Ethics in the Age of AI: An Analysis of AI Practitioners' Awareness and Challenges

AASTHA PANT, RASHINA HODA, SIMONE V. SPIEGLER, and CHAKKRIT TANTITHAMTHAVORN, Monash University, Australia
BURAK TURHAN, University of Oulu, Finland

Ethics in AI has become a debated topic of public and expert discourse in recent years. But what do people who build AI—AI practitioners—have to say about their understanding of AI ethics and the challenges associated with incorporating it into the AI-based systems they develop? Understanding AI practitioners' views on AI ethics is important as they are the ones closest to the AI systems and can bring about changes and improvements. We conducted a survey aimed at understanding AI practitioners' *awareness* of AI ethics and their *challenges* in incorporating ethics. Based on 100 AI practitioners' responses, our findings indicate that the majority of AI practitioners had a *reasonable* familiarity with the concept of AI ethics, primarily due to *workplace rules and policies*. *Privacy protection and security* was the ethical principle that the majority of them were aware of. Formal education/training was considered *somewhat* helpful in preparing practitioners to incorporate AI ethics. The challenges that AI practitioners faced in the development of *ethical* AI-based systems included (i) general challenges, (ii) technology-related challenges, and (iii) human-related challenges. We also identified areas needing further investigation and provided recommendations to assist AI practitioners and companies in incorporating ethics into AI development.

CCS Concepts: • **Software and its engineering** → *Software design engineering;*

Additional Key Words and Phrases: AI ethics, AI practitioners, awareness, challenges, survey

ACM Reference format:
Aastha Pant, Rashina Hoda, Simone V. Spiegler, Chakkrit Tantithamthavorn, and Burak Turhan. 2024. Ethics in the Age of AI: An Analysis of AI Practitioners' Awareness and Challenges. *ACM Trans. Softw. Eng. Methodol.* 33, 3, Article 80 (March 2024), 35 pages.
https://doi.org/10.1145/3635715

1 INTRODUCTION

AI technology is becoming increasingly pervasive across industries and contexts, making AI ethics more important than ever [17, 36]. There is no universal definition of AI ethics and ethical principles [26, 55]. In our study, we used the definition of AI ethics provided by Siau and Wang [50] who

Aastha Pant is supported by the Faculty of IT Ph.D. scholarship from Monash University. C. Tantithamthavorn is partially supported by the Australian Research Council's Discovery Early Career Researcher Award (DECRA) funding scheme (DE200100941).
Authors' addresses: A. Pant and S. V. Spiegler, R. Hoda, and C. Tantithamthavorn, Monash University, Wellington Road, Melbourne, Victoria 3800, Australia; e-mails: {aastha.pant, simone.spiegler, rashina.hoda, chakkrit}@monash.edu; B. Turhan, University of Oulu, Pentti Kaiteran katu 1, 90570 Oulu, Finland, P.O. Box 3000; e-mail: burak.turhan@oulu.fi.

1049-331X/2024/03-ART80 $15.00
https://doi.org/10.1145/3635715

defined it as, "*the principles of developing AI to interact with other AIs and humans ethically and function ethically in society*". We used this definition of AI ethics because it emphasises the responsibility of AI practitioners to develop AI systems that are both effective and ethical. It captures the important societal influence of AI while also acknowledging the increasing interaction between AI systems that is giving rise to new ethical challenges such as unprecedented automation of online scamming through AI controlling other AI systems [29]. It encourages the development of AI systems that contribute positively to societal well-being and mitigate potential negative consequences. The importance of ethical consideration in AI has been highlighted by several incidents in recent years [4]. For example, there have been ethical issues such as Google's **Machine learning** (**ML**) algorithm that was gender-biased against women as it associated men with **Science, Technology, Engineering, and Mathematics** (**STEM**) careers more frequently than women [45]; GitHub's use of copyrighted source code as training data for their ML-powered GitHub Copilot which is an example of a privacy issue [40]; Amazon's AI-powered recruitment tool that was gender-biased [33]. All these incidents demonstrate the repercussions of neglecting ethics in AI development, emphasising the necessity of incorporating ethical considerations in AI to prevent similar incidents in the future.

The extensive use of AI-based systems in various fields and the ethical incidents of AI-based systems have increased the number of studies in this area. Various studies have been conducted in the area of AI ethics and most studies are theoretical and conceptual in nature [6]. For instance, studies have been carried out to analyse ethical principles of AI [13, 24, 32, 50]. Studies have been conducted with the aim of enhancing the development of ethical AI-based systems or minimising ethical issues that are highly prevalent in the current era. Studies have also proposed several frameworks, models, and methods to assist AI practitioners in incorporating ethics into AI-based systems such as maturity model [56], ECCOLA method [59], AI ethics framework [12]. Likewise, several AI ethical guidelines and principles have been developed and formulated [14, 22, 35]. Despite the abundance of guidelines and studies in the field of AI ethics, various ethical problems associated with AI systems continue to persist. The mere development of guidelines is insufficient; AI practitioners[1] must also adhere to them during the development of AI-based systems. It is therefore imperative to examine whether they are following these guidelines and taking the necessary steps in the development of such systems. While conducting theoretical research is crucial, it's equally essential to conduct empirical studies to understand the views of AI practitioners on AI ethics since it ultimately relies on the practitioners to adhere to ethical principles of AI [61].

Recent studies have discussed several aspects of AI practitioners' awareness of ethics in AI. For example, Stahl et al. [52], Vakkuri et al. [58] discussed that AI practitioners were aware of the concept of ethics in AI, its importance, and its relevance. Several tools and methods have been developed to raise the awareness of AI ethics among AI practitioners [37]. Likewise, studies have been conducted to focus on investigating the challenges related to adhering to specific ethical principles of AI such as *fairness*, *accountability*, and *privacy* [31, 41, 46]. The significance of AI ethical principles and challenges associated with the implementation of those principles have been conducted through an empirical investigation with AI practitioners and lawmakers [27]. Although several studies have reported the challenges of AI practitioners in incorporating ethics in AI, most of those studies focused on investigating the challenges of AI practitioners and other stakeholders related to specific ethical principles of AI. There is a lack of research that focuses on investigating the overall challenges of AI practitioners in incorporating ethics in AI.

[1]The term "practitioners" in our study includes AI developers, engineers, specialists, experts, and designers involved in the design and development activities of AI-based systems. The terms "AI practitioners" and "AI developers" are used interchangeably throughout our study.

Previously, we conducted a **Grounded Theory Literature Review** (**GTLR**) of 38 empirical articles to gain insights into research studies that focused on investigating AI practitioners' views on AI ethics and developed a *taxonomy of ethics in AI from developers' viewpoints* spanning five categories: (i) *developer awareness*, (ii) *developer perception*, (iii) *developer need* (iv) *developer challenge*, and (v) *developer approach* [42]. In this study, we conducted a **survey** with AI practitioners to investigate two aspects (categories) including (i) *developer awareness* of ethics in AI and (ii) *developer challenge* in incorporating ethics in AI that we derived through our GTLR. The two **research questions** (**RQs**) of this study are as follows:

RQ1. How aware are AI practitioners of different aspects related to AI ethics?
To answer this RQ, we investigated: (i) the extent to which AI practitioners are aware of the "AI ethics" concept, (ii) what ethical principles AI practitioners are aware of, (iii) reasons for AI practitioners' awareness of ethics and (iv) awareness of the role of formal education/training in preparing AI practitioners to incorporate ethics in AI.

RQ2. What challenges/barriers do AI practitioners face in incorporating ethics in AI?
To answer this RQ, we investigated the degree of challenges AI practitioners face in considering and following each AI ethical principle using Australia's AI Ethics Principles[2] as a reference. We used Australia's AI Ethics Principles as a basis for investigation due to three main reasons: (i) Australia's AI Ethics Principles cover a good range of ethical issues ranging from human to environmental well-being, some of the most popular ethical concepts including fairness, privacy, transparency, as well as some more rare but critical ones such as contestability and accountability; (ii) there is no unified set of AI ethics principles that everyone around the world follows since different countries and organisations have their own; and (iii) when we first advertised the survey on social media platforms like LinkedIn and X (formerly known as Twitter), we were more likely to get participants from Australia because we are based in Australia and our connections are mostly from there. So we used Australia's AI ethics principles as we expected to gain maximum responses from the participants based in Australia. We moved to do a second round of data collection using Prolific since we didn't get the response we expected in the first round. At this stage, we kept the selected principles to maintain continuity because the Australian AI Ethics Principles cover a broad range of key areas that cut well across other sets of principles used around the world [14, 16, 22]. We also explored the overall key challenges/barriers that AI practitioners face in incorporating ethics in AI. Our survey contained 15 questions in total (12 closed-ended and 3 open-ended questions). We collected data in two rounds. In the first round, we advertised our survey on social media platforms such as LinkedIn and X and obtained 17 responses, and in the second round, we collected the data for our study using the Prolific platform and obtained 83 responses, making the total number of responses to our study 100. We used ***Socio-Technical Grounded Theory (STGT)*** *method for data analysis* [18] to analyse the qualitative data and descriptive statistics to analyse the quantitative data. The main contribution of our study is that we designed a set of recommendations for further research on AI practitioners' *awareness* and *challenges* around ethics in AI that would be beneficial for both AI practitioners and AI researchers to enhance AI practitioners' awareness of AI ethics and the challenges in incorporating ethics in AI. We also provide recommendations to AI educators.

The rest of the article is organised as follows. Section 2 presents the related work followed by the research methodology in Section 3. We present the findings in Section 4 followed by a discussion on key findings and recommendations in Section 5. Then, we provide the limitations and threats to the validity of our study in Section 6 which is followed by a conclusion in Section 7.

[2]https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles

2 RELATED WORK

In this section, we provide a summary of the previous research.

2.1 AI Practitioners' Awareness of Ethics in AI

Studies have been conducted to investigate the awareness of AI practitioners of ethics and ethical principles of AI. The majority of these studies have focused on understanding if AI practitioners are aware of the concept of AI ethics and its importance in AI development. Vakkuri et al. [62] conducted semi-structured interviews with AI practitioners and concluded that they were aware of the concept of *ethics* in AI and its importance. Stahl et al. [52] applied a case study research strategy and reported that AI practitioners were aware of the relevance of *ethics* in AI. Govia [15] conducted interviews and field observations with AI specialists to investigate the social or ethical implications of AI based on their working experience in AI development. They found that AI specialists were aware of the philosophical theories of AI ethics.

Studies have also reported the awareness of AI practitioners on different ethical principles of AI. Vakkuri et al. [62] conducted interviews with AI practitioners from five case companies to understand the practices used by them to incorporate ethics in AI. They found that the participants of all five case companies were aware of and concerned about the issues related to system *transparency*, which is an AI ethical principle. Christodoulou and Iordanou [8] conducted six focus groups with 63 AI experts to investigate the ethical issues of digital media with a focus on AI and Big Data. They reported that the participants were aware of "*Transparency*" as an ethical principle of AI. Vakkuri et al. [57] reported that AI practitioners had knowledge about the term "transparent AI" as it was one of their goals of AI development. Mark and Anya [32] discussed how participants were knowledgeable about the transparency law, which assisted them in identifying which data should be disclosed and which data should be kept confidential when creating an AI system and strived to create transparent systems. Likewise, Holstein et al. [19] conducted semi-structured interviews with 35 ML practitioners and a survey with 267 ML practitioners to investigate their challenges and needs to develop fair ML systems. They found that the majority of them were aware of "*Fairness*" principle and possessed knowledge of fairness-related issues of ML systems. Veale et al. [63] carried out interviews with 27 ML practitioners across 5 OECD countries regarding the challenges and design needs for algorithmic support in high-stakes public sector decision-making and found that the practitioners were aware of *"Fairness*" AI ethical principle and worked towards abolishing fairness issues in AI systems. The participants were also aware of "*Accountability/ Responsibility*" and its importance in the development and deployment of AI-based systems and took responsibility for any harm caused by their creations. Ibáñez and Olmeda [21] conducted 22 semi-structured interviews and 2 focus groups with AI practitioners to investigate how AI companies bridge the gap between AI ethical principles and practice. They found that "*Privacy*" is the principle that AI practitioners were most aware of and extensively discussed, with data and information privacy being a significant concern for organisations. Ryan et al. [48] conducted a multiple case study analysis to investigate the ethical concerns arising from the implementation and use of Big data and AI. They found that "*Privacy*" was the only ethical principle that the participants of all 10 case studies were aware of and discussed ethical issues related to it.

Most of the previous work has focused on understanding whether AI practitioners are aware of the concept of AI ethics and understanding AI practitioners' awareness of specific ethical principles of AI. There still remain areas that need to be investigated when it comes to the awareness of AI practitioners on AI ethics so that effective strategies can be developed to enhance the awareness of AI ethics among AI practitioners.

2.2 AI Practitioners' Challenges in Incorporating Ethics in AI

Studies have reported several challenges that AI practitioners face in incorporating ethics during the development of AI-based systems. However, the main focus of these studies was not to investigate the specific challenges faced by AI practitioners in incorporating ethics in AI. For example, Vakkuri et al. [62] and Govia [15] conducted empirical studies to explore the practices used by AI practitioners in incorporating ethics into AI development and investigate the social and ethical implications of AI, respectively. In doing that, they discovered several challenges discussed by the participants related to ethics incorporation in AI. Orr and Davis [41] conducted interviews with 21 AI practitioners to investigate how they attribute and distribute ethical responsibility for AI systems. During that process, several AI practitioners discussed the challenges they face in developing ethical AI systems including challenges related to organisational norms, legislative regulations, users, and AI machines. Ibáñez and Olmeda [21] conducted an empirical study to understand the gap between principles and practices in AI ethics. They also explored various challenges related to translating each AI ethical principle into practice. Sanderson et al. [49] carried out semi-structured interviews with AI designers and developers and discovered several challenges and insights related to translating each of Australia's AI Ethics Principles.[2]

Studies have investigated the challenges faced by AI practitioners in incorporating specific ethical principles of AI. For instance, Madaio et al. [31] conducted semi-structured interviews and workshops with 33 AI practitioners to investigate the challenges and needs of AI practitioners to assess the fairness of AI-based systems. Holstein et al. [19] conducted semi-structured interviews with 35 ML practitioners and a survey of 267 ML practitioners to investigate the team's challenge and needs in developing fairer ML systems. Rakova et al. [46] conducted semi-structured interviews with AI practitioners with a focus on algorithmic accountability and investigated common challenges, ethical tensions, and effective enablers for responsible AI initiatives. On the other hand, studies have been conducted to explore the challenges related to addressing ethical issues of AI through focus groups with AI engineers [8].

Most studies have either focused on investigating the challenges of AI practitioners in following specific AI ethical principles like fairness [31, 41] and accountability [46], explored the broader issues of addressing ethical concerns in AI [8] or their primary focus was not on investigating AI practitioners' challenges in incorporating AI ethics but have still managed to uncover certain challenges encountered by AI practitioners [15, 62]. It indicated that there is a lack of research that focuses primarily on investigating the *challenges* of AI practitioners in incorporating ethics in AI.

3 RESEARCH METHODOLOGY

We used the survey research method to conduct this study to gather insights from broader AI practitioners on different ethics-related aspects of AI, such as their *awareness* of AI ethics and *challenges* to incorporate ethics in AI [25, 65].

3.1 Survey Design

We aimed to obtain an overview of the participants' perspectives on two aspects of ethics (*awareness* and *challenges*) in AI based on their experiences through our survey (Appendix A). Figure 1 shows the steps that we followed to design the survey to achieve the objective of this study. The survey planning was carried out from August 2022 to October 2022. During this phase, the main tasks performed were defining survey goals and variables, designing the questionnaire through iterative processes, and prioritising important survey questions. Hence, we designed the survey with both open and closed-ended questions. The survey was divided into three main sections and comprised 15 questions (12 closed-ended and 3 open-ended questions). The questions focused on several areas aligned with our future studies.

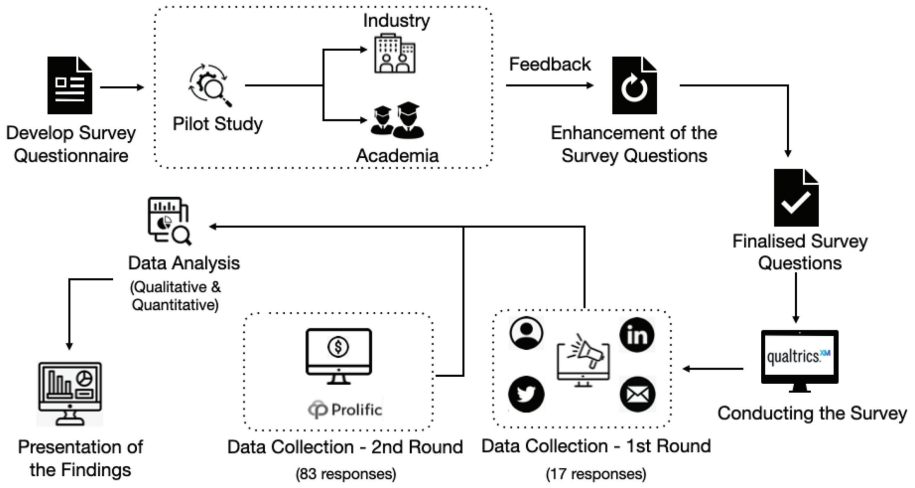

Fig. 1. Overview of the research methodology of the study.

3.1.1 Participant Information. The first section of the survey was designed to collect **basic** demographic information and current employment details of the participants. The survey was anonymous and did not record any personally identifiable information, only information about their gender, age ranges, country of residence, and educational qualifications was collected. Employment details such as job roles/titles and primary responsibilities were collected to identify whether the participants are involved in AI-based software development activities. Participants were asked about their years of experience in AI-based software development to exclude those from the study data who had no experience in that field. All the participants included in our study have at least some experience in AI-based software development.

3.1.2 Understanding the Participants' Perspectives of AI Ethics. The second and third sections were designed to get an overview of the participants' perspectives on two aspects of ethics in AI including their *awareness* of AI ethics and their *challenges* in incorporating ethics in AI. In section two, we focused on identifying how aware AI practitioners are of the concept of AI ethics and AI ethical principles (we asked the participants to rate their familiarity with the concept of "ethics" in AI and report what ethical principles of AI they are aware of based on their experience), reasons for their awareness (we provided them with a list of reasons to become aware of AI ethics that were identified in our previous study [42] and asked them to provide any other reason(s) via open-ended option in the end), and their awareness of the role of formal education/ training (ranging from "Extremely well" to "Not at all") in preparing them to incorporate ethics during the process of developing AI-based systems. The significance of formal education in promoting the ethical implementation of AI has been discussed in the literature [7].

In section three, we aimed to obtain the participants' perspectives on the challenges they encounter when incorporating ethics in AI. We used Australia's AI Ethics Principles[2] as a reference and requested participants to rate the degree of challenges ("Very Challenging" to "No experience") they experienced in adhering to the ethical principles while developing AI-based systems. Moreover, we included an open-ended question to gather the participants' opinions on the primary challenges or barriers they encounter in incorporating ethics in AI. This question enabled us to gain a range of viewpoints on the challenges faced by AI practitioners in incorporating AI ethics.

3.1.3 Pilot Study. Once the survey had been formulated, a preliminary trial was carried out with AI practitioners from our network in order to confirm the questions' clarity and comprehensibility, gauge the amount of time needed to complete the survey and gather their opinions on ways to enhance it. The survey was subsequently circulated among two AI practitioners engaged in AI development activities and two academics with previous involvement in such activities. Taking their suggestions into account, we made some modifications to the survey questions and added some definitions to enhance the clarity of some terms. After incorporating these changes, we finalised the survey and conducted our study.

3.2 Survey Sampling and Data Collection

We adopted a non-probability purposive sampling technique for our study, which entails selecting participants based on particular characteristics rather than their availability [1]. This method allowed us to specifically target our desired group of participants, who are AI practitioners engaged in AI development activities. Our survey was created using the *Qualtrics* platform and we advertised the survey as an anonymous survey link after obtaining the required ethics approval (Reference Number: 34685). The survey questionnaire can be found in Appendix A.

We carried out two rounds of data collection. The first round of data collection was executed from October 2022 to December 2022 and we advertised the survey on social media groups like LinkedIn and X, targeting AI practitioners who are engaged in AI development. We received complete survey responses from 17 AI practitioners who have some level of experience in AI development. Since we didn't receive enough responses in the initial data collection, we decided to carry out a second round. We used the Prolific platform from January 2023 to February 2023 to advertise the survey. As a result, we obtained valuable information from 83 AI practitioners who participated in the survey. The Prolific platform has useful features to filter and select participants, as well as customise monetary incentives for each of the participants. As our aim was to obtain responses from participants who had some level of experience in AI development activities, we employed the participant filtering options of "Employment Sector - Information Technology" and "Employment Status - Full-Time/Part-Time". We provided the participants who completed the survey through Prolific with a reward of 13.30 AUD. Since the survey was shared globally, we obtained responses from various countries, as illustrated in Table 2. As we did not have any particular preference for certain countries, the responses were distributed across various regions. We obtained a majority of the responses from Europe (30%), followed by the participants from other continents like Africa (28%), North America (18%), Asia (12%), and so on. Section 4.1 contains an in-depth analysis of the participants' demographics.

3.3 Data Analysis

We collected both qualitative and quantitative data through our survey. So, we used a mixed-method approach to analyse the survey data. The data analysis types (quantitative/qualitative) that we used to address each RQ in this study are presented in Table 1. We used Microsoft Excel to statistically analyse the quantitative data and organise the qualitative data. Meanwhile, we used the *STGT for Data Analysis* method to analyse the qualitative data [18], which is well-suited for analysing qualitative data, such as those acquired from open-ended text-based survey responses (or *open-text* for short). The purpose of the *STGT for data analysis*, unlike a complete STGT study, is not to develop advanced theories, but rather to identify important patterns in the qualitative data and present them as layered and/or multi-dimensional findings, along with insights and reflections. To do this, we used an *open coding* approach to developing concepts and categories with *constant comparison* of various open-text responses. This approach does not require extensive qualitative data. For example, we gathered open-text responses from 100 participants for the question, "*In your*

Table 1. Data Sources and Analysis Types used to Answer RQs (Descriptive Statistics for Quantitative Data Analysis and STGT for Qualitative Data Analysis)

RQ	Data Source	Data Analysis Type	Purpose of Analysis
RQ1	Closed-ended question	Quantitative analysis	To get an overview of participants' awareness of different aspects related to AI ethics including (i) extent of awareness of "AI ethics" concept, (ii) awareness of AI ethical principles, (iii) reasons for awareness, and (iv) role of formal education/training in preparing AI practitioners to incorporate AI ethics
RQ2	Closed-ended, Follow-up open-ended question	Quantitative analysis Qualitative analysis	To get an overview of the extent to which AI practitioners are challenged to consider and follow each AI ethical principle and the key challenges they face in incorporating AI ethics

experience, what are the main challenges or barriers in incorporating ethics in AI?". These responses were analysed using the *STGT for data analysis* approach. We applied open coding in open-text answers and developed codes as shown in Figure 2. For example, we obtained codes like "*difficulty in predicting AI outcomes*", and "*difficulty in predicting AI consequences*" through open coding. We then engaged in constant comparison, a "*process of constantly comparing derived codes within the same source and across sources to identify key patterns in the data*" [18]. This involved comparing each code with others constantly, leading to revealing patterns among them. For instance, when we compared the two codes mentioned above, the common pattern between them was related to the challenges of human beings foreseeing outcomes and consequences. We combined these two codes to create a concept of "*lack of foresight*". We employed the same approach of constant comparison with the remaining codes and obtained concepts such as "*lack of common perception*", "*lack of knowledge/understanding"*, and "*nature of humans"*. We again constantly compared these concepts to one another and established distinct categories. In this case, these four concepts shared a human-centered aspect in common, leading to their grouping within the "*human-related challenges*" category.

Similarly, we derived multiple codes and concepts that were related to the challenges of technology, as presented in Figure 2. This led to the formation of the other high-level category, namely: "*technology-related challenges*". In this way, we derived a total of three categories of challenges including **general challenges**, **technology-related challenges**, and **human-related challenges**. Detailed information on these challenges is provided in Section 4.

The survey questionnaire was designed by four authors, and all five authors participated in analysing the data and presenting the findings. We engaged in multiple rounds of discussion for each stage to reach decisions. During the analysis phase, the first author analysed the quantitative data and shared the results with the other authors. They discussed the optimal approach to presenting the findings. In addition, in the analysis of the qualitative data, the second and third authors assisted the first author. Once the qualitative data were analysed, the results, including codes, concepts, and categories for the open-ended question, were shared and discussed among all authors who also helped in presenting the findings.

Open coding, constant comparison, and memoing are the steps involved in the STGT method for data analysis. "*Basic memoing is the process of documenting the researcher's thoughts, ideas, and reflections on emerging concepts and (sub)categories and evidence-based conjectures on possible links between them*" [18]. Therefore, we created memos to document important insights and reflections

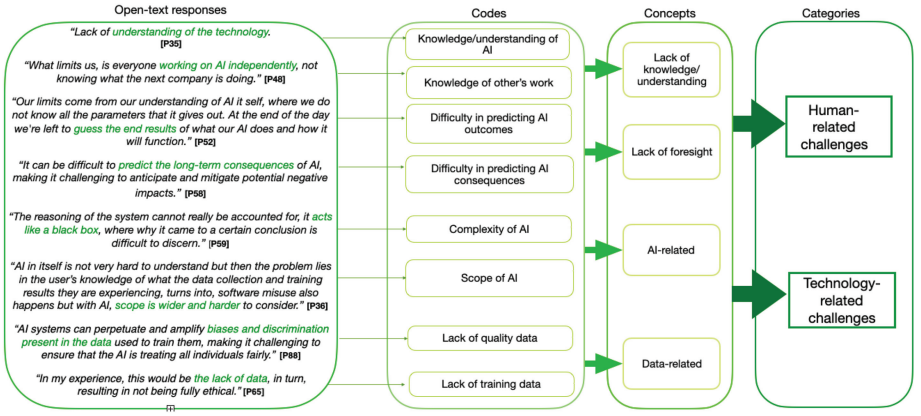

Fig. 2. Examples of STGT analysis [18] applied to qualitative data on challenges/barriers in incorporating ethics in AI.

that we discovered during the process of open coding activities. An example of a memo that we created for the concept "*lack of knowledge/understanding*" is shown below. The discussion on the key insights from memoing is presented in Section 5.3.

> **Memo on "Lack of knowledge/understanding":** "Not enough details here to gather what kind of knowledge is lacking. Some shared include lack of knowledge of AI outputs, ethical implications of AI, and lack of programming skills (although unclear what this means for incorporating ethics - or just that AI implementation is becoming very complex) Link: P97 and P36 mention limitations of human knowledge (about limits of AI - P36); limitations of humans to control and alter AI decisions (P41) and link it to possible misuse of ethical loopholes by AI (see "powerful and inescapable AI" memo). P48 notes a lack of human overview of what others are doing and what we are creating (no overview and coordination). P58 notes a lack of human ability to predict the long-term consequences of AI (no long-term view)."

4 FINDINGS

4.1 Participants' Demographics

In this section, we present the demographics of the survey participants. Table 2 summarises the overall statistics of the participants based on their gender, age, country, work experience, education, job title, and AI development-related activities.

A total of 100 AI practitioners participated in the survey. The majority of the participants were men (71%) whereas only 29% were women. The most common age group of the participants was 26 to 30 years, comprising 31% of the sample followed by the age group ranging from 20 to 25 years (25%). Only 3% of the participants were over the age of 50. Similarly, the majority of the participants (23%) were AI/Data scientists followed by AI developers (19%). A total of 16% of the participants were AI engineers followed by 11% participants who were AI/ML practitioners. The job title of only 3% of the participants was AI/ML specialist whereas 14% of the participants were in the "Other" group. Their job titles included software engineer or software developer. As our target survey participants were AI practitioners involved in AI development activities, we wanted to know the major AI development-related activities they were involved in. The results show that

Table 2. Demographics of the Survey Participants

Gender		**Work Experience in AI Development**	
Men	71%	Less than 1 year	28%
Women	29%	1 to 2 years	43%
		3 to 5 years	19%
Age		6 to 10 years	7%
20 to 25 years	25%	11 to 15 years	3%
26 to 30 years	31%	**Education**	
31 to 35 years	17%	High School	10%
36 to 40 years	16%	Bachelor Degree	56%
41 to 45 years	6%	Master Degree	22%
46 to 50 years	2%	Ph.D. or higher	7%
Above 50 years	3%	Prefer not to answer	1%
Job Title		Others	4%
AI Expert	6%	**Countries**	
AI/ML Specialist	3%	Africa	28%
AI/Data Scientist	23%	North America	18%
AI Designer	8%	Europe	30%
AI/ML Practitioner	11%	South America	3%
AI Engineer	16%	Oceania	9%
AI Developer	19%	Asia	12%
Others	14%	**AI Development-related Activities**	
		Model Requirements	29%
		Data Collection	61%
		Data Cleaning	50%
		Data Labelling	31%
		Feature Engineering	28%
		Model Training	38%
		Model Evaluation	30%
		Model Deployment	27%
		Model Monitoring	21%
		Others	5%

the majority of the participants (61%) were involved in the *Data Collection* activity followed by the *Data Cleaning* activity (50%).

4.2 RQ1 – How Aware are AI Practitioners of Different Aspects Related to AI Ethics?

4.2.1 Extent of Awareness of "AI Ethics" Concept. In our previous work (GTLR) [42], we identified five categories discussing AI practitioners' viewpoints on ethics in AI among which *"developer awareness*" is one of them. Under this category, there are multiple concepts, and developer awareness of "*AI ethics and ethical principles*" is one of the underlining concepts. The main reason to conduct the GTLR (our previous work) over a **Systematic Literature Review** (**SLR**) was due to the lack of empirical studies focusing on investigating AI practitioners' views and understanding of ethics in AI. Therefore, getting an **industry** perspective on this concept by asking them about their level of familiarity with the concept of ethics when it relates to AI development was required. The participants were given five different levels of familiarity to choose from, ranging from "Very familiar" to "Not at all familiar." As shown in Figure 3, out of all participants, 41% had a reasonable familiarity with the concept of AI ethics, followed by 33% who were somewhat familiar with the concept. Conversely, only 13 % of the participants had a high level of familiarity with ethics in AI, 12% of the participants had a low level of familiarity, and only 1% of the participants having no familiarity with the concept at all.

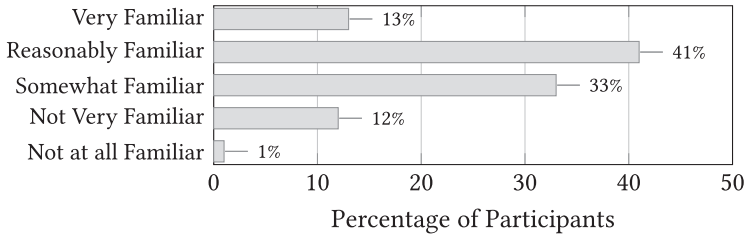

Fig. 3. AI practitioner's degree of familiarity with the concept of AI ethics.

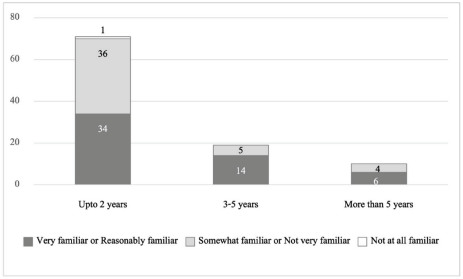

Fig. 4. Participants' familiarity with the AI ethics concept based on work experience.

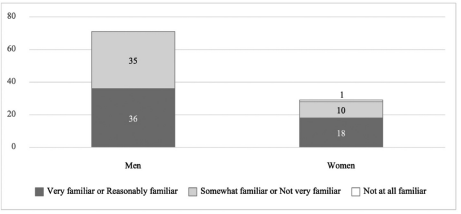

Fig. 5. Participants' familiarity with the AI ethics concept based on education.

We analysed participants' awareness of AI ethics across different factors like work experience, education, gender, and job title. Among 71 participants with up to 2 years of AI development experience, ~48% (34 people) had a good grasp of AI ethics, being either "Very familiar" or "Reasonably familiar" with the concept as shown in Figure 4 (the symbol (~) denotes approximate values). Of the 19 participants with 3–5 years and 10 participants with over 5 years of AI development experience, ~74% (14 people) and 60% (6 participants) had a strong familiarity with the concept of AI ethics, respectively.

Regarding education, among the 66 participants with high school or Bachelor's degrees, ~47% (31 individuals) were either "Very familiar" or "Reasonably familiar" with AI ethics as shown in Figure 5. Among the 22 participants with Master's degrees and the 7 participants with Ph.D. or higher degrees, ~69% (15 people) and ~86% (6 people) had a very good familiarity with the concept of AI ethics, respectively.

Regarding gender, among the 71 male participants, ~51% (36 individuals) demonstrated a strong familiarity with AI ethics, reporting either "Very familiar" or "Reasonably familiar" with the concept as shown in Figure 6. Of the 29 female participants, ~63% (18 participants) were either "Very familiar" or "Reasonably familiar" with the concept of AI ethics.

In terms of job titles, all six AI Experts and three AI/ML Specialists (100%) expressed a high level of familiarity with AI ethics, being either "Very familiar" or "Reasonably familiar" with the concept. Among the 23 AI/Data Scientists, 61% (14 individuals) had a strong familiarity with the concept of AI ethics, as shown in Figure 7.

4.2.2 Awareness of AI Ethical Principles. The ethical guidelines for AI differ depending on the country and organisation. Our previous work (GTLR) revealed that AI practitioners discussed only four specific ethical principles of AI namely "*Accountability*", "*Fairness*", "*Transparency and explainability*", and "*Privacy*". However, obtaining a broader perspective from the industry on all the

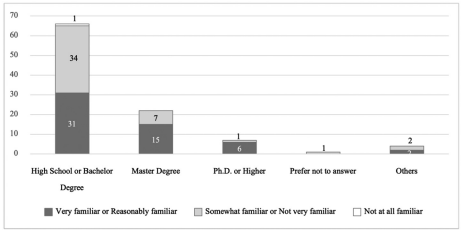

Fig. 6. Participants' familiarity with the AI ethics concept based on gender.

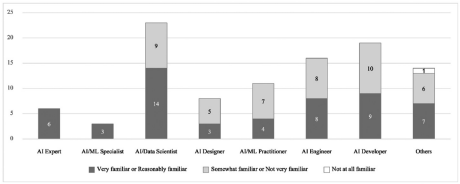

Fig. 7. Participants' familiarity with the AI ethics concept based on job title.

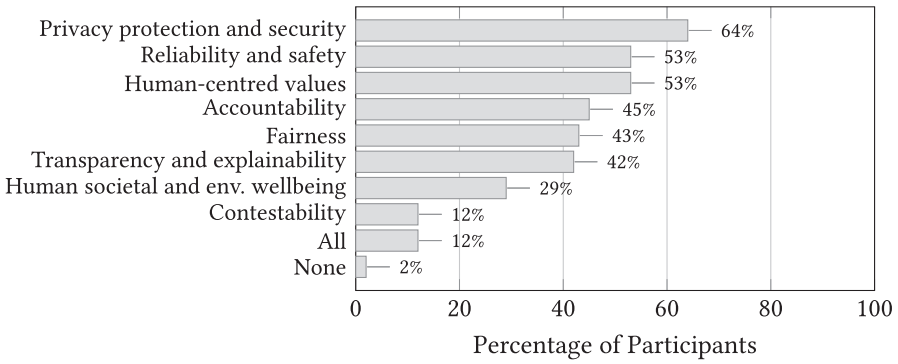

Fig. 8. AI practitioner's awareness of AI ethical principles.

ethical principles related to AI that practitioners were aware of was required. To achieve this, we referred to Australia's AI Ethics Principles[2] as a basis for investigation.

We found that the majority of the participants were aware of the AI ethical principle namely, "*Privacy protection and security*" (64%) followed by two ethical principles, "*Reliability and safety*" and "*Human-centred values*" with 53% each as shown in Figure 8. In addition, most of the participants were also aware of the three AI ethical principles including "*Accountability*" (45%) followed by "*Fairness*" (43%) and "*Transparency and explainability*" (42%). Only a small percentage of the participants (12%) were aware of all AI ethical principles whereas 2% were not familiar with any of the ethical principles of AI.

4.2.3 Reasons for Awareness. Several studies have reported the reasons that contribute to the awareness of AI ethics and ethical principles among AI practitioners such as exposure to news and media, customer complaints [19], personal interests, and experiences [21]. Therefore, we aimed to conduct further research on these reasons and also explore other factors that raise the awareness of AI practitioners regarding AI ethics and ethical principles. To achieve this, we presented a list of possible reasons and asked participants to choose the applicable ones and add any other reasons that were not listed in the options through an open-ended answer.

According to our findings, the reason for the majority of the AI practitioners' awareness of AI ethics and ethical principles was *workplace rules and policies*, accounting for 63% of responses, as shown in Figure 9. The second reason, cited by 54% of participants, was awareness of AI ethics and ethical principles through *news and media*. Likewise, the personal interests of the AI practitioners and their first-hand personal experience were also the reasons for their awareness of AI ethics and ethical principles that accounted for 49% and 48% of responses, respectively. A mere 2% of the participants provided explanations for their reasons for awareness of AI ethics and ethical

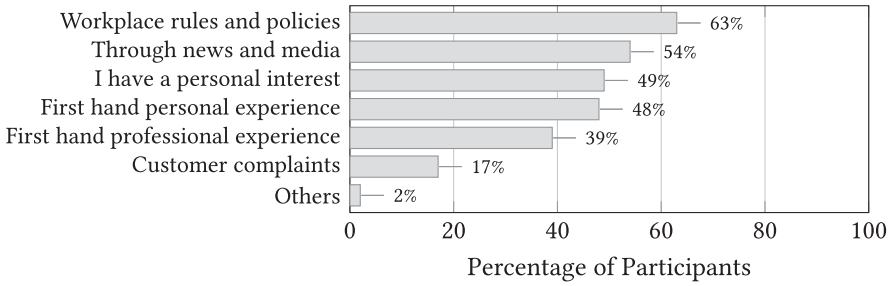

Fig. 9. AI practitioner's reasons for awareness of ethics in AI.

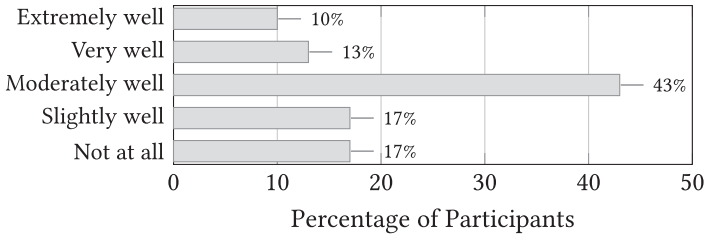

Fig. 10. Role of formal education/training in preparing AI practitioners to incorporate AI ethics.

principles through the open-ended answer option. Interestingly, all participants who provided an answer cited "University" as their reason for awareness. For instance, participant [P38] explicitly stated "University courses" as their source, while [P63] simply mentioned "University".

4.2.4 Role of Formal Education/Training in Preparing AI Practitioners to Incorporate AI Ethics. The significance of educating people about ethics in AI has been discussed in the literature [7] which was one of the recommendations in our previous work (GTLR) [42]. Therefore, our next objective was to gather AI practitioners' perspectives on how effective formal education or training helps them in preparing themselves to incorporate ethics during AI development. To accomplish this, participants were asked to rate the extent to which formal education or training help them, using a five-point scale ranging from "Extremely well" to "Not at all." Figure 10 demonstrates that the majority of participants (43%) felt that formal education or training moderately help them to incorporate ethics in AI. Conversely, 17% of the participants stated that formal education or training provides slight assistance while another 17% reported that it provides no assistance at all. It is worth noting that only a small number of participants (10%) believed that formal education or training is extremely effective in preparing them in incorporating ethics in AI.

4.3 RQ2 – What Challenges/Barriers do AI Practitioners Face in Incorporating Ethics in AI?

First, we began by asking a closed-ended question to the participants to evaluate the ethical principles of AI by assessing the challenges involved in considering and adhering to each one. We used Australia's AI Ethics Principles[2] as a reference and requested the participants to rate the degree of difficulty (ranging from "Very Challenging" to "No experience").

The results indicate that the majority of the participants (27%) found "*Human-centred values*" most challenging to adhere to while developing AI-based systems, followed by another principle "*Privacy protection and security*" (26%). In addition, 24% and 22% of the participants found "*Transparency and explainability*" and "*Reliability and safety*" very challenging principles, respectively,

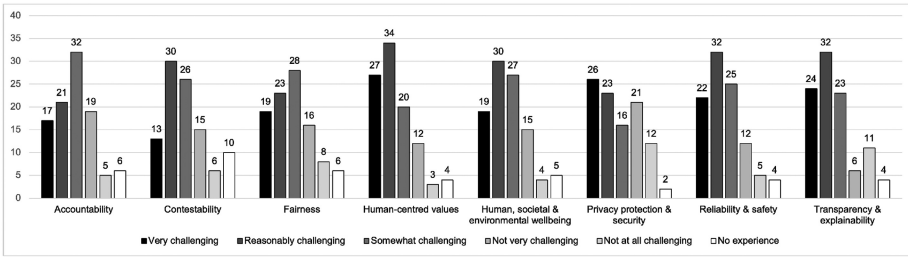

Fig. 11. AI practitioner's perceived degree of challenges in following AI ethical principles: from "Very challenging" to "No experience".

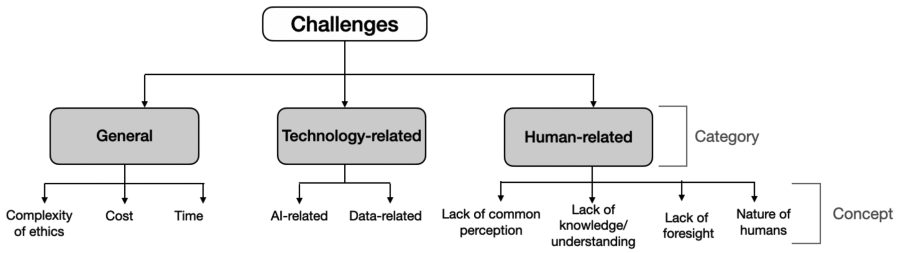

Fig. 12. Overview of the challenges/barriers in incorporating ethics in AI.

to adhere to. Besides, the majority of the participants (12%) also mentioned that "*Privacy protection and security*" is the least challenging ethical principle to adhere to during AI development. Only 3% of the participants reported that "*Human-centred values*" is the ethical principle that is least challenging to adhere to during AI development as shown in Figure 11.

We asked an open-ended question to explore the main challenges or barriers that AI practitioners face in incorporating ethics in AI and we used a bottom-up (inductive) approach for data analysis. In other words, we were not looking for mapping to principles. However, some challenges raised by the participants align with some AI ethical principles. For instance, a participant [P88] noted the challenge of ensuring transparency and accountability due to the complex decision-making process of AI systems. Likewise, the majority of the participants discussed how the biased nature of human beings makes it challenging to maintain fairness in the AI systems being developed [P1, P2, P5, P6, P7, P8, P10, P11, P13, P15, P16, P19, P20, P21, P22, P23, P26, P29, P30, P33, P34, P39, P44, P45, P54, P59, P60, P61, P63, P64, P69, P73, P75, P78, P80, P82, P83, P84, P85, P96, P99, P100].

The qualitative data was analysed using the *STGT method for data analysis*. Based on the response to the open-ended question, we categorised the main challenges of AI practitioners into three categories including (i) General challenges, (ii) Technology-related challenges, and (iii) Human-related challenges. These are explained in detail below. Figure 12 shows the overview of the challenges/barriers in incorporating ethics in AI that we obtained through the analysis of the qualitative data.

4.3.1 General Challenges.

Complexity of ethics. Quantification: The complexity of ethics arises from the multifaceted nature of ethical values and principles and makes it challenging to quantify, such as capturing, defining, and measuring them. Participants also identified this as one of the challenges in integrating ethics while developing AI-based systems. For example, participants [P72] and [P37] said:

💬 "*It is similar to values such as happiness, satisfaction, and quality of life. How to define them mathematically? In such a situation, the creators of the algorithm will probably reach for an economic, tangible, and imaginative argument - how much can be saved thanks to it. In addition, there are many exceptions to the rule that cannot be captured in the algorithm.*" – [P72]
💬 "*Measuring and building ethics red lines is challenging.*" – [P37]
Translation: Translating ethics from principles to practice can be a challenging task. While ethical principles provide a foundation for ethical decision-making, applying these principles in practice can be complex. It was considered a challenge to incorporate ethical principles into AI-based systems for most of the participants (P2, P4, P5, P6, P7, P8, P10, P11, P12, P14, P15, P16, P17, P18, P19, P21, P23, P24, P35). For example,
💬 "*The main challenge is to implement some huge concept onto program written in the programming language.*" – [P89]
Management: Various ethical principles for AI exist and differ among countries and organisations. Ensuring that these principles are balanced, connected, and monitored during the development of AI-based systems can present a challenge for AI practitioners. Some participants expressed similar views on the management of AI ethical principles and discussed the challenges they faced during AI development.
💬 "*One has to decide how to connect it with the other principles and rules we base our AI upon in order to maximise our target outcomes. Lastly, since these things are very difficult to control and audit, there is no real incentive to introduce them into the system.* " – [P30]
💬 "*A balance on all of them seems difficult as is always on the software you can't have them all.*" – [P32]
Implementation: Following each ethical principle during the development of an AI-based system could be a challenge as discussed by some of the participants (P32, P66, P71, P83).
💬 "*It's difficult to consider all branches of the main principles that should be applied to AI. This results in some gaps and scenarios being missed and only being realised once it is developed.*" – [P63]

Cost. Cost is an important factor in developing an AI-based system and it could be one of the challenges in developing ethical AI-based systems. The absence of adequate funds could be a challenge because developing such systems requires specialised skills and resources, which come at a cost and it may also impact business goals. Most of the participants (P1, P3, P5, P9, P10, P14, P15, P16, P20, P21, P22, P26, P90) reported that cost is a challenge in developing ethical AI-based systems.
💬 "*Developing ethical AI requires a specific set of skills and resources, which may be in short supply and requires a cost*" – [P58]

Time. To apply principles: Time is another factor that may impact the development of ethical AI-based systems. Most of the participants noted that lack of time is a challenge they face in considering and following ethics during the development of AI-based systems (P5, P7, P10, P11, P14, P16, P17, P18, P21, P25, P26, P29, P90).
💬 "*Some ethical principles directly impact the development timeline and require additional work which isn't required for the software to be operational, extending the development time-frame.*" – [P70]
To understand principles: It is essential to understand AI ethical principles before incorporating them into the development of AI-based systems. Some of the participants (P5, P6, P8, P9, P10, P11, P17, P18) reported that lack of time to understand the AI ethical principles is a challenge that impacts the development of ethical AI-based systems.

4.3.2 Technology-Related Challenges. Participants discussed two challenges (concepts) related to technology in incorporating ethics during AI-based system development including (i) AI-related and (ii) Data-related challenges. Each of these concepts is underpinned by multiple codes.

AI-related. Complexity: AI is a complex technology. AI is designed to learn and adapt to new situations, making it more intelligent over time. However, this also means that AI systems can become incredibly complex, making them difficult to understand and control. The complexity of AI technology could be one of the challenges for AI practitioners to follow ethics during its development. This idea was supported by some of the participants who stated:

💬 "*Systems based on artificial intelligence pursue a strictly defined goal. Make it an increase in usability. At the outset, it is necessary to describe in a mathematical way what utility is. It's hard to do. It is similar to values, such as happiness, satisfaction, and quality of life. How to define them mathematically?*" – [P72]

💬 "*The reasoning of the system cannot really be accounted for, it acts like a black box, where why it came to a certain conclusion is difficult to discern.*" – [P59]

A few other participants like [P42], [P47], and [P71] also reported the same notion.

Scope: AI has the potential to tackle a wide range of complex tasks, from recognising speech and images to analysing large datasets and making predictions based on patterns. The scope of AI has been expanding rapidly in recent years, driven by advances in ML, deep learning, and natural language processing which could be a challenge for AI practitioners to develop ethical AI-based systems. A participant [P36] shared a similar thought on AI scope.

💬 "*AI in itself is not very hard to understand but then the problem lies in the user's knowledge of what the data collection and training results they are experiencing turns into, software misuse also happens but with AI the scope is wider and harder to consider.*" – [P36]

Data-related. Lack of quality data: Quality training data is critical to the success of AI-based systems. In order for AI models to accurately learn and make predictions, they require large amounts of high-quality data. Lack of quality training data could be one of the challenges in developing ethical AI-based systems. For example, if the training data is gender or race-biased, then the AI model created will also be biased. Several participants (P58, P59, P60, P74, P76, P88) reported it as a challenge in developing ethical AI-based systems.

💬 "*One of the main challenges in developing ethical AI is ensuring that the data used to train the models is unbiased. If the data is biased, the model will be too, potentially leading to unfair or discriminatory outcomes.*" – [P58]

💬 "*AI systems can perpetuate and amplify biases and discrimination present in the data used to train them, making it challenging to ensure that the AI is treating all individuals fairly.*" – [P88]

Lack of training data: There should be sufficient training data available to create AI models. Lack of training data could be a challenge in creating ethical AI-based systems. This idea was supported by some of the participants (P65, P87).

💬 "*In my experience, this would be the lack of data, in turn, resulting in not being fully ethical*" – [P65]

4.3.3 Human-Related Challenges. Participants discussed four human-related challenges (concepts) in considering and following ethical principles during AI-based system development including (i) Lack of common perception, (ii) Lack of knowledge/understanding, (iii) Lack of foresight, and (iv) Nature of humans. Each of these concepts is underpinned by multiple codes.

Lack of common perception. Lack of common perception of ethics: The concept of "ethics" is subjective, with each person having their own unique interpretation, which can differ from one another. This difference in human perception and varying opinions on what constitutes ethics creates difficulties in creating a universal definition of the term "ethics" that hampers the incorporation of ethics in AI. Some of the participants [P27], [P93], [P71], and [P100] supported this notion.

💬 "*As I have been in conversations where people actually don't agree on ethics in the most basic way.*" – [P71]
💬 "*Difference in thinking about ethics.*" – [P93]
💬 "*A major limitation is to do with the problem that implementing ethics in AI is highly variable. Meaning that "ethics" are dependent on the individual/socially accepted practices and behaviors surrounding humans, not some predefined set of a commonly agreed set of rules or behaviors.*" – [P100]

Likewise, participants also mentioned that a lack of common perception on selecting ethical principles during AI development is a challenge. For example, P30 said:
💬 "*The major challenge is to agree upon what is an ethical principle that should be taken into account.*" – [P30]
Other participants ([P88] and [P63]) also shared similar thoughts on the challenge of selecting ethical principles during AI development.

Lack of common perception on how to follow ethics and incorporate them during AI development is also a challenge to AI practitioners. A participant [P34] mentioned:
💬 "*There isn't any standard procedure to follow and everyone has different ideas.*" – [P34]

Cause of lack of common perception of ethics: Individuals come from various backgrounds such as different societies, cultures, ethnicities, and groups. These varied backgrounds and cultural differences can influence one's perception of ethics and shape their values which makes it challenging to define the term "ethics". This notion was supported by several participants (P27, P37, P46, P72, P75, P79, P80, P97). For example, a participant [P97] stated:
💬 "*Ethics can vary based on different value systems in humans/cultures, therefore I think a big limitation with implementing ethics in AI would be taking a unified approach and applying it, especially with some ethics in particular, and also the potential for misguided ethics.*" – [P97]

Similarly, [P27] and [P37] discussed a similar idea and stated:
💬 "*Different societal groups have completely different perceptions in terms of ethical issues, therefore, creating a model which will objectively define these things for the AI will be complex.*" – [P27]
💬 "*Human values are relative.*" – [P37]
💬 "*I think human-centered values limit us a lot from reaching our potential. We must always put people forward, but this is the right way to do AI and fairness is also challenging as we deal with different people with different values.*" – [P79]

Consequence of lack of common perception of ethics: The definition of the term "ethics" of AI varies between countries and companies. This is caused by the difference in human perception of *ethics*. This variance in definitions can cause confusion for AI practitioners, who may not know which definition to adhere to when developing AI-based systems. The absence of a universal definition for the term "ethics" poses a challenge in applying it to AI-based systems. This notion has been supported by several participants (P27, P42, P45, P49, P58, P60, P88) who stated:
💬 "*The human limitations involve the general definition and perception as to what ethics actually are.*" – [P27]
💬 "*Artificial intelligence is a very complex issue, and the lack of a clear definition of ethics makes it difficult to apply in this area.*" –[P88]
💬 "*Coming up with universal guidelines, the ethics depend on the company or the organisation designing and developing the solutions and if they manipulate the idea to be in their favor.*" – [P45]

There is not only a lack of a common definition for the term "ethics" but there is also a variation in the definitions of ethical principles of AI. The list of ethical principles and their definitions can differ between companies and countries. Consequently, AI practitioners find it challenging to determine which ethical principles to adhere to while developing AI-based systems. A participant [P58] supported that lack of consensus on AI ethical principles is one of the human limitations in ethics incorporation in AI and stated:

💬 "*There is often a lack of consensus on ethical principles, which can make it difficult to develop and implement effective guidelines for AI.*" – [P58]

Lack of Knowledge/Understanding. Lack of knowledge/understanding of AI: Presently, AI technology is advancing at an accelerated pace. However, its complexity can make it challenging for humans to comprehend. Staying up-to-date with the latest AI advancements and updates can also be difficult for them which could be a limitation in incorporating ethics in AI. Some of the participants (P35, P40, P41, P58, P60, P62, P70, P77, P81, P82) noted the same. For example, [P35] and [P82] stated:

💬 "*Lack of understanding of the technology.*" – [P35]

💬 "*There are several barriers or constraints to implementing AI, but the main ones for me are the shortage of knowledge and skills and availability of technical personnel with the experience and training necessary to effectively implement and operate ethical AI solutions.*" – [P82]

Likewise, [P41] also mentioned that humans lack understanding of creating ethical AI systems as AI is powerful and rapidly evolving.

💬 "*The field of AI ethics is relatively new and rapidly evolving, and there is still much that is not understood about how to create ethical AI systems.*" – [P41]

A similar thought was shared by [P58] who said that a lack of full understanding of the ethical implications of AI leads to difficulty in developing and incorporating ethical guidelines. [P58] reported:

💬 "*Many people may not fully understand the ethical implications of AI, making it difficult to develop ethical guidelines for its development and use.*" – [P58]

A similar idea was discussed by [P70] where the participant said that AI is too advanced and practitioners' lack of knowledge of advanced programming is a limitation to incorporating ethics in AI. The participant stated:

💬 "*The limitations stem from the advancement in AI to the point where it becomes too advanced for simple programming.*" – [P70]

Cause of lack of knowledge/understanding of AI: AI is a complex and multifaceted technology. Its primary characteristics are self-learning and adaptiveness, which enable it to constantly improve its performance and decision-making abilities. However, its outcomes can be difficult for humans to comprehend, and it can sometimes be seen as a "black box". The complexity of AI has been a topic of discussion among AI experts, as it is one of the causes of humans' limited understanding of AI.

💬 "*AI decisions are not always intelligible to humans.*" – [P82]

Many participants mentioned the complexity of AI as one of the main causes of human's lack of understanding of AI (P41, P45, P70, P81).

Consequence of lack of knowledge/understanding of AI: AI has gained immense power and can sometimes surpass human capabilities, despite being created by humans. However, due to its intricate nature and humans' limited understanding of AI, there is a potential for ethical vulnerabilities to be exploited by AI. P81 stated:

💬 "*We can never tell what loopholes are there in our currently existing list of ethics. AI is super intelligent, and there may come a point where it finds a loophole and destroys us. Our limitation is that we do not know everything.*" – [P81]

Lack of knowledge of other's work: Nowadays, AI is extensively utilised in various fields like healthcare, transportation, finance, information technology, education, and many others. With the growing prevalence of AI, there is also an upsurge in the creation of AI-based systems. However, since AI practitioners are scattered across the world, they may not always be aware of each other's

progress and actions, which can hinder the implementation of ethical practices in AI due to human limitations. A participant [P48] shared similar idea:
💬 "*What limits us, is everyone working on AI independently, not knowing what the next company is doing* " – [P48]

Lack of foresight. AI outcomes: AI practitioners who developed AI-based systems are unable to predict the outcomes of the system due to the complex nature of AI. The complexity of AI not only causes a lack of understanding of AI in AI practitioners but also makes it challenging to predict AI outcomes for them. Some participants (P36, P52, P59, P88) supported this idea:
💬 "*Our limits come from our understanding of AI itself, where we do not know all the parameters that it gives out. At the end of the day, we're left to guess the end results of what our AI does and how it will function.*" – [P52]
💬 "*To consider AI's possibilities outside the scope of human possibilities or the things easily achievable as a human, it is necessary to fully consider the outcome of AI use in any field and this could become a hard task to manage with how AI is progressing.*" – [P36]
💬 "*Ensuring that AI systems are transparent and accountable for their actions can be difficult, as it can be challenging to understand how an AI system arrived at a particular decision.*" – [P88]

AI consequences: Like any other software, AI-based software may have both positive and negative consequences. One of the human limitations is the inability to predict the consequences of AI-based software. Many participants mentioned that lack of foresight of AI consequences is one of the human-related challenges in incorporating ethics in AI (P1, P2, P5, P6, P7, P9, P10, P11, P14, P15, P16, P18, P23, P41, P58, P60, P69, P81, P88).
💬 "*It can be difficult to predict the long-term consequences of AI, making it challenging to anticipate and mitigate potential negative impacts.*" – [P58]
💬 "*Difficulty in predicting future consequences*" -[P60]
💬 "*It can be difficult for developers to anticipate all of the potential consequences of an AI system, particularly if the system is highly complex or if it is being used in a new or unexpected way.*" – [P41]
💬 "*Again, not knowing everything is one day going to be a problem because these machines are destined to surpass us. The main challenge is not being able to see many years into the future and what steps we can take today to prevent a catastrophe.*" – [P81]

Nature of humans. Biased: Humanity is divided into various categories including culture, ethnicity, country of origin, and religion. This division creates diversity in the values people adopt and how they perceive things, resulting in biases. This is an inherent characteristic of human nature and represents a significant challenge in the development of ethical AI-based systems. The majority of the participants reported it as a major human-related challenge in considering and following ethical principles of AI (P1, P2, P5, P6, P7, P8, P10, P11, P13, P15, P16, P19, P20, P21, P22, P23, P26, P29, P30, P33, P34, P39, P44, P45, P54, P59, P60, P61, P63, P64, P69, P73, P75, P78, P80, P82, P83, P84, P85, P96, P99, P100).
💬 "*It's impossible to have a no-bias perspective and be conscious of all the moral implications of our work, even if we're working on a small or mid-size team.*" – [P29]
💬 "*Humans are inherently biased, and in implementing machine learning systems, some of these biases are found.*" – [P59]
💬 "*Humans are subject to their own bias which tends to seep into the logic used to build the AI.*" – [P63]
💬 "*The limitations concern the innate human view of things. A human being, no matter how hard he tries, can never be totally immune to the bias or to the psychological and cultural structures that formed him.*" – [P44]

💬 "*The main barriers in my opinion would be bias or discrimination, and potentially the philosophical challenge with regard to humans' involvement in developing AI systems.*" – [P80]

Awareness of bias transfer: Humans are inherently biased, and there is a significant likelihood that their biases may be transferred to AI-based systems during their development. Even those with good intentions may inadvertently transfer their biases, while those with bad intentions may do so intentionally. Consequently, humans may or may not be conscious of the extent to which their biases have been transferred during the development of AI-based systems. Many participants reported it as a challenge in developing AI-based systems (P38, P44, P63, P90, P94, P96, P99).
💬 "*I think the humans (behind the AI) implicit biases may be transferred to the AI be it intentionally or unintentionally.*" – [P38]
💬 "*The human bias is unknowingly being transferred in the code.*" – [P90]
💬 "*That our own ethical faults get transferred into the ethics system*" – [P94]

Ethics vs. profit: Diverse preferences and priorities are inherent traits of human beings. For instance, certain individuals may prioritise financial gain over ethical considerations, while others may prioritise ethics over profit. This human nature is one of the challenges in developing ethical AI-based systems and similar thought was shared by some of the participants:
💬 "*Humans, for example, insurance companies, care about profit. Including ethics in an AI, for example, in regard to diversity and inclusion, can be counter-intuitive or counterproductive. If statistics and the AI itself determine that people from a given ethnicity are prone to a given outcome that is adverse to business, it is counter-intuitive for us programmers to not take it into consideration if it affects what we are trying to maximise. Nature is what it is, numbers are what they are, and these things are very difficult to control and maybe, they shouldn't. It is like asking someone to imagine that gravity is non-existent and jump off a roof.*" – [P30]
💬 "*The danger lies in humans implementing AI for their own gain, and for AI practitioners to ignore ethics for profit, greed, selfishness, or any other negative reason.*" – [P100]

4.4 Summary of Key Findings

The **key findings** (**KFs**) from our survey have been summarised in Table 3. Our previous work (GTLR) [42] revealed the need for an empirical study that solely focuses on investigating AI practitioners' views and understanding of ethics in AI. We, therefore, focused our study on investigating aspects related to AI practitioners' *awareness* of AI ethics, and their *challenges* in incorporating ethics in AI-based systems.

5 DISCUSSION

We now discuss our findings and insights in light of related works.

5.1 AI Practitioners' Awareness of Ethics in AI

5.1.1 Extent of Awareness of "AI Ethics" Concept. Vakkuri et al. [60] reported that there is a lack of awareness of AI ethics among AI practitioners. Most studies have focused on understanding if AI practitioners are aware of the concept of AI ethics, and focused on developing tools and methods to raise awareness among AI practitioners [37]. An empirical study by McNamara et al. [34] reported that the ethical guidelines provided by the **Association for Computing Machinery** (**ACM**) had minimal influence on software developers, who continued to work in the same way as before and concluded that software practitioners were not well-informed on ethics. Based on that, the **Ethically Aligned Design** (**EAD**) guidelines version acknowledged that this could also be true for AI ethics [60] but there is no research to investigate how familiar AI practitioners are with the concept of AI ethics. This along with the fact that being aware of AI ethics is insufficient; a

Table 3. Key Findings (KF) of the Study

	Key Findings (KF)	Section
KF1	Majority of the AI practitioners (41%) were *reasonably* familiar with the concept of ethics in AI.	4.2.1
KF2	Few AI practitioners (13%) were very familiar with the concept of *ethics* in AI.	4.2.1
KF3	Majority of the AI practitioners were aware of some AI ethical principles including "*Privacy protection and security*" (64%), "*Reliability and safety*" (53%), and "*Human-centred values*" (53%).	4.2.2
KF4	Very few AI practitioners (12%) were aware of *all* the ethical principles of AI.	4.2.2
KF5	The reason for the majority of AI practitioners' (63%) awareness of AI ethics and ethical principles was *workplace rules and policies*.	4.2.3
KF6	Very few AI practitioners (2%) reported *university* as the reason for their awareness of AI ethics and ethical principles.	4.2.3
KF7	Majority of AI practitioners (43%) believed that formal education or training *moderately* help in preparing them to adhere to AI ethical principles during AI development.	4.2.4
KF8	Majority of the participants (27%) reported "*Human-centred values*" as the most challenging ethical principle to adhere to during AI development.	4.3
KF9	AI practitioners encountered *General challenges*, *Technology-related challenges*, and *Human-related challenges* in incorporating ethics in AI.	4.3
KF10	Majority (42%) of the AI practitioners reported *biased nature of humans* as a key human-related challenge in incorporating ethics in AI.	4.3.3

thorough understanding of the concept is crucial for AI practitioners to ensure that AI development is conducted in a responsible and ethical manner motivated us to explore this topic. As a result, we carried out a survey involving 100 AI practitioners, revealing that most participants (41%) possess a *reasonable* level of familiarity with the concept of AI ethics. This suggests that there is still a deficiency in the efforts required to enhance awareness of "AI ethics" among AI practitioners.

Similarly, our data revealed that participants with over 2 years of work experience in AI development had a greater awareness of AI ethics compared to those with 2 or fewer years of experience as reported in Section 4.2.1. One possible explanation for our result could be that as the level of experience increases, the familiarity with AI ethics might also increase.

Furthermore, our results indicated that participants with a Ph.D. or higher degree exhibited the highest level of familiarity with AI ethics, followed by those with a Master's degree, while individuals with a high school or Bachelor's degree had the lowest level of familiarity as reported in Section 4.2.1. A plausible explanation for our findings may be that there might be a tendency for familiarity with the concept of AI ethics to rise with higher levels of education.

Our results also indicated that female participants had a greater awareness of the concept of AI ethics compared to male participants as reported in Section 4.2.1. A potential justification for our results could be that females might be more interested in learning about AI ethics than males.

Likewise, our data revealed that job titles like AI Experts, AI Specialists, and AI/Data Scientists had higher familiarity with the concept of AI ethics as compared to other job titles as reported in Section 4.2.1. One possible explanation of our result could be that AI Experts, AI Specialists, and AI/Data Scientists might be expected to be familiar with AI ethics as part of their job duties.

5.1.2 Awareness of AI Ethical Principles. Various companies, such as Microsoft [35], Google [14], and IBM [22], have their own ethical guidelines on AI development, outlining the ethical principles that AI-based systems should be developed based on, such as transparency, fairness, privacy, and so on. These guidelines serve to steer AI practitioners toward ethical AI development and ensure that the systems they develop align with all these principles. This implies that AI practitioners must be aware of and possess adequate knowledge of these ethical principles of AI before developing them. However, research shows that AI practitioners are aware of only specific ethical principles of AI such as *accountability/responsibility*, *privacy*, *fairness*, and *transparency and explainability*. For example, Vakkuri et al. [62] and Mark and Anya [32] concluded that AI practitioners in their respective studies were aware of the "transparency" ethical principle of AI. AI developers were aware of the ethical principle of "fairness" in AI and strived to eliminate any issues related to it [19]. According to Veale et al. [63], participants understood the importance of accountability in AI systems and took responsibility for any harm caused by their creations. According to Rothenberger et al. [47], the principle of "responsibility" was deemed highly relevant and influential among the other ethical principles in AI. "*Privacy*" was another ethical principle that AI practitioners were aware of and extensively discussed, with data and information privacy being a significant concern for organisations [21, 48].

Christodoulou and Iordanou [8] reported that participants discussed the challenges related to *transparency*, *privacy*, *fairness*, and *accountability* only when they were asked about the challenges in addressing ethical issues in AI. It indicates that AI practitioners were either not aware of other ethical principles of AI or they didn't face any challenges related to them. We reviewed empirical studies that focused on understanding AI practitioners' views on AI ethics in our previous work (GTLR) [42]. We found that AI practitioners discussed only four ethical principles of AI including *transparency and explainability*, *privacy*, *fairness*, and *accountability/responsibility*. The study conducted by Christodoulou and Iordanou [8] and our previous work (GTLR) [42] both indicate that AI practitioners primarily discussed a limited set of AI ethical principles. These include *privacy protection and security*, *accountability/responsibility*, *transparency and explainability*, and *fairness*. However, it is unclear whether the research context and questions asked were responsible for the AI practitioners only discussing these four ethical principles. Our study asked specifically about a set of AI ethical principles and we report those findings (KF3 and KF4), however, this is limited in applicability by our limited survey size of 100 participants.

We used Australia's AI Ethics Principles[2] and our results show that "*Privacy protection and security*" is the ethical principle that most AI practitioners (64%) were aware of. This suggests that, regardless of the specific research context, the majority of AI practitioners possess knowledge about the ethical principle of AI concerning "*Privacy protection and security*". We also found that only 12% of the participants were aware of *all* ethical principles of AI.

5.1.3 Reasons for Awareness. *Organisational pressure* [63], *laws and regulations* [62], *personal interest and experience* [21], *customer complaints* and *negative media coverage* [19] were some of the reasons for AI practitioners' awareness of AI ethics reported in the literature. We consolidated the reasons cited in previous studies and asked survey participants about the reason for their awareness and we included an open-text option at the end to allow participants to share any reasons that were not included in the provided list. *Workplace rules and policies* were cited by the majority of the participants (63%) as the reason for being aware of AI ethics which was discussed in one of the studies [62]. Likewise, most of the participants reported *personal interest and experience*, *news and media*, and *customer complaints* as their reasons for awareness of AI ethics.

On the other hand, previous studies did not mention *first-hand professional experience* as a reason for AI practitioners' awareness of AI ethics, however, our study revealed that it is indeed a

significant factor (39%). In addition, according to our findings, a small number of participants (2%) identified *university* as a reason for their awareness of AI ethics, which was not previously reported in the literature. However, it is not surprising that only 2% of the participants reported this since AI ethics is a fairly new addition to the university curriculum (where it is part of the curriculum) and, in our study, a big part of the participants were older than 30 years so they not have been introduced to this in their university education.

5.1.4 Role of Formal Education/Training in Preparing AI Practitioners to Incorporate AI Ethics. It has been highlighted in the literature that the topic of "AI ethics" must be incorporated into the curriculum to make students aware of the concept of AI ethics [2, 3, 5]. Although the importance of formal education/training has been highlighted in the literature, there is a lack of research that shows to what extent formal education or training assists AI practitioners to incorporate ethics in AI. Our survey discovered that the majority of participants (43%) believe that formal education/training *moderately* aids in preparing them to incorporate ethics in AI. From our results, it can be inferred that formal education or training plays a role in preparing AI practitioners to incorporate ethics in AI, but their significance may not be paramount.

5.2 AI Practitioners' Challenges in Incorporating Ethics in AI

Studies reported the challenges of AI practitioners in adhering to specific ethical principles of AI during AI development such as transparency [21, 49], accountability [53], and fairness [19]. Through our study, we investigated AI practitioners' degree of challenges in considering and following all ethical principles of AI using Australia's AI Ethics Principles.[2] Our results indicated that the majority of the participants (27%) find "*human-centered values*" the most challenging ethical principle to adhere to. This finding contradicts the findings of other studies as those studies did not report about "*human-centered values*". In fact, none of the empirical studies discussing the challenges related to specific AI ethical principles mentioned "*human-centered values*" as the most challenging ethical principle to adhere to, rather they discussed challenges related to other ethical principles such as *transparency*, *accountability*, and *fairness*. However, this discrepancy may be due to the differences in ethical principles across countries or organisations, as we used Australia's AI Ethics Principles[2] in our survey but recruited participants from around the world as discussed in Section 6.

Likewise, AI practitioners encountered challenges in *conceptualising ethics* [58], dealing with various tensions and tradeoffs between AI ethical principles when AI practitioners had to incorporate specific ethical principles of AI [49], addressing issues such as *highly general principles, vague principles, lack of technical understanding* that impacted the development of ethical AI-based systems [27] and translating AI ethical principles into practice [21, 36]. According to Mittelstadt [36], one challenge to the principled approach to AI ethics is the abstract and vague nature of the principles themselves. Similarly, Whittlestone et al. [64] found that terms related to the ethical and societal aspects of algorithms, data, and AI lack consistent meanings across different contexts. This inconsistency creates ambiguity and makes it difficult to clearly define various principles in this domain. In addition, the study concluded that the involvement of various stakeholders with varying interests leads to vague principles, such as fairness or respecting human dignity in AI, which lack the specificity needed for practical guidance. Hagendorff [17] reported that the current AI ethics efforts have mainly given us broad principles that sound good in theory but are vague and lack concrete guidance. They fall short in tackling the deeper ethical and political challenges inherent in concepts like *fairness* and *privacy*. Some participants in our study highlighted the challenge of developing ethical AI-based systems due to the complex nature of ethics. They discussed the difficulty of quantifying ethics during development, attributing it to the multifaceted nature of

ethical values and principles. Interestingly, they did not specifically mention the abstract or vague nature of ethical principles as a significant challenge.

Likewise, Mittelstadt [36] reported that there is a lack of empirically proven methods to translate principles into practice in real-world development contexts. Most of the participants in our study also noted the challenge of translating ethical principles into practice when developing AI-based systems, but we lack insight into whether this challenge directly stems from the lack of such methods. Future work can explore this challenge and its reasons in depth.

These findings align with our study, where participants faced similar difficulties in defining and conceptualising ethics due to its subjective and complex nature, incorporating all AI ethical principles into development because there are numerous ethical principles that need to be considered, translating AI ethical principles to practice (see Section 4.3.1) and lacking knowledge of AI systems while developing ethical AI-based systems (see Section 4.3.3).

Although some of the findings of our study are similar to the previous studies, there are some findings that differ from the previous studies. Specifically, we obtained various challenges that are related to humans (AI practitioners) which impact the development of ethical AI-based systems. For example, *lack of common perception, lack of knowledge and understanding of various aspects (like AI, other's work), lack of foresight, and nature of humans* are the challenges that we explored through our study. It is important to understand AI practitioners' limitations because these practitioners are responsible for designing and developing AI systems that have a significant impact on society [41]. Orr and Davis [41] proposed that there is a need for further research to investigate the limitations that AI practitioners possess when it comes to incorporating ethics in AI. Since their study was based in Australia, the authors recommended exploring this issue among a broader range of AI practitioners. Therefore, due to the research gap, we were motivated to conduct a survey to determine the overall challenges that AI practitioners face when it comes to developing ethical AI-based systems and "*human-related challenges*" are the new findings that we obtained through our study.

5.3 Insights

From the analysis of the open-ended responses and memos taken while employing the *STGT for data analysis* approach and literature, we have discovered a number of noteworthy **insights**. These primary findings, along with our observations, may be used as recommendations for future research.

5.3.1 Biased Nature of Humans (based on participants' responses and supported by literature): Our findings show that the majority of the participants (42%) mentioned the *biased nature of humans* as a key human-related challenge in incorporating ethics in AI. Research has also been conducted to explore the role of human biases and their impact on the development of software systems. According to a study by Dominguez-Catena et al. [11], machine bias can also be caused by the programmer's biases derived from their cultural background and the programming environment's context. Soleimani et al. [51] reported that one of the reasons for a biased AI system is HR managers' and AI developers' biased assumptions of different aspects that lead to biased decisions. Likewise, Cowgill et al. [9] also concluded that if the programmers developing ML algorithms are highly non-representative, they may exhibit biases that are passed onto the algorithms they create.

Hiring diverse people in the AI teams has been shown to help minimise these biases. For instance, Zowghi and da Rimini [66] mentioned that making sure teams are diverse, fair, and inclusive is essential for reducing risks and getting different perspectives from AI designers and engineers. Similarly, various guidelines for developing ethical AI systems have emphasised the importance of not just involving AI experts but also engaging a diverse range of people throughout the AI

development process to ensure it's ethical [54]. Enhancing the diversity of AI development teams is a step to mitigate biases from AI systems and achieve fair and equitable AI development [10]. Creating a diverse development team from the outset is also recommended as a means to help mitigate algorithmic bias by smoothing out prediction errors across subgroups of developers [39].

5.3.2 Lack of Consensus on Definitions (based on Literature). The term "*ethics*" has been defined by different people. For example, "*ethics*" is defined as "*the moral principles that govern the behaviors or activities of a person or a group of people*" [38]. Likewise, Iacovino [20] defined *ethics* as "*the way an individual behaves and the values they hold.*", whereas Payne and Joyner [43] defined *ethics* as "*a system of value principles or practices and the ability to determine right from wrong.*" There are various definitions available for AI ethics and ethical principles. Vakkuri and Abrahamsson [55] concluded that despite ongoing academic discourse on the connection between AI and ethics for many years, there is still no widely accepted definition or consensus on what AI ethics entails or how it should be labeled. Despite numerous papers and diverse keywords from various fields regarding AI ethics, it remains a difficult task to define the field accurately. For example, Siau and Wang [50] defined AI ethics as "*the principles of developing AI to interact with other AIs and humans ethically and function ethically in society*", whereas [44] defined it as "*the ability of a machine to behave morally, without invoking its moral motivations.*" We used the definition of AI ethics provided by Siau and Wang [50] and the definition of ethics provided by Nalini [38] in our study as both of these definitions highlight the importance of ethical considerations in shaping behavior, whether in the realm of AI development or in human actions.

There is also a lack of consensus on the ethical principles of AI. For example, the definition of ethical principles varies in different parts of the world. One of the ethical principles included in Australia's AI Ethics Principles[2] and European Commission's Ethics Guidelines is related to "Diversity, non-discrimination and fairness". Australia's AI Ethics Principles[2] defined "Fairness" in AI as *"AI systems should be inclusive and accessible and should not involve or result in unfair discrimination against individuals, communities, or groups.*" On the other hand, the European Commission's Ethics Guidelines [16] defined "Diversity, non-discrimination, and fairness" as "*AI systems should consider the whole range of human abilities, skills, and requirements, and ensure accessibility and should focus on (i) avoidance of unfair bias, (ii) accessibility and universal design and (iii) stakeholder participation*".

According to our survey findings, the absence of a shared understanding of the term "ethics" and the lack of agreement on the definitions of "AI ethical principles" make it challenging to develop ethical AI-based systems. According to the participants, the main cause of this challenge is the varying perceptions of human beings on those terms. Other causes of the lack of consensus on definitions of "ethics" and "ethical principles" and their impact on the development of AI-based systems should be explored in more depth so that mitigation measures may be devised.

5.4 Recommendations

Based on the findings from our survey, we offer some recommendations for the AI industry, the AI research community, and AI educators. Figure 13 summarises the KFs of the study.

Recommendations for Practice

— *Emphasising workplace rules and policies*: Most participants (63%) perceived that *workplace rules and policies* were the reasons for their awareness of AI ethics and ethical principles as discussed in Section 4.2.3. Therefore, we suggest that AI companies create policies that encompass all aspects of AI development, including ethical considerations such as fairness, accountability, transparency, and so on, to enhance the awareness of AI ethics among AI

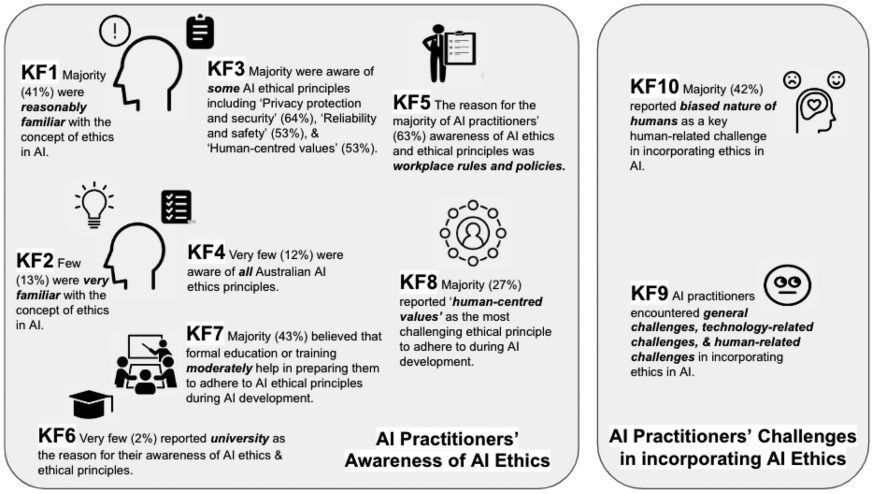

Fig. 13. A summary of the KFs of the study.

practitioners. AI companies should also make it mandatory for AI practitioners to comply with all company policies to ensure adherence to ethical principles during AI development.

— *Practising diversity and inclusion in the hiring processes*: Our survey findings validate that the primary human-related challenge encountered by AI practitioners during ethics incorporation in AI is the "biased nature of humans" (Section 4.3.3). Since the majority of AI practitioners reported it as a challenge, diversity, and inclusion in the hiring panels or decision-making processes could be a potential solution to mitigate the biased nature of humans to ensure that their teams are representative of the diverse communities that they serve. This may help to reduce the likelihood of biased perspectives being integrated into AI systems.

🎓 Recommendations for Research- Future Work

— *Investigating solutions for the challenges of integrating ethics in AI*: A set of challenges that AI practitioners encounter when incorporating ethical principles into AI systems were identified which include *general challenges*, *technology-related challenges*, and *human-related challenges*, as discussed in Section 4.3. AI researchers can concentrate on exploring solutions to overcome these challenges faced by AI practitioners, which will improve the integration of ethics into AI for AI practitioners. In the same vein, we recommend that AI practitioners analyse the challenges they encounter when integrating ethical principles into AI, based on the challenges identified in our study. Doing so will help them gain a better understanding of their strengths and weaknesses when incorporating ethics into AI.

— *Investigate the "human-centered values" principle in-depth*: The focus of our survey study was to find out the AI practitioner's degree of challenges in considering and following specific AI ethical principles during AI development (see Section 4.3). Our survey confirmed that *human-centered values* is the ethical principle that is most challenging to consider and follow during AI development among the ethical principles listed in *Australia's AI Ethics Principles*. Future research should focus on investigating the challenges associated with the adherence of *human-centered values* during AI development and design solutions to overcome those challenges. Research could also be conducted to explore the potential tradeoffs between different ethical principles and the ways in which human values can be effectively integrated into AI systems.

— *Better understanding the human-related challenges that AI practitioners face when incorporating ethics in AI*: We identified a set of human-related challenges that AI practitioners face when considering and following ethical principles in AI that include *lack of common perception*, *lack of knowledge/ understanding*, *lack of foresight*, and *nature of humans* (Section 4.3.3). Researchers can delve into each of these limitations and explore solutions to help AI practitioners overcome these limitations, ultimately improving the incorporation of ethics in AI.
— *Exploring reasons for AI practitioners' awareness of specific ethical principles*: We found that the majority of the AI practitioners were aware of only a few ethical principles of AI including "*Privacy protection and security*" (64%), "*Reliability and safety*" (53%), and "*Human-centred values*" (53%) and very few practitioners (12%) were aware of *all* ethical principles of AI, as shown in Section 4.2.2. Researchers can conduct further studies to explore the reasons why AI practitioners are aware of only specific ethical principles of AI and how they may affect the development of ethical AI-based systems.
— *Exploring the terms "morals and ethics" and their impact on AI development*: We found that the survey participants in our study used the term "moral" frequently when asked about the challenges in incorporating ethics in AI. According to the participants, a person's moral perception has an impact on the incorporation of ethics in AI, demonstrating the interdependence of morality and ethics. Since moral philosophy is a very broad field, future research can be conducted to delve deeper into these terms (morals and ethics) and how they impact the development of ethical AI-based systems.

Recommendations for Education

— *Working toward including the topic of "AI ethics" in a curriculum*: Very few participants (2%) mentioned that *university courses* helped them to be aware of AI ethics and ethical principles, as discussed in Section 4.2.3. Therefore, we recommend that AI educators should include the topic of "AI ethics" in their curriculum to ensure that students are aware and knowledgeable about ethical considerations related to AI development. Likewise, effective training programs related to the incorporation of AI ethical principles can be organised to aid students in developing ethical AI-based systems.
— *Enhancing formal education/training programs*: In our study, a significant number of participants (43%) indicated that formal education/training plays a moderate role in equipping AI practitioners with the necessary skills to incorporate ethics into AI practices, as presented in Section 4.2.4. We recommend that AI educators concentrate on improving the quality of formal education/training initiatives to effectively assist AI practitioners in incorporating ethics and upholding ethical standards throughout the process of AI development.

6 LIMITATIONS AND THREATS TO VALIDITY

A limitation of our study is that it is based on the responses from only 100 participants. Among 100 participants, the majority of the participants reported that they are involved in data collection (61%), data cleaning (50%), and model training (38%), which indicates that we collected data from our target participants. We did not consider the team size, organisation size details, AI software domains, or information on the frequency or amount of performing AI development activities since our main focus was to gain AI practitioners' insights into their understanding of different aspects of ethics in AI, which is a limitation of this study. AI encompasses various fields such as ML, natural language processing, data science, and so on. We did not ask the participants what type of AI-based software they developed. Thus, further studies can focus on understanding them. To ensure that our survey is manageable for AI practitioners, we added some fundamental demographic questions.

In addition, we decided to investigate more intricate demographic and work-related environments in future research.

Similarly, in Section 3, we presented descriptive statistics regarding the factors influencing AI practitioners' familiarity with AI ethics. However, a limitation of our study lies in its limited sample size, consisting of only 100 participants. Consequently, our findings lack statistical significance. Therefore, future research should aim to gather data from a more extensive sample and test the link between factors such as work experience, education, job title, and gender of AI practitioners and their level of familiarity with the concept of AI ethics.

Instead of asking survey participants to provide their own description of AI ethics, we provided them with the AI ethics definition authored by Siau and Wang [50]. This decision was made due to concerns about the survey's length [30], and to avoid introducing an additional open-ended question, especially one to begin with, as this can be an easy point of quitting on the survey [25]. On the other hand, asking people to define AI ethics is not a trivial pursuit. It is not easy to put into words, in meaningful ways, as to what it means and this could further deter people from continuing. Even if people provided their own definitions, it would have still resulted in another set of limitations arising from disparate definitions for the purposes of analysing the remaining answers. By listing a definition, we wanted to clarify where we stand on the issue and build a shared understanding for answering the remaining questions. In other words, this was a research design tradeoff with pros and limitations in both cases. It is important to acknowledge this limitation of our study.

We included concise descriptions of each AI ethical principle in the survey to help participants comprehend their definitions. We refrained from presenting lengthy and exhaustive explanations within the survey which might have discouraged participants from completing it [28]. In order to mitigate the likelihood of participants opting out due to information overload, we opted for a condensed explanation of each AI ethical principle while providing a link to Australia's AI Ethics Principles list within that question for the participants who may have wanted to follow through for more information on AI Ethics Principles. Despite this approach, our study has the limitation that participants' selection of ethical principles could have been influenced if they did not understand the ethical principles provided and did not access the link to Australia's AI Ethics Principles.

Likewise, one of the key findings of our study is that only 2% of participants mentioned "university" as their reason for awareness of AI ethics and principles. Nevertheless, a limitation of our study is that a big part of the participants were over 30 years old, which may have influenced the data we collected since AI ethics is a new topic in the university curriculum if at all added, as explained in Section 5.1.3.

Similarly, the majority of our participants were data scientists (23%), and "Privacy protection and security" was the principle most respondents (64%) were aware of. The prevalence of data scientists in our pool may have influenced this finding because data scientists typically work with data which, in turn, has privacy issues.

Regarding the number of survey participants, although 190 participants started answering our survey questions (as shown in Qualtrics records), only 104 participants completed it. The target participants for the survey were AI practitioners involved in the development of AI-based systems, so we had to exclude four participants who were students, teachers, researchers, or who did not have experience in AI development. As a result, we had to include the responses of only 100 participants in our survey. All the authors were involved in all stages except that only the first, second, and third authors were involved in the analysis stage of the qualitative data. For the quantitative data, the first author conducted the analysis and shared it with all the other authors to discuss each followed step and technique.

After multiple discussions, the team finalised the best methods for presenting the findings of the qualitative study (as described in Section 3.3). We had multiple conversations regarding the

analysis, findings, and methods for presenting the results in order to minimise bias. There is a potential risk to the research's internal validity when using the payment for the second round of data collection. Nevertheless, we carefully examined previous studies [23] and decided to use Prolific. We included two *attention check questions* in between the survey questions to check if the participants were paying attention while answering the survey. Payments for participants were only approved after confirming that they were part of our target participant group, answered both the *attention check questions* correctly, and provided responses to every question.

7 CONCLUSION

Understanding AI practitioners' views on ethics in AI has been highlighted in the literature but the lack of empirical research on investigating AI practitioners' views on AI ethics motivated us to conduct this study. This study contributes to understanding the industry perspective on the *awareness* of ethics in AI and the *challenges* in incorporating ethics into AI-based systems.

We explored four aspects related to the AI practitioners' *awareness* of ethics in AI through our study. The aspects are (i) the extent to which AI practitioners are aware of the concept of ethics in AI, (ii) the ethical principles of AI that AI practitioners are aware of, (iii) reasons for AI practitioners' awareness of ethics and (iv) AI practitioners' awareness of the role of formal education or training in preparing them in incorporating AI ethics. We captured AI practitioners' insights through closed-ended questions and the data were analysed using descriptive statistics for analysis. Our results show that the majority of the participants are **moderately** aware of the concept of ethics in AI and ***privacy protection and security*** is the principle that the majority of the participants are aware of. Our results also indicate that **workplace rules and policies** play a major role in AI practitioners' awareness of ethics in AI and only a **few** AI practitioners thought that formal education or training is extremely helpful for them in incorporating ethics in AI.

Similarly, through an open-ended question, we obtained data on the key challenges that AI practitioners face in incorporating ethics in AI, and through a closed-ended question, we obtained insights on the degree of challenges faced by AI practitioners specific to implementing each AI ethical principle. We analysed the open-text answers using the *STGT method for data analysis* and closed-ended answers through descriptive statistics for data analysis and categorised the challenges into three sections which include, *general challenges*, *technology-related challenges*, and *human-related challenges*. We found that the majority of the participants believe that the **biased nature** of human beings is a major challenge in developing ethical AI-based systems. We also found that the majority of the participants find the incorporation of ***human-centered values*** the most challenging ethical principle during AI development. This study's results provide valuable insights into the industry's perspective on their *awareness* and *challenges* related to AI ethics and its incorporation. The AI research community will gain a better understanding of how AI practitioners view ethics in AI and the challenges they encounter while considering and following ethical principles during AI development. In addition, the study identified areas that require further investigation to benefit the industry, and AI practitioners can use these findings to improve their understanding and incorporation of AI ethics during AI development.

APPENDICES

A APPENDIX A: SURVEY QUESTIONS

Section A: Demographic Information

1. What is your current job title?
   — AI Engineer
   — AI/ Data Scientist

— AI/ML Specialist
— AI Expert
— AI/ML Practitioner
— AI Developer
— AI Designer
— Prefer not to answer
— Others:

2. How old are you?
— Below 20
— 20-25
— 26-30
— 31-35
— 36-40
— 41-45
— 46-50
— Above 50
3. How would you describe your gender?
— Man
— Woman
— Prefer to self-describe as:
— Prefer not to answer
4. What is your country of residence?
5. What is the highest degree or level of education you have completed?
— High School
— Bachelor's degree
— Master's degree
— Ph.D. or Higher
— Prefer not to answer
— Others:
6. What activities are you involved in? Select **all** that apply.
— Model requirements
— Data collection
— Data cleaning
— Data labeling
— Feature engineering
— Model training
— Model evaluation
— Model deployment
— Model monitoring
— Others:
7. How many years of experience do you have in AI-based software development?
— No experience
— Less than 1 year
— Between 1 to 2 years
— Between 3 to 5 years
— Between 6 to 10 years
— Between 11-15 years

— Between 16-20 years
— Over 20 years

Section B: AI Practitioners' Awareness of Ethics in AI "Ethics in AI refers to the principles of developing AI to interact with other AIs and humans ethically and function ethically in society." *From K. Siau and W. Wang, "Artificial intelligence ethics: Ethics of AI and ethical AI," Journal of Database Management, vol. 31, no. 2, pp. 74–87, 2020*

8. How familiar are you with the concept of ethics as it relates to AI development?
   — Very familiar
   — Reasonably Familiar
   — Somewhat familiar
   — Not very familiar
   — Not at all familiar
9. What made you aware of "ethics in AI"? Select all that apply.
   — Workplace rules and policies
   — Customer complaints
   — First-hand personal experience (e.g., as a software user)
   — First-hand professional experience (e.g., as an AI practitioner)
   — Through news and media
   — I have a personal interest in this
   — Not applicable
   — Others:

Attention-check question: The AI ethics test you are about to take part in is very simple, when asked for the most discussed ethical principle of AI, you must select "Fairness". This is an attention check.

Based on the text you read above, which ethical principle have you been asked to enter?

— Accountability
— Fairness
— Contestability
— Reliability and safety

10. Which of the following AI ethical principles are you aware of? Select **all** that apply. *These are a selected list of the majority of the principles considered around the world. (Australia's AI ethics principles: https://www.industry.gov.au/data-and-publications/australias-artificial-intelligence-ethics-framework/australias-aiethics-principles)*
    — Accountability: people identifiable and accountable for AI system outcomes
    — Contestability: timely process to allow people to challenge the AI system use/outcomes
    — Fairness: inclusive and accessible system
    — Human-centered values: respect human rights, diversity & autonomy of individuals
    — Human, societal, and environmental well-being: benefit individuals, society, and environment
    — Privacy protection and security: respect & uphold privacy rights & ensure data security
    — Reliability and safety: reliably operate in accordance with their intended purpose
    — Transparency and explainability: transparency & responsible disclosure to help people understand AI impacts & engagement
    — All
    — None
    — Others:

Table 4. Degree of Challenges in Considering and Following AI Ethical Principles

	Very	Reasonably	Somewhat	Not very	Not at all	No experience
Accountability	o	o	o	o	o	o
Contestability	o	o	o	o	o	o
Fairness	o	o	o	o	o	o
Human-centred values	o	o	o	o	o	o
Human, societal & environmental well-being	o	o	o	o	o	o
Privacy protection & security	o	o	o	o	o	o
Reliability & safety	o	o	o	o	o	o
Transparency & explainability	o	o	o	o	o	o

11. How well do you think your formal education/training prepared you to implement ethics in AI?
    — Extremely well
    — Very well
    — Moderately
    — Slightly
    — Not at all

Section C: AI Practitioners' Challenges of Incorporating Ethics in AI

12. In your experience, how challenging is it to consider and follow the following ethical principles when developing AI-based software solutions? (Please choose one option for each ethical principle in Table 4.)
    Attention-check question: In Australia's AI ethics principles list, how many ethical principles are included? Please select "8". This is an attention check.
    — 6
    — 7
    — 8
    — 9
13. In your experience, what are the main challenges or barriers to incorporating ethics in AI? (Open-text answer)
14. Based on your experience, is there anything else about ethics in AI you would like to share? (voluntary)
15. If you would like to participate in an interview on this topic with us, please share your name and email address (voluntary).

REFERENCES

[1] Sebastian Baltes and Paul Ralph. 2022. Sampling in software engineering research: A critical review and guidelines. *Empirical Software Engineering* 27, 4 (2022), 94. DOI: https://doi.org/10.1007/s10664-021-10072-8

[2] Veronika Bogina, Alan Hartman, Tsvi Kuflik, and Avital Shulner-Tal. 2021. Educating software and AI stakeholders about algorithmic fairness, accountability, transparency, and ethics. *International Journal of Artificial Intelligence in Education* 32 (2021), 808–833. DOI: https://doi.org/10.1007/s40593-021-00248-0

[3] Jason Borenstein and Ayanna Howard. 2021. Emerging challenges in AI and the need for AI ethics education. *AI and Ethics* 1, 1 (2021), 61–65. DOI: https://doi.org/10.1007/s43681-020-00002-7

[4] Nick Bostrom and Eliezer Yudkowsky. 2018. The ethics of artificial intelligence. In *Artificial Intelligence Safety and Security*, Roman V. Yampolskiy (Ed.). Chapman and Hall/CRC, 57–69.

[5] Jiyoo Chang and Christine Custis. 2022. Understanding implementation challenges in machine learning documentation. In *Proceedings of the ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. Number 16, ACM, New York, NY, 1–8. DOI: https://doi.org/10.1145/3551624.3555301

[6] Shruthi Sai Chivukula, Aiza Hasib, Ziqing Li, Jingle Chen, and Colin M. Gray. 2021. Identity claims that underlie ethical awareness and action. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13. DOI: https://doi.org/10.1145/3411764.3445375

[7] Shruthi Sai Chivukula, Chris Rhys Watkins, Rhea Manocha, Jingle Chen, and Colin M. Gray. 2020. Dimensions of UX practice that shape ethical awareness. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13. DOI: https://doi.org/10.1145/3313831.3376459

[8] Eleni Christodoulou and Kalypso Iordanou. 2021. Democracy under attack: Challenges of addressing ethical issues of AI and big data for more democratic digital media and societies. *Frontiers in Political Science* 3 (2021), 682945. DOI: https://doi.org/10.3389/fpos.2021.682945

[9] Bo Cowgill, Fabrizio Dell'Acqua, Samuel Deng, Daniel Hsu, Nakul Verma, and Augustin Chaintreau. 2020. Biased programmers? Or biased data? A field experiment in operationalizing AI ethics. In *Proceedings of the 21st ACM Conference on Economics and Computation*. New York, NY, 679–681. DOI: https://doi.org/10.1145/3391403.3399545

[10] Anne A. H. de Hond, Marieke M. van Buchem, and Tina Hernandez-Boussard. 2022. Picture a data scientist: A call to action for increasing diversity, equity, and inclusion in the age of AI. *Journal of the American Medical Informatics Association* 29, 12 (2022), 2178–2181. DOI: https://doi.org/10.1093/jamia/ocac156

[11] Iris Dominguez-Catena, Daniel Paternain, and Mikel Galar. 2022. Assessing demographic bias transfer from dataset to model: A case study in facial expression recognition. arXiv:2205.10049. Retrieved from https://arxiv.org/abs/2205.10049

[12] Amitai Etzioni and Oren Etzioni. 2017. Incorporating ethics into artificial intelligence. *The Journal of Ethics* 21 (2017), 403–418. DOI: https://doi.org/10.1007/s10892-017-9252-2

[13] Jessica Fjeld, Nele Achten, Hannah Hilligoss, Adam Nagy, and Madhulika Srikumar. 2020. Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. *Berkman Klein Center Research Publication* 2020-1 (2020). DOI: https://doi.org/10.2139/ssrn.3518482

[14] Google. 2022. Google AI. Retrieved April 10, 2023 from https://ai.google/principles/

[15] Leah Govia. 2020. Coproduction, ethics and artificial intelligence: A perspective from cultural anthropology. *Journal of Digital Social Research* 2, 3 (2020), 42–64. DOI: https://doi.org/10.33621/jdsr.v2i3.53

[16] High-Level Expert Group. 2019. Ethics Guidelines for Trustworthy AI. Retrieved September 6, 2023 from https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[17] Thilo Hagendorff. 2020. The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines* 30, 1 (2020), 99–120. DOI: https://doi.org/10.1007/s11023-020-09517-8

[18] Rashina Hoda. 2021. Socio-technical grounded theory for software engineering. *IEEE Transactions on Software Engineering* 48, 10 (2021), 3808–3832. DOI: https://doi.org/10.1109/TSE.2021.3106280

[19] Kenneth Holstein, Jennifer Wortman Vaughan, Hal Daumé III, Miro Dudik, and Hanna Wallach. 2019. Improving fairness in machine learning systems: What do industry practitioners need?. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–16. DOI: https://doi.org/10.1145/3290605.3300830

[20] Livia Iacovino. 2002. Ethical principles and information professionals: Theory, practice and education. *Australian Academic & Research Libraries* 33, 2 (2002), 57–74. DOI: https://doi.org/10.1080/00048623.2002.10755183

[21] Javier Camacho Ibáñez and Mónica Villas Olmeda. 2022. Operationalising AI ethics: How are companies bridging the gap between practice and principles? An exploratory study. *AI & SOCIETY* 37, 4 (2022), 1663–1687. DOI: https://doi.org/10.1007/s00146-021-01267-0

[22] IBM. 2022. AI Ethics. Retrieved April 8, 2023 from https://www.ibm.com/artificial-intelligence/ethics

[23] Jirayus Jiarpakdee, Chakkrit Kla Tantithamthavorn, and John Grundy. 2021. Practitioners' perceptions of the goals and visual explanations of defect prediction models. In *Proceedings of the 2021 IEEE/ACM 18th International Conference on Mining Software Repositories*. IEEE, 432–443. DOI: https://doi.org/10.1109/MSR52588.2021.00055

[24] Anna Jobin, Marcello Ienca, and Effy Vayena. 2019. The global landscape of AI ethics guidelines. *Nature Machine Intelligence* 1, 9 (2019), 389–399. DOI: https://doi.org/10.1038/s42256-019-0088-2

[25] Mark Kasunic. 2005. *Designing an Effective Survey*. Technical Report. Carnegie Mellon University, Software Engineering Institute Pittsburgh, PA.

[26] Emre Kazim and Adriano Soares Koshiyama. 2021. A high-level overview of AI ethics. *Patterns* 2, 9 (2021). DOI: https://doi.org/10.2139/ssrn.3609292

[27] Arif Ali Khan, Muhammad Azeem Akbar, Muhammad Waseem, Mahdi Fahmideh, Aakash Ahmad, Peng Liang, Mahmood Niazi, and Pekka Abrahamsson. 2022. AI ethics: Software practitioners and lawmakers points of view. *IEEE Transactions on Computational Social Systems* 10, 6 (2023), 2971–2984. DOI: 10.1109/TCSS.2023.3251729

[28] Barbara A. Kitchenham and Shari L. Pfleeger. 2008. Personal opinion surveys. In *Guide to Advanced Empirical Software Engineering*, Forrest Shull, Janice Singer, and Dag I. K. Sjøberg (Eds.). Springer, 63–92.

[29] Will Knight. 2023. Scammers used ChatGPT to Unleash a Crypto Botnet on X. Retrieved September 6, 2023 from https://www.wired.com/story/chat-gpt-crypto-botnet-scam/

[30] Johan Linaker, Sardar Muhammad Sulaman, Martin Höst, and Rafael Maiani de Mello. 2015. Guidelines for conducting surveys in software engineering v. 1.1. Vol. 50. Lund University.

[31] Michael Madaio, Lisa Egede, Hariharan Subramonyam, Jennifer Wortman Vaughan, and Hanna Wallach. 2022. Assessing the fairness of AI systems: AI practitioners' processes, challenges, and needs for support. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–26. DOI: https://doi.org/10.1145/3512899

[32] Ryan Mark and Gregory Anya. 2019. Ethics of using smart city AI and big data: The case of four large European cities. *The ORBIT Journal* 2, 2 (2019), 1–36. DOI: https://doi.org/10.29297/orbit.v2i2.110

[33] Nicole Martin. 2018. Are AI Hiring Programs Eliminating Bias or Making it Worse? Retrieved August 2, 2022 from https://www.forbes.com/sites/nicolemartin1/2018/12/13/are-ai-hiring-programs-eliminating-bias-or-making-it-worse/?sh=552bb0cc22b8

[34] Andrew McNamara, Justin Smith, and Emerson Murphy-Hill. 2018. Does ACM's code of ethics change ethical decision-making in software development?. In *Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. 729–733. DOI: https://doi.org/10.1145/3236024.3264833

[35] Microsoft. 2022. Microsoft Responsible AI Standard. Retrieved April 20, 2023 from https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1%3aprimaryr6

[36] Brent Mittelstadt. 2019. Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence* 1, 11 (2019), 501–507. DOI: https://doi.org/10.1038/s42256-019-0114-4

[37] Jessica Morley, Anat Elhalal, Francesca Garcia, Libby Kinsey, Jakob Mökander, and Luciano Floridi. 2021. Ethics as a service: A pragmatic operationalisation of AI ethics. *Minds and Machines* 31, 2 (2021), 239–256. DOI: https://doi.org/10.1007/s11023-021-09563-w

[38] B. Nalini. 2020. The Hitchhiker's Guide to AI Ethics. Retrieved July 15, 2022 from https://towardsdatascience.com/ethics-of-ai-a-comprehensive-primer

[39] Madelena Y. Ng, Supriya Kapur, Katherine D. Blizinsky, and Tina Hernandez-Boussard. 2022. The AI life cycle: A holistic approach to creating ethical AI for health decisions. *Nature Medicine* 28, 11 (2022), 2247–2249.

[40] Dare Obasanjo. 2021. Abandoning GitHub. Retrieved August 9, 2022 from https://twitter.com/youyuxi/status/1411824059780849675

[41] Will Orr and Jenny L. Davis. 2020. Attributions of ethical responsibility by Artificial Intelligence practitioners. *Information, Communication & Society* 23, 5 (2020), 719–735. DOI: https://doi.org/10.1080/1369118X.2020.1713842

[42] Aastha Pant, Rashina Hoda, Chakkrit Tantithamthavorn, and Burak Turhan. 2023. Ethics in AI through the developer's view: A grounded theory literature review. arXiv:4916508. Retrieved from https://arxiv.org/abs/4916508

[43] Dinah Payne and Brenda E. Joyner. 2006. Successful US entrepreneurs: Identifying ethical decision-making and social responsibility behaviors. *Journal of Business Ethics* 65 (2006), 203–217. DOI: https://doi.org/10.1007/s10551-005-4674-3

[44] Thomas M. Powers and Jean-Gabriel Ganascia. 2020. *The Ethics of the Ethics of AI*. Oxford University Press, Oxford.

[45] Marcelo O. R. Prates, Pedro H. Avelar, and Luís C. Lamb. 2020. Assessing gender bias in machine translation: A case study with Google translate. *Neural Computing and Applications* 32 (2020), 6363–6381. DOI: https://doi.org/10.1007/s00521-019-04144-6

[46] Bogdana Rakova, Jingying Yang, Henriette Cramer, and Rumman Chowdhury. 2021. Where responsible AI meets reality: Practitioner perspectives on enablers for shifting organizational practices. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–23. DOI: https://doi.org/10.1145/3449081

[47] Lea Rothenberger, Benjamin Fabian, and Elmar Arunov. 2019. Relevance of ethical guidelines for artificial intelligence-A survey and evaluation. In *Proceedings of the 27th European Conference on Information Systems*.

[48] Mark Ryan, Josephina Antoniou, Laurence Brooks, Tilimbe Jiya, Kevin Macnish, and Bernd Stahl. 2021. Research and practice of AI ethics: A case study approach juxtaposing academic discourse with organisational reality. *Science and Engineering Ethics* 27, 16 (2021), 1–29. DOI: https://doi.org/10.1007/s11948-021-00293-x

[49] Conrad Sanderson, David Douglas, Qinghua Lu, Emma Schleiger, Jon Whittle, Justine Lacey, Glenn Newnham, Stefan Hajkowicz, Cathy Robinson, and David Hansen. 2023. AI ethics principles in practice: Perspectives of designers and developers. *IEEE Transactions on Technology and Society* 4, 2 (2023), 171–187. DOI: https://doi.org/10.1109/TTS.2023.3257303

[50] Keng Siau and Weiyu Wang. 2020. Artificial intelligence ethics: Ethics of AI and ethical AI. *Journal of Database Management* 31, 2 (2020), 74–87. DOI: https://doi.org/10.4018/JDM.2020040105

[51] Melika Soleimani, Ali Intezari, Nazim Taskin, and David Pauleen. 2021. Cognitive biases in developing biased Artificial Intelligence recruitment system. *Hawaii International Conference on System Sciences* 54 (2021), 5091–5099.

[52] Bernd Carsten Stahl, Josephina Antoniou, Mark Ryan, Kevin Macnish, and Tilimbe Jiya. 2022. Organisational responses to the ethical issues of artificial intelligence. *AI & SOCIETY* 37, 1 (2022), 23–37. DOI: https://doi.org/10.1007/s00146-021-01148-6

[53] Tara Qian Sun and Rony Medaglia. 2019. Mapping the challenges of artificial intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly* 36, 2 (2019), 368–383. DOI: https://doi.org/10.1016/j.giq.2018.09.008

[54] Ruchika Tulshyan. 2022. *Inclusion on Purpose: An Intersectional Approach to Creating a Culture of Belonging at Work*. MIT Press. DOI: https://doi.org/10.7551/mitpress/14004.001.0001

[55] Ville Vakkuri and Pekka Abrahamsson. 2018. The key concepts of ethics of artificial intelligence. In *Proceedings of the 2018 IEEE International Conference on Engineering, Technology and Innovation*. IEEE, 1–6. DOI: https://doi.org/10.1109/ICE.2018.8436265

[56] Ville Vakkuri, Marianna Jantunen, Erika Halme, Kai-Kristian Kemell, Anh Nguyen-Duc, Tommi Mikkonen, and Pekka Abrahamsson. 2021. Time for AI (ethics) maturity model is now. arXiv:2101.12701. Retrieved from https://arxiv.org/abs2101.12701

[57] Ville Vakkuri, Kai-Kristian Kemell, and Pekka Abrahamsson. 2019. Ethically aligned design: An empirical evaluation of the resolvedd-strategy in software and systems development context. In *Proceedings of the 2019 45th Euromicro Conference on Software Engineering and Advanced Applications*. IEEE, 46–50. DOI: https://doi.org/10.1109/SEAA.2019.00015

[58] Ville Vakkuri, Kai-Kristian Kemell, and Pekka Abrahamsson. 2019. Implementing ethics in AI: Initial results of an industrial multiple case study. In *Product-Focused Software Process Improvement. PROFES*, Xavier Franch, Tomi Männistö, and Silverio Martínez-Fernández (Eds.). Lecture Notes in Computer Science, Springer, 331–338. DOI: https://doi.org/10.1007/978-3-030-35333-9_24

[59] Ville Vakkuri, Kai-Kristian Kemell, and Pekka Abrahamsson. 2020. ECCOLA- A method for implementing ethically aligned AI systems. In *Proceedings of the 2020 46th Euromicro Conference on Software Engineering and Advanced Applications*. IEEE, 195–204. DOI: https://doi.org/10.1109/SEAA51224.2020.00043

[60] Ville Vakkuri, Kai-Kristian Kemell, Marianna Jantunen, and Pekka Abrahamsson. 2020. "This is just a prototype": How ethics are ignored in software startup-like environments. In *Proceedings of the 21st International Conference on Agile Software Development*. Springer International Publishing, 195–210. DOI: https://doi.org/10.1007/978-3-030-49392-9

[61] Ville Vakkuri, Kai-Kristian Kemell, Joni Kultanen, and Pekka Abrahamsson. 2020. The current state of industrial practice in artificial intelligence ethics. *IEEE Software* 37, 4 (2020), 50–57. DOI: https://doi.org/10.1109/MS.2020.2985621

[62] Ville Vakkuri, Kai-Kristian Kemell, Joni Kultanen, Mikko Siponen, and Pekka Abrahamsson. 2019. Ethically aligned design of autonomous systems: Industry viewpoint and an empirical study. *Electronic Journal of Business Ethics and Organization Studies* 27, 1 (2022), 4–15.

[63] Michael Veale, Max Van Kleek, and Reuben Binns. 2018. Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. In *Proceedings of the 2018 CHI conference on Human Factors in Computing Systems*. 1–14. DOI: https://doi.org/10.1145/3173574.3174014

[64] Jess Whittlestone, Rune Nyrup, Anna Alexandrova, Kanta Dihal, and Stephen Cave. 2019. Ethical and societal implications of algorithms, data, and artificial intelligence: A roadmap for research. Nuffield Foundation, London.

[65] Claes Wohlin, Per Runeson, Martin Höst, Magnus C. Ohlsson, Björn Regnell, and Anders Wesslén. 2012. *Experimentation in Software Engineering*. Springer Science & Business Media, Doedrecht.

[66] Didar Zowghi and Francesca da Rimini. 2023. Diversity and inclusion in artificial intelligence. arXiv:2305.12728. Retrieved from https://arxiv.org/abs/2305.12728

Received 26 May 2023; revised 13 November 2023; accepted 20 November 2023