines and principles for the use of data by public and private-sector bodies?
3) What does an effective local data ecosystem looks like?

The second roundtable was at U.K. national level, featuring not only local SMEs and Policy Makers but also Members of Parliament and the House of Lords, and key national stakeholders such as the Centre for Data Ethics and Innovation (CDEI), Visa, British Standards Institute, and the Greater Manchester Combined Authority (GMCA). The second roundtable ($n = 18$) focused on how parliament and government could work to develop local data strategies as part of a wider effort to make the U.K. a world leader in ethical, data-driven technologies. It also analyzed current links between central government, regulators, local and combined authorities, and industry, and considered how those links could be developed over the coming years. The discussion focused on how to develop place-based approaches to data ethics; the role for regulators and government bodies; the feasibility of an "Ethical AI kitemark," which organizations should lead on ethical AI policies at the national and regional level; and what challenges exist with regard to bringing these bodies together.

Following the roundtables (between August 2020 and March 2021), a series of supplementary follow-up interviews were conducted by Policy Connect with selected participants to explore some of the emergent themes in greater depth. Summary reports from both roundtable events and the interviews were produced by Policy Connect and cross-checked by this study's authors (Crockett and Colyer) for accuracy, identified emergent themes, and indicators of agreement, disagreement, and consensus among participants.

2) Event 2. Greater Manchester AI Foundry: The Greater Manchester AI Foundry [41], with £3 million ERDF funding, is a three-year research and innovation project which commenced in July 2020. The aim of the Foundry is to increase SME performance by placing AI research and innovation at the center of business growth through practical knowledge transfer from AI academic research into industry. SMEs go through two phases: 1) Phase 1 is a series of workshops on AI development from a business perspective and 2) Phase 2 is a technical assist to develop a prototype AI solution. The objective is that research acts as a technology accelerator for new products and services based on AI. Given the importance of the development of ethical technology, a pilot workshop was given in early 2021 to the first cohort of SME participants ($n = 20$) to enable SMEs to gain an understanding of ethical, social, and legal perspectives of AI and data privacy, and also to facilitate practical ethics into the technical assists. The workshop was not intended to provide any legal advice, rather it was designed to showcase best practice in ethics and regulatory compliance. The first workshop was positively received and a full workshop was developed and embedded with a second cohort in June 2021. In the full workshop, SMEs were actively encouraged to look at the impact and assess the risks of their AI product or service in light of the newly proposed EU regulation [36]. The workshops introduced a variety of ethical toolkits and activities with SMEs including datasheets for datasets [42], consequence scanning [43], conducting a data privacy impact assessment [44], and examining the risk to stakeholders of an AI recruitment tool using padlet [45]. Feedback on adoption of potential tools and barriers to use was obtained through Q and A and discussion during and after the workshop. Workshop members were also asked to complete a longitudinal ethical AI practice survey [46]. Feedback was anonymized and collated and thematic coding was undertaken to identify ethical concerns and barriers.

B. Part 2: Review of Practical "Ethical" Toolkits

Our review of toolkits covers academic, organizational, government, and gray literature sources. The search strategy employed the following primary keywords: (*toolkit, resource, guidelines, guidance, checklist, methodology, method, activity, process, framework, workflow, approach*); (*ethical, responsible, trustworthy, trusted, data, data ethics, tech ethics*); and [*artificial intelligence* (*AI*)*, machine learning* (*ML*)]. Our toolkit dataset was created by using the primary keywords to perform searches on Google Scholar and Scopus and gray online literature on Google from 2017 to July 5th, 2021. Our toolkit dataset was also cross-checked with work published by Morley *et al.* [23] and Moltzau [38], who produced a full typology of identified methods and tools (up to mid-July 2019) which were limited to helping developers, engineers, and designers of ML apply ethics within their roles. In comparison, our review takes on a more holistic view in analyzing toolkits that are also used to initiate engagement with wider public stakeholders to explain decisions and build trust. *Inclusion criteria* were documents (checklists, guidelines, activities) including those published by public and private sectors, governments, and international bodies and the toolkit language was English. *Exclusion criteria* were legal frameworks, opinion articles and speeches. Once a list of toolkits that met the inclusion criteria was obtained (referred to as the EAI toolkit dataset), each toolkit was evaluated and coded independently by expert researchers in the field of AI and ethics

TABLE I
GROUP *B*: COMMON ETHICAL PRINCIPLES

Criterion No	Ethical Principal
B_1	AI should not be used to harm or kill any human and respect human rights
B_2	AI must always be fair, unbiased and transparent in the decision-making process
B_3	AI systems and solutions should always operate within the law and have human accountability
B_4	Data Governance and Data Privacy should be incorporated into the AI life cycle
B_5	Humans should always know when they have interactions with an AI system
B_6	AI systems should be inclusive to all human–centered AI design
B_7	Appropriate levels of explainability should always be provided on AI decision making
B_8	Humans must always be in the loop when an AI is making a decision that affect other humans
B_9	Humans responsible for designing, developing and operating AI systems should be competent in the skills and knowledge required
B_{10}	AI systems should be sustainable and work to benefit humans, the society and the environment
B_{11}	AI systems should be inclusive to all

TABLE II
GROUP *C*: STAGES OF THE AI PRODUCT LIFE CYCLE

Criterion No	Criterion Name/Description
C_1	**Conceptualization**: includes imagineering, defining aims, objectives, desiderata, cost/benefit of new AI products and services and conducting a risk assessment
C_2	**Data Preparation and Exploration**: e.g., collection, curation, feature engineering, cleaning, feature selection, and sampling
C_3	**Model Building and Evaluation**
C_4	**Deployment and Monitoring**

TABLE III
GROUP *D*: RESPONSIBLE AI THEMES

Criteria No	Criterion Name/Description
D_1	**Robustness**
D_2	**Fairness** (includes bias)
D_3	**Transparency**
D_4	**Accountability**
D_5	**Explainability**
D_6	**Privacy**
D_7	**Safety**
D_8	**Impact** (both positive and negative, on society)
D_9	**Inclusivity of the toolkit (in general)**: incorporation of needs from stakeholders with different roles (e.g., managerial, data protection officer), motivations, technical expertise (e.g., machine learning engineers, senior management), and cognitive equity (for example, that it was inclusive to people with varying levels of educational attainment)
D_{10}	**Inclusivity w.r.t to General Public**: as D_9 but, more specifically, the extent to which the conception of the toolkits included and offered consultation with the general public

TABLE IV
GROUP *E*: PRACTICAL APPLICATION ASPECTS

Criteria No	Criterion Name/Description
E_1	**Exemplars:** case studies, examples of what-good-looks-like, among others.
E_2	Quick Read e.g. too-long-didn't-read; short, accessible, practical, quick-start type of guidance for application of the principles.
E_3	**Stakeholders Inclusivity**: does the toolkit address different types of stakeholders such as technical, managerial, and end user (e.g., customer)?
E_4	**Feasibility** of applying the toolkit with respect to a **typical SME skillset**.
E_5	**Feasibility** of applying the toolkit with respect to **resources** such as workload, personnel, budget at SMEs.
E_6	**Recommendations of AI Techniques**: e.g., does the toolkit make concrete recommendations for data management and machine learning methods?
E_7	**Recommendations on Personnel Training**
E_8	**Evidence of adoption** of the toolkit by an SME

based on four groups of criteria, shown in Tables I–IV. For each toolkit, its source (academic, organizational, business, and gray) was recorded, along with publication year, whether it was open source, and the country of origin.

Criteria in Group *E* were determined on the basis of the findings of the two SME engagement events reported in Section IV – analysis of SME engagement events.

A modified nominal group approach to coding was adopted [39], [40]. The first round of coding involved three experts in the fields of AI, ethics, and business engagement, independently evaluating two-thirds of the EAI toolkit dataset with each toolkit being evaluated by two experts initially. A structured spreadsheet containing links to the toolkits and the 33 criteria for coding was given to each expert to evaluate and code independently. Each criterion was coded according to a three-point Likert scale with values in (01, 2) indicating, respectively, *weak*, *moderate*, and *strong* levels of support by a toolkit for a given criterion. For example, if a toolkit strongly addressed B_{10} – AI systems should be sustainable and work to benefit humans, the society, and the environment – then it was scored as 2; if it moderately or partially addressed that criterion, it was scored 1; and if support for the criterion was largely or completely absent, then it was scored as 0.

The first round of independent coding revealed a 72% agreement across 33 criteria; 18% of criteria indicated that there was a disagreement with one expert coding 0 and another scoring 1 or 2; in 10% of cases, both experts agreed that the toolkit contained at least some evidence of the criteria, but the experts disagreed on how much (scoring 1 or 2). When adopting a percentage agreement approach [39] there is no agreed threshold for consensus, and it is up to the researchers to judge what represents acceptable agreement for a particular study. A second round of independent expert coding was then instigated for all toolkits where there was significant disagreement for any criteria, defined as when one expert scored 0 and the other expert either 1 or 2; these toolkits were fully coded by a third expert in an attempt to establish majority agreement. The level of agreement between the three experts was then recorded in a

structured spreadsheet for 77 toolkits. There was a good majority agreement between the two experts for 89% of the 33 criteria scored across the 77 toolkits. Experts were unable to reach a majority agreement on all criteria across all toolkits in only 1% of cases. The most common disagreement between the coders was on the interpretation of B_{10} – AI systems should be sustainable and work to benefit humans, the society, and the environment (6 out of 77 toolkits) and on the toolkit coverage of C_4 – deployment and monitoring (6 out of 77 toolkits).

IV. Analysis and Discussion

A. Analysis of SME Engagement Events

Event 1: For event 1, analysis of the first roundtable revealed that ethical and legal issues surrounding "data" and not "AI" needed to be resolved first before the wider ethical aspects of AI could be addressed. This was true for both public and private sector organizations. The key themes emerging from the roundtables were as follows:

1) ethical guidelines and principles should be simple and flexible and should be much more than a checklist;
2) practical guidance on how to apply data and ethical AI principles should be usable;
3) mechanisms were needed to support practical guidance (training, resource support) in partnership with local authorities;
4) data-driven technology strategies should be developed in partnership with all stakeholders;
5) SMEs should have access to "resource knowledge sharing" to make effective and ethical use of AI and ML.

The main output of the Event 1 study was a report *Our Place, Our Data: Involving Local People in Data and AI-Based Recovery* [47], which made five recommendations to the U.K. government, including that local authorities should work in partnership with businesses (including SMEs) and academic institutions to develop data-driven technology strategies to develop innovative AI services and products which have citizen engagement at the heart of the creation process.

Event 2: The analysis of Event 2 was based on Q and A during the two cohort sessions and follow-ups in 1:1 virtual meetings. SMEs referred to the following Information Commissioner's Office (ICO) guidance: What are the accountability and governance implications of AI? [48], guidance on AI and data protection [44], data protection impact assessments [44], what do we need to do to ensure lawfulness, fairness, and transparency in AI systems? [45], and how do we ensure individual rights in our AI systems? [49]. They noted these documents as long and complicated, and provided no practical advice or methods on how to apply them. The key message was that toolkits/guidance needed to be simpler. One SME data scientist stated that they "*did not know some of this existed*" emphasizing the general lack of awareness. SMEs thought that training or free consultancy was required to help them understand and apply legal guidance in relation to AI and data. Three SMEs also thought that in general, ICO guidance was "*subject to interpretation.*" Positive feedback was received about the use of consequence scanning [43] as a useful way to think about harms and risks of a product at conceptualization, but in general SMEs said whether they would be used in practice was based on whether they had available resource. They had no strong opinion about the benefits of involving the public, for example, as a stakeholder in an activity such as consequence scanning. Despite growing consensus on the benefits of public involvement to build trust in AI tech [50], [51], SMEs indicated that they were not sure how to involve the public and that the real benefits of consulting with the public was not clear. Two SMEs suggested that successful case studies would benefit them. The SMEs thought that the toolkits presented were useful, but they needed time to learn how to use them – not only just one-off training but also how to practically apply them in their own business.

Summary: From these two events, the barriers to SMEs adopting toolkits were identified as follows.

1) Availability of resources to SMEs (people and time), current skills, and training requirements.
2) Skepticism about the benefits of public stakeholder involvement in the design of new products and services.
3) Lack of understanding around governance of responsibility and accountability regarding AI development and implementation outcomes.
4) The lack of audit and compliance and legal frameworks.
5) Need for practical training and upskilling regarding ethics, data and legal frameworks, and managing liabilities.
6) Challenges associated with communication with users – different language for different stakeholders.
7) Serious implications for a business in terms of liability. What are the consequences of noncompliance?

B. Toolkit Analysis

Following the methodology described in Section III, a total of 77 toolkits were identified which met the inclusion criteria. 30 of these toolkits were from 2021, while the earliest was from 2017. A total of 51% of toolkits were from the US, 23% were from the U.K. and there was representation from South America, China, Denmark, Saudi Arabia, Germany, and Ireland, in addition to three toolkits which were classed as global. The process for analyzing toolkits can be defined as follows.

1) All toolkits were scored using the groups of criteria B to E (see Tables I to IV) according to a three-point Likert scale with values in (0, 1, 2) indicating, respectively, weak, moderate, and strong level of support by a toolkit for a given criterion. As explained in Section III, these are the combined scores from the interannotator coding and agreement process.
2) For the analysis of the criteria, we derived an n by m matrix R (see supplementary material), where n is the number of toolkits ($n = 77$) and m is the number of criteria considered ($m = 33$).
3) Each cell in R contains one of (0, 1, 2, D), with D standing for a disagreement among coders.
4) From R, we derive a mean score for a toolkit (i.e., a row) or a criterion (i.e., a column) by taking the mean of its empirical probability distribution (epdf) (excluding disagreements). More specifically, let X be either a row or

a column in M, which is assumed to be a discrete random variable. Then, epdf(X) = (p_0, p_1, p_2), where pi is the probability of the score i in (0, 1, 2).

Table V located in the appendix, displays the statistical summary of scores across the 77 toolkits, ranked on the basis of their coverage of criteria groups *C, D,* and *E,* where p_0, p_1, and p_2 are the values of the epdf, shown as percentages, of the Likert scores on the criteria, and m is the number of criteria assessed. Group B is not included in Table V as it considerably overlaps with responsible AI themes in Group D. We opted for the latter, given that it provides a more fine-grained analysis of tool coverage. For example, B_2 – AI must always be fair, unbiased, and transparent in the decision-making process – is covered by D_2 – fairness (including bias) and D_3 (transparency).

The top-ranking toolkit was Microsoft's *Responsible Innovation: A Best Practices Toolkit* [111]. While this toolkit was targeted at developers, it had a strong focus on identifying potential negative consequences of technology on humans. The toolkit features three elements. The first, judgment call – a game and team-based activity that explores all of Microsoft's AI principles [128] through scenario imagining where the aim is for participants to write product reviews for different stakeholders accessing the impact and harms. Harms modeling – a framework for product teams based on the four pillars of responsible innovation (*"injuries, denial of consequential services, infringement on human rights, and erosion of democratic and societal structures"*[111]) – is designed for teams to look at real world impacts of technology. Finally, community jury, defined as an adaptation of the citizen jury [111] brings together the product team and user stakeholders to discuss various product artifacts, deliberate and cocreate new technologies over a 2–3-h session. This toolkit had moderate to strong coverage across all criteria *B, C,* and *D*. However, it did not contain any exemplars E_1, and had no training guides E_7, which is a key requirement for SMEs. That said, its uniqueness is its ability to engage the public, seek consensus, and opinion, and it is forward-thinking in terms of providing practical guidance that is applicable to a wide range of businesses/organizations. Ranked second was the U.K. government's *Data Ethics Framework Guidance*, published in 2020, which focuses on responsible and ethical use of data in the public sector [114]. While the emphasis is on the public sector, the guidance is targeted at all stakeholders who use or interact with data, including policy makers and data scientists. Similar to [111], the emphasis is on defining and understanding the public benefit of any "data project" including human rights, understanding potential consequences, compliance with law and diversity in the development team. The toolkit provides a set of questions which are scored on a Likert scale based on clarity and understanding with respect to a specific project. The framework also covers algorithms and outputs in relation to AI and is applicable to all stages of the AI life cycle. This toolkit also did not provide any examples of practical application E_1 and is less inclusive in its approach by not involving wider publics as stakeholders E_3. The toolkit did not offer any specific training E_8.

Table V also highlights the lowest ranking toolkits [70], [97], and [125], none of which provided strong evidence of coverage across any of the criteria. For example, Covington is a global law firm, based in USA. Its toolkit [125] claims to provide practical guidance for "*the evolving regulatory landscape*" with an emphasis on USA, U.K., and EU. The guidance is in the form of overviews, summaries of news articles, and a white paper with links to recent AI legislation articles and to the ICO/Alan Turing Explaining AI Decisions' toolkit [83]. On the basis of our findings across the two SME engagement events, SMEs requested more training in order to understand the implications of legal frameworks and this toolkit would be difficult for them to practically apply as it is more a means of monitoring evolving regulation and legislation.

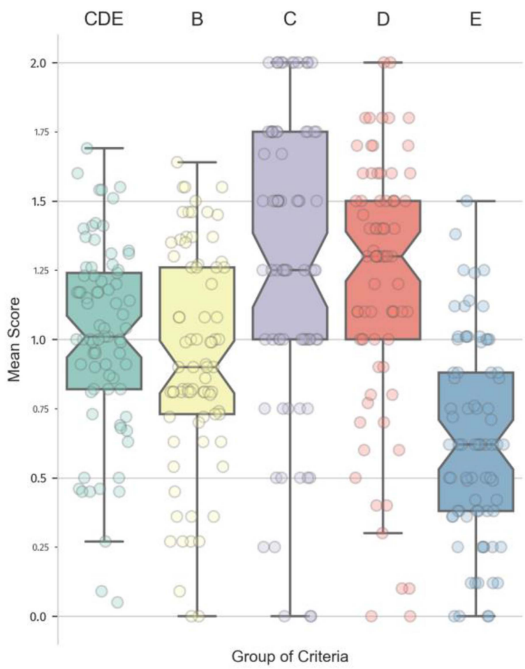

Fig. 1. Boxplot showing mean score distributions of independent expert ranked criteria over Likert scale [0, ..., 2].

Fig. 1 shows the distribution of mean scores by groups of criteria. For example, one can see that criteria E (the practical application aspects for SMEs) has the lowest median and overall coverage by the toolkits (each, represented as a data point). Each plot represents one toolkit. This confirms the largely consensus view arising from our two events that in spite of the existence of toolkits to support responsible and ethical AI, most still lack adequate instructions and training to facilitate adoption. Many require significant time and specialist skills for implementation due to their length

Analysis has shown that no single toolkit covers all criteria, as indicated in Table V ($p_0 > 0$ in all columns). Consequently, each set of criteria will now be analyzed independently to assess criterion coverage and highlight those toolkits with the highest ranked coverage. This will help SMEs to select toolkits that best align with their business culture and values, and the stage they are at in developing their own ethical policies and procedures.

1) Common Ethical Principles (Group B*):* Fig. 2 shows the toolkit coverage of the ethical principles B_1,..., B_{11}. Clearly, B_2 – AI must always be fair, unbiased, and transparent in the decision-making process receives the highest coverage across all toolkits. This is closely followed B_3 – AI systems should always

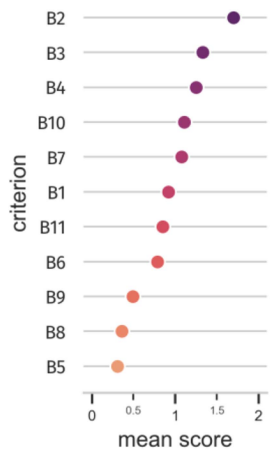

Fig. 2. Ranking of Group *B* criteria on mean score.

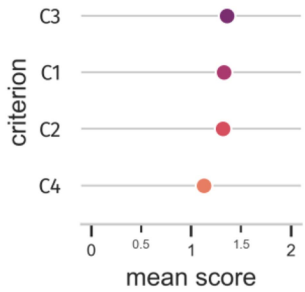

Fig. 3. Ranking of Group *C* criteria on mean score.

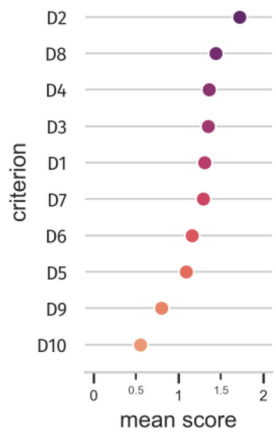

Fig. 4. Ranking of Group *D* criteria on mean score.

operate within the law and have human accountability and B_4 – data governance and data privacy should be incorporated into the AI life cycle. These findings align with predominant global ethical principles [4]. Of least coverage was B_5, humans should always know when they have interactions with an AI system, which is only highlighted by toolkits [74], [116], [118], [120], and [126] and B_8 – a human should always be in the loop for automated decision-making, covered by [101], [112], and [126]. Toolkit [126] (ranked 33 overall) stands out in this group. Titled "*Application Guide for the Ethical Assessment of AI for Actors within the Entrepreneurial Ecosystem,*" the toolkit is an open source guide published by the Inter-America Development Bank in May 2021. Its interdisciplinary approach to ethical self-assessment covers all stages on the AI life cycle, governance, and security with a focus on human involvement in AI systems. The guide has a three-stage assessment to determine the level of human involvement based on the impact that the system has on a human's life. The toolkit helps organizations define associated key performance indicators, risk mitigation, and even develop emergency responses following analysis of all conceivable scenarios.

2) Stages of AI Product Life Cycle (Group C*):* Fig. 3 shows the toolkit coverage for the four stages of the AI life cycle: 1) conceptualization C_1; 2) data preparation C_2; 3) exploration, model building, and evaluation C_3; and 4) deployment and monitoring C_4. Analysis showed that toolkits were less likely to cover the audit and compliance stage of the life cycle, compared to the other stages, presumably because few regulatory frameworks or standards are yet approved. For example, to date, out of the IEEE P7000 standards in development, only the IEEE 7010-2020 – IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being [14] is available on subscription only. Only toolkits [55], [56], [65], [70], [83], [85], [95], [101], [104], and [107] covered the whole life cycle, but to varying degrees. Toolkits [56] and [107] ranked, respectively, third and fifth overall against all criteria (see Table II). Agile ethics for AI (HAI) [56] is a Trello board which contains a series of boards covering scope, data audit, training, analysis, feedback, calibrate (optimal AI for increased uptake), augmentation (e.g., upskilling and training), and "people and the environment" which addresses accountability in AI deployment. Each board contains a series of "TO DOs" with specific resources, all available as open source. The World Economic Forum's AI Procurement in a Box: Workbook [107] is a lengthy tool kit (54 pages) that features a series of questions and risk matrices and mapping tools covering the full AI life cycle. It is intended for businesses seeking to procure AI solutions. It also features a user manual with a strong emphasis on how to define the public benefit of AI while assessing risks in the early stages of conceptualization. The toolkit provides guidance on how to address both the technical and ethical limitations of data, clearly addressing the impact of bias.

3) Responsible AI Themes (Group D*):* Fig. 4. shows the toolkit coverage for the responsible AI themes: Robustness D_1, fairness D_2, transparency D_3, accountability D_4, explainability D_5, privacy D_6, safety D_7, impact D_8, inclusivity of the toolkit (in general) D_9, and inclusivity w.r.t. general public inclusion as a stakeholder D_{10}. Examination of Group *D* criteria allows for more fine-grained analysis than within the more general ethical principles (see Fig. 2) and we expected to see the similarity with ethical principle B_2 and fairness D_2 with regard to coverage. Ninety-five percent of all toolkits moderately or strongly addressed the issue of fairness, with 88% also addressing the impact of AI technology on society D_8. Accountability D_4, both in terms of the processes of developing responsible technology

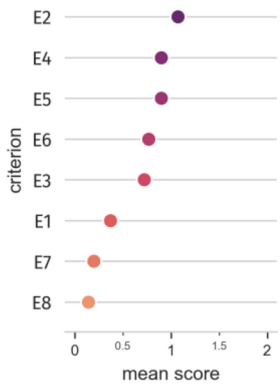

Fig. 5. Ranking of Group *E* criteria on mean score.

and the decision outcome, quality of the data and the model produced, also had moderate to strong coverage in 89% of toolkits. More than half (53%) of the toolkits failed to include the public voice, in any codesign or coproduction process to seek their opinions (D_{10}) and only 62% of toolkits were moderately inclusive to the requirements and needs of a wide range of stakeholders (i.e., data scientists, software developers, managers, CEOs) (D_9). The Action-Oriented AI Policy Toolkit for Technology Audits by Community Advocates and Activists [122], Agile Ethics for AI (HAI) [56], the JUST AI reflection prototype [82], Microsoft's – Responsible Innovation: A Best Practices Toolkit [111], U.K. governments, Data Ethics Framework Guidance [114], and the Royal Society – Democratizing decisions about technology toolkit [120] were the only toolkits to have strong coverage of public inclusivity embedded within the toolkit objectives. As reported in Ouchchy *et al.* [129], public opinion is critical in the acceptance and adoption of new technology. Other work [130] has recommended that businesses including ethical value statements on trusted webpages; the inclusion of both ethicists and the public in new technology discussions could avert negative media responses and reputational damage to businesses. The importance of the role of the public stakeholder is also highlighted in policy road maps [32] and proposed regulation [36].

4) Practical Application Aspects (Group E*):* Fig. 5 displays the ranked criteria in relation to different aspects regarding the practical application of the toolkits. Only 27% of the toolkits were coded as being equivalent to "quick start" guidance E_2. Sixty-nine percent of toolkits and their associated websites provided no exemplars or case studies of how to practically apply the toolkit; only 6% provided at least one example of adoption E_1. Coverage of stakeholders' inclusivity E_4 within the toolkit was scored as weak (27%), moderate (56%), and strong (17%). Analysis showed that toolkits were designed with specific audiences in mind, for example, the technical community (data scientists, programmers, and data analysts) where the focus was on criteria such as bias and fairness in both data quality and model generation. There were few toolkits that had end users and public inclusivity in mind, suggesting that the trajectory of practical application of toolkits is behind emerging legislation and wider discourse around building trust through public involvement [120]. The feasibility of practical application of toolkits w.r.t. to SME resources (workload, personnel, and budgets) E_5 was ranked similar to E_4. This indicated that SMEs would have to make a moderate to high investment to apply toolkits and embed ethical values and processes into business operations. Eighty-three percent of toolkits provided no training opportunities such as step-by-step instructions, user guides or checklist on how to practically use the toolkit. A strong emphasis on training E_7 could only be found in IEEE Ethical Aligned Design [65] and The Royal Society – Democratizing decisions about technology toolkit [120]. The following toolkits covered some aspects of training: [56], [60], [70], [88], [99], [102], [104], [107], [108], [114], and [120]. An observation was that toolkits that were focused on the conceptualization stage of the AI life cycle and/or had more stakeholder inclusivity included some form of training.

Finally, evidence of adoption of a specific toolkit by SMEs' E_8 was barely evident to nonexistent in 90% of toolkits. This suggests that either toolkits have not been designed with SMEs in mind, the barriers to practical application are too high, or toolkits are simply not being evaluated and publicized through practical use cases. Digital Catapult's Machine Intelligence for Business [88] (ranked 24th in Table II) has published a short case study on Loomi – an AI assistant which builds trust through ethical transparent design [129]. Loomi, also the name of the SME featured in the case study, utilized Digital Catapult's ethics framework to reposition "the product using ethics as a key differentiator." IDEO's toolkit (ranked 16th in Table II) highlights the benefits of human-centered design using its Design Kit [64] in a series of humanitarian case studies.

Across the criteria in this category E_1,..., E_8, DotEveryone's Consequence Scanning toolkit [43], ranked 21st (Table II), exhibited moderate to strong coverage of all criteria. This open-source toolkit, developed in U.K., allows businesses and organizations (regardless of size) to examine, debate, risk assess, and mitigate the potential consequences of their product/service on society, communities, and the environment. A manual is provided (27 pages), with minimal resources required. The tool is employed at the conceptualization stage, with all stakeholders taking part, although public stakeholders are not specifically mentioned (D_{10}). A strong facilitator is needed which may be a barrier for SMEs, but a session can last as little as 90 min. The tool has been reportedly adopted by SalesforceUX [130] as a way to bring design risks out into the open.

C. Discussion

This article has evaluated and analyzed 77 toolkits that cover different aspects of the ML/AL life cycle and common ethical principles, responsible AI themes, such as bias and fairness, and degrees of practical application. Consequently, every organization should be able to find one or more toolkits that fit with their working practices, culture, and to complement their organizational values. Although Table II ranked Microsoft's Responsible innovation: A Best Practices Toolkit [111] as the number one toolkit with regard to our criteria (*C*, *D*, and *E*), it still

has limitations in its practical application by SMEs. Therefore, this research concludes that there is not a toolkit currently in existence that overcomes all the barriers and fully meets all the needs of SMEs identified in the analysis of the two SME engagement events. SMEs struggle with long, wordy, and technical documents. They require case studies, clear compelling stories of benefits, and step-by-step instruction manuals on how to use and embed toolkits into operations (and how much time/cash it will cost).

There was a good distribution across the toolkits of all the ethical principles (criteria *B*). Greatest coverage (mean of 1.64) was the Data Ethics Impact Assessment (ranked 17th in Table II) [91] which comprised a 16-page questionnaire designed for organizations to integrate the assessment of data ethics and the impacts of their AI on humans and society within their development and operational processes. The 56 questions cover aspects of transparency, equality, data governance, sustainability, accountability, and human-centered design and centered, drawing on DataEthics.eu's principles of data ethics. In contrast, Nesta's Civic Al Toolkit [121], which focused on using AI and data to address climate crisis and the Online Ethics Canvas [127], had little to no coverage. Results concluded that few toolkits addressed all 11 principles, and none were considered to fully address all 11 by any expert coder. Therefore, organizations will probably need to use more than one toolkit to get comprehensive coverage.

Detailed analysis in Section IV revealed that toolkits [55], [56], [65], [70], [83], [85], [95], [101], [104], and [107] covered the whole AI life cycle, but to varying degrees. Experts agreed that 24% of toolkits did not cover audit and compliance and this may be due to the current lack of AI legislation, regulation, and ethics standards. However, the proposed EU Regulation on AI [132] is likely to have a significant impact on future toolkit development, as it is being described by the Global Centre for Data Innovation as the "*most restrictive regulation of AI*" in the world. The expert coders agreed that 80% of toolkits analyzed in this study placed emphasis on getting things right the first time, i.e., at the point of AI product or service conceptualization, and can be seen as proactive in determining the consequences and harms a potential product could have on humans and society.

Analysis across the responsible AI themes (criteria *D*) indicates that the vast majority of toolkits covered aspects of fairness and the impact of AI. While these are core values in developing ethical and responsible AI, SMEs do need to ensure that they address all themes across the AI life cycle through culture change, rather than becoming fixated on bias and fairness to the detriment of other themes. It is unsurprising that so few toolkits strongly emphasize the importance of citizen representation in their toolkit application. Only 8% of all toolkits strongly advocated the participation of citizens, with 53% relying only on internal stakeholders to take part. An absence of public involvement, especially in the new AI product/service conceptualization phase, leads to flaws in design thinking due to a lack of diversity and inclusivity, which leads to narrower perspectives. Consequently, a great business idea, with no public license to operate, can ultimately lead to reputational damage and loss of revenue. For example, Deloitte reported that a lack of inclusivity in the conceptualization stage of a smart city design resulted in a negative impact as people in wheelchairs were unable to access eye-level retina scanners that require the person to be standing [133]. Section IV highlighted only six toolkits featuring citizen inclusivity. SMEs urgently need to find ways to engage and involve more diverse teams including people outside of their organizations, such as the general public. Our SME engagement events found that this activity is typically beyond their resources and skillset; they also raised concerns about intellectual property rights and trade secrets being disclosed. Put simply, SMEs need support and advice on how to engage effectively. The Community Jury proposed within Microsoft's Responsible innovation: A best practices toolkit [111] is a good example of citizen engagement in the AI life cycle. The caveat is that it was designed by and for a large corporate and not an SME. Setting up such a jury may be daunting and resource intensive for an SME; we propose setting up city or regional juries, focused on ethical AI tech, as part of collective approach, where SMEs could present novel ideas and seek public opinion on design solutions. Ultimately, SMEs should seek to cocreate and codesign with citizens to build trust and obtain the public license to operate, but this is a significant step change to current operations.

Our analysis also highlighted the lack of exemplars or case studies by those organizations who have developed the toolkits. There was little evidence of adoption and virtually none involving SMEs. This is not to say they haven't been involved, but stories, outcomes, analyses, benefits, and outcomes are not in the public domain. This is a key knowledge gap that should be addressed to close the gap between ethical principles and practice. Toolkit developers could produce publicly accessible case studies to thoroughly document the journey and the impacts of adopting ethical practices. This is crucial to lower resistance, leverage investment, and gain the trust and attention of SMEs to invest their limited resources in upskilling and training their employees on AI ethics.

Guidance on how to train people to use the toolkits is another significant challenge. Our analysis indicated that 83% of toolkits did not provide any training material on how to practically implement the tool within the organization. While the overall majority of toolkits are open source and in the public domain, some organizations did offer consultation opportunities for a fee [113], [124], [125]. However, this is not enough, particularly for SMEs, if they do not come with comprehensive training and support materials.

It is important to note that many of these toolkits have been designed for specific and narrow purposes, with no intention to support all possible dimensions of ethical AI, not least because many were produced while ethical frameworks were still under development. For example, IBM's 360 Fairness tool [78] was conceived to focus on evaluating bias and the fairness of algorithms, with no explicit regard for any assessment of eventual outcomes from decisions supported by said algorithms. At the other end of the spectrum, AINow's Algorithmic Impact Assessment toolkit [53] is "designed to support affected communities and stakeholders as they seek to assess the claims made about these systems, and to determine where – or if – their use is acceptable." It is therefore good to bear in mind that SMEs may need to deploy two or more toolkits to fully capture all dimensions of ethical operations.

V. Conclusion

This research aimed to address two research questions as follows:

1) first, to understand the AI ethics landscape from the SME perspective (and uncover any existing barriers to adoption);
2) second, to evaluate and identify existing toolkits that are suitable for practical application by SMEs.

Two SME engagement events were conducted that identified a number *of common barriers to ethical AI* adoption by SMEs on the themes of: 1) resources (people and time); 2) practical business-focused training and upskilling on ethical and responsible AI; 3) data and AI governance infrastructures; 4) citizen engagement; 5) applicability of legal frameworks (data and AI) and how to apply them; and 6) audit, compliance, and liability. Next, a comprehensive review provided a picture of the current state of the market in availability of toolkits for embedding AI ethical frameworks and governance into an SME culture. Our key findings are summarized as recommendations to both the SME and academic communities.

There is no one-size-fits-all toolkit that provides guidance sufficient to cover all ethical principles and themes around responsible and ethical AI. Toolkits vary in their feasibility to implement. It is recommended that SMEs select toolkits based on their current capacity, resources, and ethical awareness levels – focusing initially at the conceptualization stage of the AI life cycle and then extending throughout.

Academics engaged in knowledge transfer projects with businesses should also share good ethical practices, policies, procedures and approval templates from their universities. While established processes governing research ethics are different, for example, in terms of the data processed and controlled, and differences in legal basis according to GDPR, they can help inform the private sector and provide cross pollination of good ethical practices. In this article, ethical AI toolkits have been analyzed from an SME perceptive; however, evaluation of criteria *B*, *C*, and *D* provides a useful reference to the academic community, who may wish to embed the use of toolkits into their ethics approvals and evaluations of research projects. Finally, this analysis contributes a useful teaching resource for courses that include AI ethics and/or data and AI governance, to enable future data scientists and analysts to operationalize practical data and AI ethics within their future employment settings.

Our next step is to produce an easy online tool to help SMEs select the best toolkits to implement/inform practice based on coverage, ease of implementation, and stage in their ethical AI evolution as a company. Our proposed online selection tool will be a curated database that will allow SMEs to provide their own rating across different categories following a similar methodology to ours in this article. They will also be able to propose and categorize new toolkits to add to the database as and when they become available, given the high level of activity in this domain. The tool will be cocreated with SMEs and citizen stakeholders and be flexible to incorporate legislation changes and provide a go-to resource kit.

Appendix

TABLE V
Toolkit Coverage of Criteria *C*, *D*, and *E*

ID	Ref	m	p_0	p_1	p_2	mean	rank
72	**[111]**	22	13	5	82	1.69	1
75	**[114]**	22	13	14	73	1.6	2
6	**[56]**	22	9	27	64	1.55	3
78	**[116]**	22	14	18	68	1.54	4
68	**[107]**	22	5	36	59	1.54	4
24	**[70]**	22	4	41	55	1.51	6
73	**[112]**	22	13	32	55	1.42	7
70	**[109]**	22	18	23	59	1.41	8
18	**[65]**	22	18	23	59	1.41	8
13	**[60]**	22	5	50	45	1.4	10
58	**[99]**	22	4	55	41	1.37	11
74	**[113]**	22	14	36	50	1.36	12
61	**[102]**	22	18	32	50	1.32	13
64	**[104]**	22	14	41	45	1.31	14
60	**[101]**	22	14	41	45	1.31	14
17	**[64]**	22	9	55	36	1.27	16
49	**[91]**	22	19	36	45	1.26	17
41	**[84]**	22	19	36	45	1.26	17
80	**[118]**	20	15	45	40	1.25	19
40	**[83]**	22	31	14	55	1.24	20
9	**[43]**	22	18	41	41	1.23	21
7	**[57]**	22	27	23	50	1.23	21
82	**[120]**	19	37	5	58	1.21	23
45	**[88]**	21	9	62	29	1.2	24
38	**[81]**	22	23	36	41	1.18	25
28	**[74]**	22	19	45	36	1.17	26
30	**[76]**	22	19	45	36	1.17	26
81	**[119]**	22	19	45	36	1.17	26
55	**[96]**	22	19	45	36	1.17	26
23	**[69]**	22	19	45	36	1.17	26
36	**[79]**	21	28	29	43	1.15	31
26	**[72]**	22	27	32	41	1.14	32
88	**[126]**	22	32	23	45	1.13	33
47	**[90]**	22	23	41	36	1.13	34
8	**[58]**	22	32	27	41	1.09	35
69	**[108]**	22	27	41	32	1.05	36
14	**[61]**	22	32	32	36	1.04	37
27	**[73]**	22	23	50	27	1.04	37
15	**[62]**	22	22	55	23	1.01	39
33	**[77]**	22	22	55	23	1.01	39
43	**[86]**	22	18	64	18	1	41
5	**[55]**	21	43	14	43	1	41
52	**[94]**	22	32	36	32	1	41
22	**[68]**	22	18	64	18	1	41
57	**[98]**	22	32	36	32	1	41
50	**[92]**	22	28	45	27	0.99	46
25	**[71]**	22	28	45	27	0.99	46
53	**[95]**	22	36	32	32	0.96	48
85	**[123]**	20	25	55	20	0.95	49
67	**[103]**	22	32	41	27	0.95	49
79	**[117]**	20	20	65	15	0.95	51
65	**[105]**	22	50	9	41	0.91	52
66	**[106]**	22	41	27	32	0.91	52
34	**[78]**	21	38	33	29	0.91	54
86	**[124]**	22	37	36	27	0.9	55
2	**[52]**	22	36	41	23	0.87	56
46	**[89]**	21	33	48	19	0.86	57
84	**[122]**	22	41	36	23	0.82	58
19	**[66]**	22	41	36	23	0.82	58
10	**[59]**	22	41	36	23	0.82	58
71	**[110]**	22	37	45	18	0.81	61
77	**[115]**	22	41	45	14	0.73	62
51	**[93]**	22	46	36	18	0.72	63
44	**[87]**	22	45	41	14	0.69	64
16	**[63]**	22	50	32	18	0.68	65
37	**[80]**	21	33	67	0	0.67	66
12	**[42]**	22	55	27	18	0.63	67
42	**[85]**	22	68	14	18	0.5	68
62	**[103]**	22	68	14	18	0.5	68
59	**[100]**	22	68	18	14	0.46	70
3	**[53]**	22	59	36	5	0.46	71
29	**[75]**	22	64	27	9	0.45	72
39	**[82]**	22	64	27	9	0.45	72
83	**[121]**	22	64	27	9	0.45	72
76	**[127]**	22	73	27	0	0.27	75
56	**[97]**	22	91	9	0	0.09	76
87	**[125]**	22	95	5	0	0.05	77

Acknowledgment

The authors would like to thank Policy Connect and the APPGDA for their work in the inquiry that led to the Our Place Our Data Report [47] and the open source communities that we

used to conduct the data processing, analysis, and visualization such as Seaborn [133], Matplotlib [134], Pandas [135], and Jupyter Lab.

REFERENCES

[1] European Commission, "General data protection regulation," *Recital*, vol. 71, pp. 119–114, 2018. [Online]. Available: https://gdpr-info.eu/

[2] European Commission, "Ethics guidelines for trustworthy AI," 2019. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[3] United Nations, "A framework for ethical AI at the United Nations," 2021. [Online]. Available: https://unite.un.org/sites/unite.un.org/files/unite_paper_-_ethical_ai_at_the_un.pdf

[4] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of AI ethics guidelines," *Nat. Mach. Intell.*, vol. 1, pp. 389–399, 2019.

[5] BAAI, "Beijing AI principles," 2019. [Online]. Available: https://www.baai.ac.cn/news/beijing-ai-principles-en.html

[6] R. Vought, "Regulation of artificial intelligence applications," 2020. [Online]. Available: https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf

[7] OECD, "Principles on AI," 2019. [Online]. Available: https://www.oecd.org/going-digital/ai/principles/

[8] UNESCO, "Draft text of the recommendation on the ethics of artificial intelligence," *UNESCO Digit. Library*, 2021. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000377897

[9] U.K.-Gov, "Data ethics and AI guidance landscape," 2020. [Online]. Available: https://www.gov.uk/guidance/data-ethics-and-ai-guidance-landscape

[10] P. Cihon, M. J. Kleinaltenkamp, J. Schuett, and S. D. Baum, "AI CERTIFICATION: Advancing ethical practice by reducing information asymmetries," *IEEE Trans. Technol. Soc.*, to be published, doi: 10.1109/TTS.2021.3077595.

[11] U. S. Government, "Guidance for regulation of artificial intelligence applications," 2020. [Online]. Available: https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf

[12] Australia Government, "Australia's artificial intelligence ethics framework," 2019. [Online]. Available: https://www.industry.gov.au/data-and-publications/australias-artificial-intelligence-ethics-framework

[13] IEEE, "The ethics certification program for autonomous and intelligent systems (ECPAIS)," 2020. [Online]. Available: https://standards.ieee.org/industry-connections/ecpais.html

[14] IEEE, "Ethics in action in autonomous and intelligent systems," *IEEE P7000 Standards Projects*, 2021. [Online]. Available: https://ethicsinaction.ieee.org/p7000/

[15] D. Schiff, J. Borenstein, J. Biddle, and K. Laas, "AI ethics in the public, private, and NGO sectors: A review of a global document collection," *IEEE Trans. Technol. Soc.*, vol. 2, no. 1, pp. 31–42, Mar. 2021, doi: 10.1109/TTS.2021.305212.

[16] A. Kumar, B. Finley, B. T, S. Tarkoma, and P. Hui, "Sketching an AI marketplace: Tech, economic, and regulatory aspects," *IEEE Access*, vol. 9, pp. 13761–13774, 2021, doi: 10.1109/ACCESS.2021.3050929.

[17] C. A. Pownall, "Understanding the reputational risks of AI," *IAAIC Repository*, 2021. [Online]. Available: https://docs.google.com/spreadsheets/d/1Bn55B4xz21

[18] European Commission, "SME definition," [Online]. Available: https://ec.europa.eu/growth/smes/sme-definition_en

[19] North American Industry Classification System, "SME definition," *US Census Bur.*, 2020. [Online]. Available: https://www.census.gov/eos/www/naics/development_partners/devpartners.html

[20] The World Bank, "Small and medium enterprises (SMEs) finance," 2021. [Online]. Available: https://www.worldbank.org/en/topic/smefinance

[21] D. Leslie, "Understanding artificial intelligence ethics and safety," The Alan Turing Institute, 2019. [Online]. Available: https://www.turing.ac.uk/sites/default/files/2019-06/understanding_artificial_intelligence_ethics_and_safety.pdf

[22] L. Floridi *et al.*, "AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations," *Minds Mach.*, vol. 28, pp. 689–707, 2018.

[23] J. Morley, L. Floridi, L. Kinsey, and L. A. Elhalal, "From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices," *Sci. Eng. Ethics*, vol. 26, pp. 2141–2168, 2020.

[24] M. Astobiza, M. Toboso, M. Aparicio, and D. López, "AI ethics for sustainable development goals," *IEEE Technol. Soc. Mag.*, vol. 40, no. 2, pp. 66–71, Jun. 2021.

[25] T. Araujo, N. Helberger, S. Kruikemeier, and C. H. De Vreese, "AI we trust? Perceptions about automated decision-making by artificial intelligence," *AI Soc.*, vol. 35, no. 3, pp. 611–623, 2020.

[26] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," *ACM Comput. Surv.*, vol. 54, no. 6, 2021, [Online]. Available: https://doi.org/10.1145/3457607

[27] Centre for Data Ethics and Innovation, Review into bias in algorithm decision-making, 2020. [Online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/957259/Review_into_bias_in_algorithmic_decision-making.pdf

[28] A. Bibal, M. Lognoul, A. De Streel, and A. B. Frénay, "Legal requirements on explainability in machine learning," *Artif. Intell. Law*, vol. 29, no. 2, pp. 149–169, 2021.

[29] N. Burkart and M. F. Huber, "A survey on the explainability of supervised machine learning," *J. Artif. Intell. Res.*, vol. 70, pp. 245–317, 2021.

[30] Office for AI and World Economic Forum, "AI procurement guidelines," 2020. [Online]. Available: https://www.gov.uk/government/publications/guidelines-for-ai-procurement

[31] U.K.-Gov Department of Health and Social Care, "A guide to good practice for digital and data-driven health technologies – updated 2021," 2021. [Online]. Available: https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology

[32] U.K. Government, "AI roadmap," 2021. [Online]. Available: https://www.gov.uk/government/publications/ai-roadmap

[33] U.K. Government, "Public sector guidance: Ethics, transparency and accountability framework for automated decision-making," 2021. [Online]. Available: https://www.gov.uk/government/publications/ethics-transparency-and-accountability-framework-for-automated-decision-making/ethics-transparency-and-accountability-framework-for-automated-decision-making

[34] U.K. Government, "Government response to the consultation on the National Data Strategy," 2021. [Online]. Available: https://www.gov.uk/government/consultations/uk-national-data-strategy-nds-consultation/outcome/government-response-to-the-consultation-on-the-national-data-strategy

[35] K. Crockett, *Adaptive Psychological Profiling from Non-Verbal Behaviour – "Why are Ethics Just not Enough to Build Trust?", 2021, Women in Computational Intelligence*, A. Smith, Eds. New York, NY, USA: Springer, 2021.

[36] European Union, "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, 2021. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

[37] D. Leslie, C. Burr, M. Aitken, J. Cowls, M. Katell, and M. Briggs, "AI, human rights, democracy and the rule of law: A primer prepared for the Council of Europe," The Alan Turing Institute, 2021. [Online]. Available: https://www.turing.ac.uk/research/publications/ai-human-rights-democracy-and-rule-law-primer-prepared-council-europe

[38] Al. Moltzau, "A typology of AI ethics tools, methods and research," 2019. [Online]. Available: https://towardsdatascience.com/a-typology-of-ai-ethics-tools-methods-and-research-cacdea134503

[39] J. Saldaña, *The Coding Manual For Qualitative Researchers*, Thousand Oaks, CA, USA: Sage, 2021.

[40] K. M. MacQueen, E. McLellan-Lemal, K. Bartholow, and B. Milstein, "Teambased codebook development: Structure, process, and agreement," in *Handbook For Team-Based Qualitative Research*, G. Guest and K. M. MacQueen, Eds. Lanham, MD, USA: AltaMira Press, pp. 119–135, 2008.

[41] GM AI Foundry, 2021. [Online]. Available: https://gmaifoundry.ac.uk/about/

[42] T. Gebru *et al.*, "Datasheets for datasets," 2018, *arXiv:1803.09010.* [Online]. Available: https://arxiv.org/abs/1803.09010

[43] DotEveryone, "Consequence scanning," 2018. [Online]. Available: https://doteveryone.org.uk/project/consequence-scanning/

[44] ICO, "Data protection impact assessments," 2020. [Online]. Available: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments/

[45] ICO, "What do we need to do to ensure lawfulness, fairness, and transparency in AI systems?," 2020. [Online]. Available: https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/guidance-on-ai-and-data-protection/what-do-we-need-to-do-to-ensure-lawfulness-fairness-and-transparency-in-ai-systems/

[46] MMU, "Ethical AI practice survey," 2021. [Online]. Available: https://mmu.onlinesurveys.ac.uk/ai-ethics-survey

[47] Policy Connect, "Our place our data: Involving local people in data and AI based recovery," 2021. [Online]. Available: https://www.policyconnect.org.uk/research/our-place-our-data-involving-local-people-data-and-ai-based-recovery

[48] ICO, "What are the accountability and governance implications of AI?," 2020. [Online]. Available: https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/guidance-on-ai-and-data-protection/what-are-the-accountability-and-governance-implications-of-ai/

[49] ICO, "How do we ensure individual rights in our AI systems?," 2020. [Online]. Available: https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/guidance-on-ai-and-data-protection/how-do-we-ensure-individual-rights-in-our-ai-systems/

[50] N. Aoki, "The importance of the assurance that 'humans are still in the decision loop' for public trust in artificial intelligence: Evidence from an online experiment," in *Computers in Human Behavior*. New York, NY, USA: Elsevier, 2021, Art. no. 106572. [Online]. Available: https://doi.org/10.1016/j.chb.2020.106572

[51] A. Jacovi, A. Marasović, T. Miller, and Y. Goldberg, "Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI," in *Proc. 2021 ACM Conf. Fairness, Accountability, Transparency*, 2021, pp. 624–635.

[52] a3i, "The trust in AI framework," 2018. [Online]. Available: http://a3i.ai/trust-in-ai

[53] AI Now Institute, "Algorithmic accountability policy toolkit," 2018. [Online]. Available: https://ainowinstitute.org/aap-toolkit.pdf

[54] The Institute for Ethical AI & ML, "AI-RFX procurement framework," 2018. [Online]. Available: https://github.com/EthicalML/XAI

[55] T. Arnold and M. Scheutz, "The 'big red button' is too late: An alternative model for the ethical evaluation of AI systems," *Ethics Inf. Technol.*, vol. 2, no. 1, pp. 59–69, 2018, doi: 10.1007/s10676-018-9447-7.

[56] HAI, "Agile ethics for AI trello board," 2021. [Online]. Available: https://trello.com/b/SarLFYOd/agile-ethics-for-ai-hai

[57] FAT/ML, "Principles for accountable algorithms and a social impact statement for algorithms," 2020. [Online]. Available: https://www.fatml.org/resources/principles-for-accountable-algorithms

[58] N. Diakopoulos, D. Trielli, and G. Lee, "Algorithm tips," 2018. [Online]. Available: http://algorithmtips.org/

[59] Z. Epstein *et al.*, "TuringBox: An experimental platform for the evaluation of AI systems," in *Proc. 27th Int. Joint Conf. Artif. Intell. Demos*, 2018, pp. 5826–5828.

[60] J. Glenn, "Futures wheel," 2018. [Online]. Available: http://ethicskit.org/futures-wheel.html

[61] P. Hall and N. Gill, "An introduction to machine learning interpretability," *O'Reilly*, 2019. [Online]. Available: https://www.h2o.ai/wp-content/uploads/2019/08/An-Introduction-to-Machine-Learning-Interpretability-Second-Edition.pdf

[62] Ethics Kit, "Ethics toolkit," 2021. [Online]. Available: http://ethicskit.org/tools.html

[63] The Data Nutrition Project. [Online]. Available: https://datanutrition.org/

[64] IDEO.ORG, "Design kit," [Online]. Available: https://www.designkit.org/case-studies

[65] IEEE, "Ethically aligned design," 2018. [Online]. Available: https://ethicsinaction.ieee.org/

[66] OPAL, "Open algorithms," 2021. [Online]. Available: https://www.opalproject.org/about-opal

[67] Moral Machines, 2018. [Online]. Available: https://www.moralmachine.net/

[68] M. Mitchell *et al.*, "Model cards for model reporting," in *Proc. Conf. Fairness, Accountability, Transparency*, 2019, pp. 220–229.

[69] The Federation, "New economic impact model," 2019. [Online]. Available: http://ethicskit.org/downloads/economy-impact-model.pdf

[70] ODI, "Data ethics canvas," 2021. [Online]. Available: https://theodi.org/article/the-data-ethics-canvas-2021/

[71] C. Oxborough, E. Cameron, A. Rao, A. Birchall, A. Townsend, and C. Westermann, "Explainable AI: Driving business value through greater understanding," *Price Waterhouse Cooper*, 2018. [Online]. Available: https://www.pwc.co.uk/audit-assurance/assets/explainable-ai.pdf

[72] D. Peters and R. A. Calvo, "Beyond principles: A process for responsible tech," 2019. [Online]. Available: https://medium.com/ethics-of-digital-experience/beyond-principles-a-process-for-responsible-tech-aefc921f7317

[73] D. Peters, R. A. Calvo, and R. M. Ryan, "Designing for motivation, engagement and wellbeing in digital experience," *Front. Psychol.*, vol. 9, 2018, Art. no. 797.

[74] D. Reisman, J. Schultz, K. Crawford, and M. Whittaker, "Algorithmic impact assessments: A practical framework for public agency accountability," *AINow*, 2018. [Online]. Available: https://ainowinstitute.org/aiareport2018.pdf

[75] Responsible AI Licenses, 2021. [Online]. Available: https://www.licenses.ai/

[76] Royal Society and British Academy, "Data management and use: Governance in the 21st century," 2018. [Online]. Available: https://royalsociety.org/-/media/policy/projects/data-governance/data-management-governance.pdf

[77] B. C. Stahl and D. Wright, "Ethics and privacy in AI and big data: Implementing responsible research and innovation.," *IEEE Secur. Priv.*, vol. 16, no. 3, pp. 26–33, May/Jun. 2018. [Online]. Available: https://doi.org/10.1109/MSP.2018.2701164

[78] IBM, "IBM 360 fairness," 2019. [Online]. Available: https://www.ibm.com/blogs/research/2019/08/ai-explainability-360/

[79] EthicalML, "XAI library," 2018. [Online]. Available: https://github.com/EthicalML/xai

[80] W. Zhao, "Improving social responsibility of artificial intelligence by using ISO 26000," in *Proc. Published Under License by IOP Publishing Ltd IOP Conf. Ser.: Mater. Sci. Eng., Vol. 428, 3rd Int. Conf. Automat., Control Robot. Eng.*, 2018, Art. no. 012049. [Online]. Available: https://iopscience.iop.org/article/10.1088/1757-899X/428/1/012049

[81] M. Zook *et al.*, "Ten simple rules for responsible big data research," 2017. [Online]. Available: https://collaborate.princeton.edu/en/publications/ten-simple-rules-for-responsible-big-data-research

[82] Ada Lovelace Institute, "JUST AI reflection prototype," 2021. [Online]. Available: https://www.adalovelaceinstitute.org/project/just-ai-reflection-prototype/

[83] Information Commissioners Office and Alan Turing Institute, "Guidance on explaining AI decisions," 2020. [Online]. Available: https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-ai/

[84] Unbiased, "AI4DM toolkits," 2020. [Online]. Available: http://proboscis.org.uk/6473/special-offer-on-unbias-ai4dm-toolkits/

[85] OECD, "Public consultation on the OECD framework for classifying AI systems," 2021. [Online]. Available: https://oecd.ai/classification

[86] E. Blasch, J. Sung, and T. Nguyen, "Multisource AI scorecard table for system evaluation," 2021, *arXiv:2102.03985*. [Online]. Available: https://arxiv.org/abs/2102.03985

[87] Microsoft, "Allovus design, fairlearn: A toolkit for assessing and improving fairness in AI," 2020. [Online]. Available: https://www.microsoft.com/en-us/research/uploads/prod/2020/05/Fairlearn_WhitePaper-2020-09-22.pdf

[88] Digital Catapult, "Machines for machine intelligence," 2021. [Online]. Available: https://www.digicatapult.org.uk/for-startups/acceleration-programmes/machine-intelligence-garage

[89] ACLU Washington, "Algorithmic equity toolkit," 2019. [Online]. Available: https://www.aclu-wa.org/AEKit

[90] AI Global, "Responsible AI design assistant beta," 2021. [Online]. Available: https://oproma.github.io/rai-trustindex/

[91] Data ethics, "Data ethics impact assessment," 2021. [Online]. Available: https://dataethics.eu/wp-content/uploads/dataethics-impact-assessment-2021.pdf

[92] Deon, "An ethics checklist for data scientists," 2021. [Online]. Available: https://deon.drivendata.org/

[93] M. Arnold *et al.*, "FactSheets: Increasing trust in AI services through supplier's declarations of conformity," *IBM J. Res. Develop.*, vol. 63, no. 4/5, pp. 6:1–6:13, Jul.–Sep. 2019, doi: 10.1147/JRD.2019.2942288.

[94] Google, "Playing with AI fairness tool," 2021. [Online]. Available: https://pair-code.github.io/what-if-tool/ai-fairness.html

[95] Google, "Explainable AI beta tools and frameworks," 2021. [Online]. Available: https://cloud.google.com/explainable-ai/

[96] VDE Bertelsmann Stiftung, "From principles to practice – an interdisciplinary framework to operationalise AI ethics," 2020. [Online]. Available: https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/WKIO_2020_final.pdf

[97] B. Goefring, F. Rossi, and D. Zaharchuk, "Advancing AI ethics beyond compliance from principles to practice," *IBM*, 2020. [Online]. Available: https://www.ibm.com/downloads/cas/J2LAYLOZ

[98] ICO, "Guidance on AI and data protection," 2021. [Online]. Available: https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/guidance-on-ai-and-data-protection/

[99] EthicalOS, "Ethical OS starter checklist," 2021. [Online]. Available: https://ethicalos.org/
[100] I. D. Raji *et al.*, "Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing," in *Proc. Conf. Fairness, Accountability, Transparency*, 2020, pp. 33–44.
[101] The Institute for Ethical AI and Machine Learning, "The AI-RFX procurement framework," 2021. [Online]. Available: https://ethical.institute/rfx.html
[102] Design Ethically, "A library of resources and toolkits to help you integrate ethical design into your practice," 2021. [Online]. Available: https://www.designethically.com/toolkit
[103] M. Madaio, L. Stark, J. Vaughan, and H. Wallach, "Co-designing checklists to understand organizational challenges and opportunities around fairness in AI," in *Proc. CHI Conf. Humans Factors Comput. Syst.*, 2020, pp. 1–14.
[104] Price Water House Cooper, "Responsible AI diagnostic tool," 2021. [Online]. Available: https://www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/what-is-responsible-ai.html
[105] Smart Dubai AI Systems, "Ethics self-assessment tool," 2021. [Online]. Available: https://www.smartdubai.ae/self-assessment
[106] Aequitas, "Bias and fairness audit toolkit," 2021. [Online]. Available: https://github.com/dssg/aequitas
[107] World Economic Forum, "AI procurement in a box: Workbook," 2020. [Online]. Available: http://www3.weforum.org/docs/WEF_AI_Procurement_in_a_Box_Workbook_2020.pdf
[108] World Economic Forum, "Empowering AI leadership, an oversight toolkit for boards of directors," 2020. [Online]. Available: https://spark.adobe.com/page/RsXNkZANwMLEf/
[109] 510, "F.A.C.T score for responsible AI," 2020. [Online]. Available: https://www.510.global/f-a-c-t-score-for-responsible-ai/
[110] Microsoft, "InterpretML a toolkit for understanding machine learning models," 2020. [Online]. Available: https://www.microsoft.com/en-us/research/uploads/prod/2020/05/InterpretML-Whitepaper.pdf
[111] Microsoft, "Responsible innovation: A best practices toolkit," 2020. [Online]. Available: https://docs.microsoft.com/en-us/azure/architecture/guide/responsible-innovation/
[112] S. Vallor, "An ethical toolkit for engineering/design practice, 2018. [Online]. Available: https://www.scu.edu/ethics-in-technology-practice/ethical-toolkit/
[113] AI Ethics Lab, "Ethics training," 2019. [Online]. Available: https://aiethicslab.com/training/
[114] U.K. Government, "Data ethics framework," 2018. [Online]. Available: https://www.gov.uk/government/publications/data-ethics-framework
[115] Linklaters, "A toolkit for artificial intelligence (AI) projects," 2021. [Online]. Available: https://www.linklaters.com/en/insights/thought-leadership/artificial-intelligence-toolkit/ethical-safe-legal---a-toolkit-for-artificial-intelligence-projects
[116] Rolls Royce, "The Aletheia framework," 2021. [Online]. Available: https://www.rolls-royce.com/sustainability/ethics-and-compliance/the-aletheia-framework.aspx
[117] NetHope, "Artificial intelligence (AI) ethics for nonprofits toolkit," 2020. [Online]. Available: https://solutionscenter.nethope.org/artificial-intelligence-ethics-for-nonprofits-toolkit
[118] Open Roboethics Institute, "AI ethics assessment toolkit," 2019. [Online]. Available: https://openroboethics.org/ai-toolkit/
[119] D. Anderson, J. Bongaguro, M. McKinney, A. Nicklin, and J. Wiseman, "Ethics & algorithms toolkit: A risk management framework for governments (and other people too!)," 2018. [Online]. Available: https://ethicstoolkit.ai/
[120] RSA, "Democratizing decisions about technology: A toolkit," 2019. [Online]. Available: https://www.thersa.org/globalassets/reports/2019/democratising-decisions-tech-report.pdf
[121] Nesta, "Civic AI toolkit," 2021. [Online]. Available: https://www.nesta.org.uk/toolkit/civicai/
[122] P. M. Krafft *et al.*, "An action-oriented AI policy toolkit for technology audits by community advocates and activists," in *Proc. ACM Conf. Fairness, Accountability, Transparency*, 2021, pp. 772–781.
[123] Omidyar Network, "Ethical explorer," 2020. [Online]. Available: https://ethicalexplorer.org/additional-resources-for-ethical-explorers/
[124] IDEO, "AI & ethics: Collaborative activities for designers," 2019. [Online]. Available: https://www.ideo.com/post/ai-ethics-collaborative-activities-for-designers
[125] Covington, "Artificial intelligence toolkit," 2021. [Online]. Available: https://www.cov.com/en/practices-and-industries/industries/artificial-intelligence/toolkit
[126] Inter-American Development Bank, "Ethical assessment of AI for actors within the entrepreneurial ecosystem," 2021. [Online]. Available: https://publications.iadb.org/publications/english/document/Ethical-Assessment-of-AI-for-Actors-within-the-Entrepreneurial-Ecosystem-Application-Guide.pdf
[127] EthicsCanvas.org, "Online ethics canvas," 2021. [Online]. Available: https://www.ethicscanvas.org/index.html
[128] Microsoft, "Responsible AI principles," 2021. [Online]. Available: https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3aprimaryr6
[129] L. Ouchchy, A. Coin, and V. Dubljević, "AI in the headlines: The portrayal of the ethical issues of artificial intelligence in the media," *AI Soc.*, vol. 35, pp. 927–936, 2020.
[130] SalesforceUX, "How to run a consequence scanning workshop," 2020. [Online]. Available: https://medium.com/salesforce-ux/how-to-run-a-consequence-scanning-workshop-4b14792ea987
[131] Digital Catapult, "Loomi: The artificial intelligence assistant," 2021. [Online]. Available: https://www.digicatapult.org.uk/for-startups/success-stories/loomi
[132] Center for Data Innovation, "How much will the artificial intelligence act cost Europe?," 2021. [Online]. Available: https://www2.datainnovation.org/2021-aia-costs.pdf
[133] Y. Murphy, S. Garg, B. Sniderman, and T. Buckley, "Ethical technology use in the fourth industrial revolution," 2019. [Online]. Available: https://www2.deloitte.com/us/en/insights/focus/industry-4-0/ethical-technology-use-fourth-industrial-revolution.html
[134] M. L. Waskom, "Seaborn: Statistical data visualization," *J. Open Source Softw.*, vol. 6, no. 60, 2021, Art. no. 3021.
[135] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Comput. Sci. Eng.*, vol. 9, no. 03, pp. 90–95, 2007.
[136] W. McKinney, "Data structures for statistical computing in python," in *Proc. 9th Python Sci. Conf.*, 2010, pp. 51–56.

Keeley Crockett (Senior Member, IEEE) has over 20 years' experience in research and development of computational intelligence algorithms and applications, including adaptive psychological profiling, fuzzy systems, dialogue systems, and educational tutoring systems. She is currently a Professor of computational intelligence with Manchester Metropolitan University, Manchester, U.K., and Coacademic lead with the ERDF-funded Greater Manchester AI Foundry.

Prof. Crockett is the current Chair of the IEEE Taskforce on Ethical and Social Implications of Computational Intelligence and IEEE Women in Computational Intelligence and a STEM Ambassador.

Edwin Colyer (Senior Member, IEEE) is an Impact and Engagement Manager with the Research and Knowledge Exchange Directorate, Manchester Metropolitan University, Manchester, U.K. He is responsible for the mobilization of research knowledge, connecting and collaborating with stakeholders, potential users, and beneficiaries to drive evidence-informed societal change.

Luciano Gerber is a Senior Lecturer of computer science with Manchester Metropolitan University, Manchester, U.K. He is currently working on developing AI solutions for SMEs within the ERDF-funded Greater Manchester AI Foundry. His research interests include fundamental and domain-agnostic applied data science and machine learning.

Annabel Latham (Senior Member, IEEE) is a Senior Lecturer with the Department of Computing and Mathematics, Manchester Metropolitan University, Manchester, U.K. Her research interests include conversational agents, intelligent tutoring systems, affective computing, and ethics of AI in education, user profiling, and computational intelligence.

Dr. Annabel is the Chair of IEEE U.K. and Ireland Women in Engineering, Chair of the IEEE CIS Education Repository subcommittee, and a STEM Ambassador.