AI and Ethics
https://doi.org/10.1007/s43681-024-00458-x

ORIGINAL RESEARCH

The ethical wisdom of AI developers

Tricia A. Griffin[1] · Brian P. Green[2] · Jos V.M. Welie[3,4]

Received: 15 July 2023 / Accepted: 5 March 2024

Abstract
This paper explores ethical wisdom in the artificial intelligence (AI) developer community. Despite robust literature about the need for virtue ethics approaches in AI development, little research has directly engaged with the developer community about their progress in this regard. We have thus conducted semi-structured interviews with a worldwide cohort of 40 developers, which focused on their awareness of ethics issues, how they navigate ethical challenges, and the barriers they encounter in developing ethical wisdom. We find developers are largely aware of the ethical territories they must navigate and the moral dilemmas they personally encounter, but they face limited and inconsistent resources for ethical guidance or training. Furthermore, there are significant barriers inhibiting the development of ethical wisdom in the AI developer community, including the industry's fixation on innovation, the narrow scope of technical practice, limited provisions for reflection and dialogue, and incentive structures that prioritize profits and prestige. The paper concludes by emphasizing the need to address the gap in domain-specific ethical skill and provides recommendations for organizations, educators, and the AI developer community.

Keywords Ethical AI · Professional ethics · Virtue ethics · Practical wisdom · Artificial intelligence ethics · Data science ethics · Machine learning ethics

> Nobody is going to ask you, while you are developing it, nobody is going to ask you how ethical issues were considered. If people complain, now they'll come back to you [and ask], 'Why have you done this?' But we don't have explicit guidelines on how to do it. So, it's up to the developers to think [about] the best way. That's how it works – 3943_Robotics.

1 Introduction

The ongoing real and potential harms generated by artificial intelligence (AI) deployments [1, 2] have been met with persistent calls for systems that comply with a multitude of human values, like fairness, responsibility, and justice [3]. Yet, achieving this goal has proven challenging [4, 5]. This is in part because these principles are difficult, if not impossible, to code [6]. But also because they are embodied concepts. They are lived by people and negotiated over and over again as contexts and cultures change. As a result, AI developers[1] must know something (indeed, a lot) about fairness, responsibility, and justice before they can hope to incorporate those principles into

✉ Tricia A. Griffin
t.griffin@maastrichtuniversity.nl

[1] Faculty of Science and Engineering, Maastricht University, Zwingelput 4, Maastricht 6211 KH, The Netherlands

[2] Markkula Center for Applied Ethics, Santa Clara University, California, USA

[3] Faculty of Science and Engineering, Maastricht University, Maastricht, The Netherlands

[4] St. André International Center for Ethics and Integrity, Saint André d'Olérargues, France

[1] We use the term "AI developer" throughout this paper to cover a spectrum of specializations practiced by the developers interviewed for this study. We do this for the sake of readability and not to reduce a rich and complex set of skills to a single term. Our participants identified as practitioners of artificial intelligence (AI), machine learning (ML), data science (DS), deep learning (DL), robotics (RO), and research science (RS).

Published online: 20 March 2024

automated systems.[2] This requires wisdom, not just technical skill, and it is the subject of this paper.

This paper is not about theoretical wisdom. It is about practical wisdom. It is about the kind of wisdom that can only be gained through judicious awareness, lived experience, directed learning, and self-reflection. In short, we are concerned here with what traditionally have been called virtues. We contend that cultivating the virtues in the AI developer community will have certain compatibilities with ethical approaches currently being utilized in practice, such as consequentialism, deontology, principlism, and casuistry. This compatibility is germane because it is virtues that society wants to see reflected in the outputs of automated systems, and so those virtues need to be understood at a practically wise level by the practitioners tasked with building them.

The state of practical wisdom in the AI developer community is, at present, difficult to assess. This is both because of a lack of empirical data about this population and because the complex nature of practical wisdom makes measuring it challenging [8, 9]. Nevertheless, this paper attempts to provide a grounded account of practical wisdom in the AI developer community and how to facilitate further development in this regard. We rely on insights from semi-structured interviews with 40 developers, which revealed how aware they are of ethics issues in their domain, how they navigate the ethical issues they personally experience, if (and where) they seek help, and the formidable currents that work against their gaining more ethical wisdom.

We begin with a brief overview of ethical wisdom in occupations. We then explain our qualitative methodology for the interviews we conducted with developers. The third section will discuss relevant findings about practical wisdom in the AI community, organized into four themes: ethical sensitivity, navigating ethical territories, ethics training, and barriers to practical wisdom. In the discussion section, we consider the implications of these findings, including how developers may be personifying cultural beliefs about the liberating power of technology. Following a brief description of the study's limitations, we conclude with implementation guidance for organizations, educators, and the AI developer community.

2 Ethical wisdom

Rather than present an exhaustive review of the literature about ethical wisdom, we will here limit our discussion to how it manifests in contemporary occupations, and specifically computer engineering from which AI development emerged. In most occupational settings, ethical practice is thought to be achieved via compliance, or rules-based practices. This understanding is concerned with *what it is good to do*. And the answer to that question is often found in complying with specific sets of rules, such as company policy, government laws, and industry regulations. An ethically richer understanding seeks to derive such rules from a utilitarian determination of the greatest good for the greatest number of people or from a deontological determination of duties, whether grounded in reason (as Immanuel Kant proposed) or more commonly captured in oaths and codes of ethics.

Indeed, for computer engineering, as for most occupations, what it is good to do is often expressed in codes of ethics. These codes usually call their members to uphold the law, but they also establish standards of practice that should apply when the law is silent, or when common sense and occupational norms are no longer adequate [10]. The code of ethics from the Association for Computing Machinery (ACM) [11], for example, states that computing professionals "have a special responsibility to provide objective, credible evaluations and testimony" and that they should "provide full disclosure of all pertinent system capabilities [and] limitations" among other guidance. Likewise, the code of ethics for the Institute of Electrical and Electronics Engineers (IEEE) [12] tells computing professionals to "improve the understanding by individuals and society of the capabilities and societal implications of conventional and emerging technologies" and to "hold paramount the safety, health, and welfare of the public."

Ethics codes specifically written for AI development are numerous. Nearly 100 ethics codes currently purport to guide how practitioners ought to ply their trade [3]. It is not surprising that the field sought to address the repeated ethical crises of AI deployments by first issuing these codes of ethics. The practice of building automated systems is rules-based, and it stood to reason that rules- or principles-based ethics could help guide such a practice. Yet, these codes have so far proven inadequate for the task of building ethical AI. As we noted in our introduction, the principles espoused in these codes are contextual. Knowing how to apply them in practice requires prudence, or the ability to find the right way to do things in a given situation, and that in turn requires some skill in working through ethical issues.

[2] This echoes a similar argument made by Carl Mitcham [7] regarding the inadequacy of engineering to achieve the ends of public health, safety, and welfare. He argues that while engineering aims for these good ends, engineers themselves receive no special training or knowledge that would make them competent to make judgements about health, safety, or welfare. We argue this is also the case for AI developers, who receive little or no training or knowledge about the myriad principles that they are nonetheless expected to code into algorithms.

This brings us to yet another way to think about ethics in occupations, which is as a character-based practice. In this understanding, we are interested in *who it is good to be*, and here we find ourselves in the realm of virtue ethics. Virtue ethics is concerned with the intellectual, emotional, and psychological habits that culminate in a person of practical wisdom [13]. That is, someone for whom doing the right thing at the right time and in the right way has become a matter of habit [14, 15]. This does not mean that virtues are instinctive. On the contrary, a central, and often overlooked, feature of virtue ethics is that one must deliberately practice it. To become virtuous, a person must decide to be fair in all their dealings, cultivate a community of like-minded practitioners, and apply themselves to both the study and practice of fairness. Practical wisdom thus encompasses a person's moral disposition, education, and level of experience (as evidenced both by how long they have been in occupational practice and the issues they have confronted before), as well as the network of people they rely on, and a commitment to self-reflection/improvement.

As mentioned in our introduction, the status of the virtues in the computer engineering and AI development community is, at present, difficult to determine. This is not to say there is no conversation about it. Numerous scholars have been calling for a more robust emphasis on virtue ethics in both engineering practice broadly and in AI development [16–20]. Prominent textbooks on engineering and computer ethics (for example, [21–23]) likewise reference the importance of virtue ethics in some way. Yet, the dominant perception of *who AI developers actually are* still tend to describe them in cynical terms – as either savant-like geniuses "possessed of an Olympian brilliance and productivity" [24] or as entitled "tech bros" who pursue their own interests without regard for, and sometimes purely to spite, others [25, 26]. This disconnect has arguably been fostered by Western notions of character development that prize radical individualism, libertarian freedom, and the idea of the "heroic engineer" [27].

We must emphasize that these two accounts of being ethical in occupations – *what it is good to do* and *who it is good to be* – are not mutually exclusive. As a practical matter, both are always operative. They are also interdependent. With every action we take we also become a certain kind of person; and the kind of person we are influences the actions we will (or will not) take. Furthermore, in both accounts the focus is traditionally on the intentions and actions of an individual actor; an "I" whose choices can be rationally explained as good, bad, or something in between.

This traditional focus on individual actions has also made it notoriously difficult to judge the ethics of individual actors in computer engineering (and in engineering practice broadly), because there is often not just one "I" who is making decisions. As Basart and Serra [28] have pointed out, "Engineers are not a singularity inside engineering; they exist and operate as a node in a complex network of mutual relationships with many other nodes." Likewise, in AI practice there can be dozens or even hundreds of developers, all building pieces of a system without ever seeing the full picture.[3] This "problem of many hands" [29] is compounded by the fact that developers see themselves as "constrained agents" [30] who have a great deal of autonomy bestowed on them because of their technical expertise, yet who must still enact the decisions of others with more authority (e.g., their superiors or a client). We will take up this problem again in our conclusion.

3 Methodology

This study proceeded from a constructivist worldview, which posits that people learn by applying new information or experiences to past knowledge and beliefs. As noted in Cobern [31], *knowing* something involves more than just taking in facts or theories. Those facts and theories must also *make sense* in the context of our lives and experiences. Applying this understanding of learning to the study of ethical wisdom in the AI developer community requires that we engage directly with developers both about what (if anything) they know about ethics in their domain and whether/how they make sense of it. We do this both to provide empirical evidence about practical wisdom in the AI developer community, and to better understand what tools and structures may be necessary if they are to fulfill the moral tasks and responsibilities before them.

As described in Griffin et al. [32], semi-structured interviews with 40 AI developers were conducted between February 2022 and October 2022.[4] An international cohort of participants was recruited via LinkedIn for 45-minute interviews about the ethics of being an AI developer. We limited recruitment to developers actively

[3] While our analysis here is limited to practitioners directly coding AI systems, we also acknowledge that there are tens of thousands and even millions of people contributing to the functioning of these systems in the form of data cleaners, reinforcement learning practitioners, and even the public.

[4] This paper relies on findings from the same set of interviews reported in Griffin et al. [32]. That paper focused on the *ethical agency* of AI developers. This paper covers themes related to *ethical wisdom*, which were not addressed in the first paper.

involved in the design, development, and maintenance of AI systems. Interview candidates' LinkedIn profiles were closely reviewed to ensure that they had experience working with natural language models, recommender systems, chatbots, robotics, vision/multimodal sensing, deep learning, or artificial neural networks. Participant demographics are detailed in Table 1.

To reduce the risk of reprisal from an employer for participating in this study, participants were guaranteed anonymity and were also asked to speak only for themselves (not as a representative of any former or current employer). Additionally, all communication was conducted via LinkedIn, personal email, or encryption-based messaging. Unless the participant insisted otherwise, interviews took place outside of normal working hours to avoid any overlap with their official work duties. There was also a mutual agreement between participants and the interviewer to not discuss any specific projects, deployments, or companies.

An interview guide was used to direct the discussions, but not all questions were asked to each participant and depending on participant responses, additional follow-up questions were asked. All interviews were conducted by a single researcher (Griffin). Unless the participant declined, interviews were recorded for the sole purpose of obtaining an accurate transcript and then permanently deleted within 48 h. For the two instances in which participants declined to be recorded, answers were logged manually. Transcripts were anonymized and saved under a random number plus a generic job description (e.g., 7534_MachineLearning). To further assure participants that they would remain anonymous, we also guaranteed that the full transcripts would not be made publicly available. The study received ethics approval from the university's research ethics committee.

4 Findings

What follows is a summary of insights about ethical wisdom gained from our interviews. Themes included cover ethical sensitivity, navigating ethical territories, ethics training, and barriers to ethical wisdom.

4.1 Ethical sensitivity

Ethical sensitivity is the ability to recognize an ethical issue when it arises, and it is important in occupations because if a decision is not recognized as ethical in nature, moral reasoning will not be called upon to address it [33]. This issue is heightened in AI development because the developers interviewed for this study very often see the choices they make as purely technical in nature [32]; and when professionals are focused narrowly on technical issues they tend to overlook ethical ones [34, 35]. We began at the macro level, asking whether participants were aware of the ethics debates happening in their field. Responses fell into four categories: no awareness, vague awareness, specific awareness, and nuanced awareness.

Only three (3) participants reported having *no awareness* of ethics debates in the AI field. Of these, one expressed an interest in learning more while the other two indicated that they actively avoided it or preferred to focus on "learning and trying to implement AI, not thinking about ethics," as participant 4933_DataScience said. Five (5) developers gave *vague* responses. They could name an issue, like bias,

Table 1 Participant demographics

Sex		**Region**			**Industry**	
Female	19		*Employed*	*Origin*	Academia	07
Male	21	Africa	09	09	Agriculture	05
					Civil Eng.	01
		Asia	01	08	Commerce	02
Ethnicity		Australia Pacific	01	01	Defense	01
BIPOC	25	Central & South America	04	04	Entertainment	01
					Finance	05
		Europe	09	06	Healthcare	07
White	15	North America	16	12	Human Resources	01
					Insurance	01
					Technology	07
Highest Degree		**Primary Expertise**			Transportation	02
Bachelor	04	Artificial Intelligence	03			
		Data Science	13			
		Deep Learning	02			
Master	14	Machine Learning	14			
PhD	22	Research	03			
		Robotics	05			

but also admitted that their knowledge was limited to what they heard in the mainstream press. Participant 1855_Robotics, for example, briefly mentioned a dilemma a self-driving car might have to negotiate before saying, "I do, I guess, read the news and when stories become really popular, I do follow them, but I don't really sit down and say, 'Now I am gonna do some research on ethics.'".

Most responses ($n=22$) fell into the *specific* awareness category. Here, participants could name one or more issues along with some limited detail about why those issues were ethically relevant. These respondents spoke chiefly about bias in data sets. They pointed to biases that emerge in facial recognition technology, setting insurance premiums, mortgage lending, allocation of care, and criminal justice. When speaking about these biases, one participant, 2182_MachineLearning, noted that developers were not initially thinking about what they were doing in terms of bias, saying, "Early papers on [bias] were kind of eye opening because before that we were all just running data experiments."

Ten (10) participants demonstrated a more *nuanced* understanding of what was at stake and how many layers of debate are active. Responses in this category covered varying gradients of ethical detail. Participant 4807_ArtificialIntelligence, for example, spoke at length about the complexities of privacy, including whether algorithms should access personal data only if it benefited the user or if they should also do so to contribute to company analytics. This participant also spoke about the dangers of turning over too much decision-making authority to "increasingly complicated and opaque algorithms" and whether the big tech companies were too big to fail. Meanwhile, participant 7591_DataScience offered a sociological critique of information systems as they are currently deployed, saying, "We've digitized everything, so every system has a new logic to it. I would call it an a-human logic. You can't argue with the computer the way you can argue with another person."

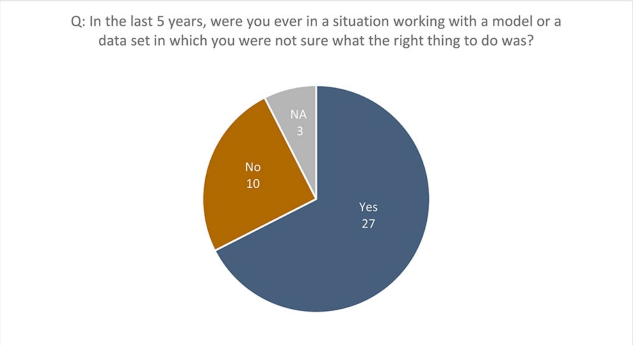


Fig. 1 Developers' personal encounters with ethics questions

4.2 Navigating ethical territories

As an extension of ethical sensitivity, we also explored whether and how developers navigated the ethical dilemmas they personally encountered. We began by asking if, at any time in the previous five years, the developer was not sure what the ethical thing to do was while working with a data set or a model. As shown in Fig. 1, twenty-seven (27) said they had been in such a situation and ten (10) said they had not. Three (3) were not asked the question either because of time limitations or because the conversation went in a different direction.

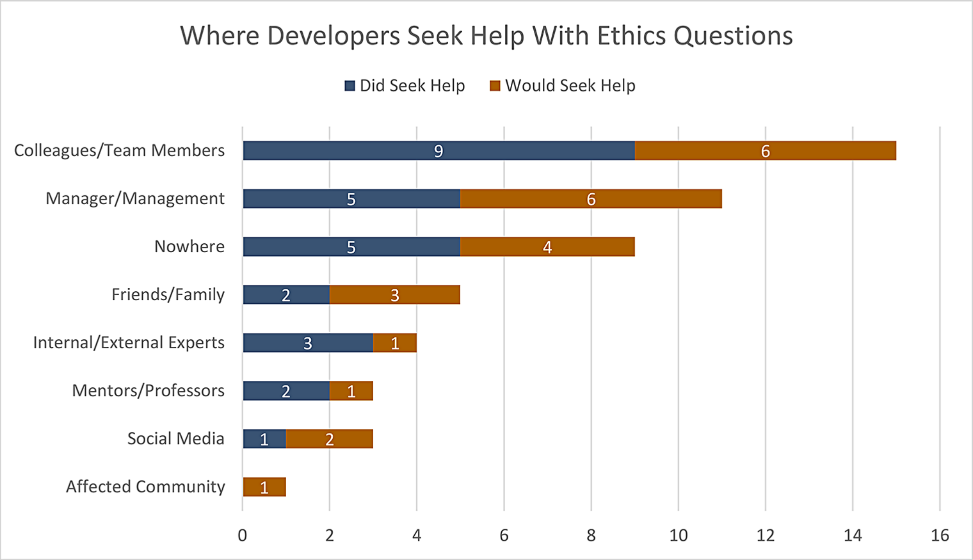
Participants who responded affirmatively were asked if they sought help for that situation, and if so, where. If they answered in the negative, they were asked to hypothesize such a situation and then to discuss where they might go for help. Figure 2 illustrates that most participants ($n=15$) had (or would) go to trusted colleagues or other team members for help with an ethical dilemma. The next most common source was to seek help from their manager or to escalate it through management ($n=11$). Nine (9) said there was nowhere to go for help. Four (4) had access to internal or external experts who could help them navigate ethical issues. The remaining sources of help ranged from friends and family ($n=5$) to mentors ($n=3$) to social media ($n=3$). Notably, only one (1) participant said they would seek help from the community affected by the system, and none said they would (or had) consult(ed) an ethics code.

Verbatims shown in Table 2 provide a side-by-side sample of the issues developers in this study faced and where they went for help. In most cases, these situations were classic dilemmas. The developer had a choice between two or more options and could only choose one. Less common were cases of ethical distress, wherein the developer knew what the right course of action was, but were prevented from doing so.

The range of dilemmas participants in this study experienced are as diverse as the uses of the systems they build. There was no common theme. In some cases, the dilemmas were more personal in nature, like for 1855_Robotics, who struggled with the knowledge that their robotics research might mean members of their community would eventually lose their jobs. This practitioner did not seek help negotiating this issue, because the incentives in place were designed to support the researcher's own productivity, not necessarily the community's wellbeing.

Others, like 4970_ResearchScience, recalled being faced with managing the changing winds of ethics guidance. This participant pointed out that where guidance once recommended against collecting race data if it was not necessary for the functioning of a product, it now encourages companies to ensure that their product performs equitably for

Fig. 2 Where developers go for help with ethics questions
Note: Most participants gave more than one answer to this question. The numbers, therefore, do not add up to 40, nor do they align neatly with those in Figure 1



Q: If yes to question illustrated in Figure 1: Where did you go for help with this situation? *If no:* Where might you go for help if you were to find yourself in a situation in which you didn't know what the right thing to do was?

all races. To satisfy the new guidance, this developer would have to infer race from other data, a practice that has its own ethical issues. This developer, in contrast to the first, did have internal policy experts to go to for help.

Of the four (4) developers who reported experiencing ethical distress, three (3) said they left the team or the company because of that situation. These participants reported there was nowhere to go for help and that whistleblowing was not a viable option for them. The fourth developer in this category expressed a similar, if more generalized, articulation of the whistleblower's distress, saying:

> The way it is, you do what the companies want. You abide by those rules, and if you break those rules you're blacklisted. It is highly likely you're not going to get hired again simply because they don't know when you're going to want to play ball or not, even if the public thinks highly of you for reporting that kind of information – 4807_ArtificialIntelligence.

This participant did not leave their place of employment despite knowing they were in this predicament because they reported having positive experiences working through other ethical issues in their team. Yet, this participant also pointed to a deeper issue facing many practitioners, which is whether developers' ethical responsibility ought to be limited to the proper functioning of the technology or if it should extend to the impacts of it. This participant offered the following as an illustrative example:

> In some cases, technologists' hands are tied. Everyone seems to be blaming [various platforms] for spreading misinformation and letting people be misled. Well, there's no regulation saying that they have to take it down, and there are no guidelines on what they should or shouldn't take down. No one is telling them this and they have to figure out how to navigate that whole space by themselves. […] So, ok, then the onus is on the consumer to just deal with it – 4807_ArtificialIntelligence.

4.3 (Anything but) ethics training

While virtues can be understood as moral habits, they are not merely the result of repetitive actions. We already pointed out that a necessary condition of behaving virtuously is ethical sensitivity, or being aware that one's practice domain is ethically charged. But the practitioner must also be able to navigate this terrain with skill and know when and where to seek advice. This takes discernment, knowledge, and reflection, or what Welie et al. [36] called "moral competencies". As is true of all competencies, moral competencies are not simply acquired; they generally require education and training. We thus asked whether developers had been exposed to any training or education in ethics, and further whether they thought they ought to be exposed to it.

When asked about previous exposure to ethics training, eight (8) said they pursued ethics education on their own. This typically took the form of independent reading or participation in ethics discussions at industry conferences. Twelve (12) said they received some ethics training as part of their university education, but only one of these said it was a course solely devoted to ethics. The remaining respondents

Table 2 Ethical dilemmas and ethical distress experienced by AI/ML/DS developers

Dilemma	**ID***	**Seeking Help**
"I can't comfortably talk about my robotics research. When I wrote a grant proposal it was tricky to find ways to make my argument. I had to find ways to not sound like robots are going to take away jobs but compliment the workforce."	1855 RO	"I didn't really go anywhere to get help or advice on solving the discomfort. The only advice I would get is how to make my proposal sound like it's a good thing. Really, the incentive is what are you inventing and how many publications you have. So, I didn't seek advice on how to navigate the discomfort."
"There are tech companies that cannot collect race data, because why would you? You don't need it to deliver a website, right? And [now] we're trying to do fairness work so that we can make sure our product is equitable for all races, but [we] don't have that data, so what happens then? Should I be inferring race?"	4970 RS	"We have policy people […] [who work with] AI researchers from philosophy, ethics, and social science backgrounds […] [Their] guidance documents are what we follow internally. Now, we can challenge this. There is a healthy debate internally, but if you are rushed, you will just follow whatever the policy guidance says. So, we basically rely on those teams being good."
Distress	**ID**	**Seeking Help**
"I definitely had ethical questions about some projects. In most cases, it still feels like it wasn't really in my hands. I can circle around people and jump and wave my arms, but in the end, nothing was done and I [was] seen as adversarial. And then I changed teams."	7433 DL	"There are few places you can go for help. Some things you have to escalate to the top level [to] force people on the product side to at least give you access so you can show there's a problem. It's very difficult to raise an ethics question without having any kind of proof; it's very difficult to have proof without access, and all access is regulated."
"Most recently, there was an ethical issue that I was pretty sure what we needed to do, but due to time constraints from the company and commitments we had made to customers, I wasn't allowed to do those things. So, it became less of a 'Let's make sure this is ethically in the right direction,' and more like, 'Let's see how we can make this look good on paper and still release it on time.'"	44377 DS	"I complained about it to coworkers. It's hard to know where to go. I feel like, if I talked to some media person as a whistleblower then this company is going to sue me for violation of an NDA. If I try to talk to people higher up in the company, I'm going to be ridiculed and/or fired. It just doesn't seem like there's a lot of room for recourse."
"I was in this credit thing where you put up a model and it tells you how much money you can [lend] to a specific person. The bank […] their business was to give credit cards to people who had no previous credit experience. So, you have zero data on [a person's] credit behavior. What you had was just demographics. I had a lot of doubts about it because you could see if you actually analyze the data, it loaned a lot less to women than to men."	7015 AI	"I went to my boss. I was like, 'Look, I think this might be weird.' But she said, 'Ah, it is what it is.' So, there was no answer for it […] you're not supposed to say that this is happening with the model. You just say, "It's the algorithm." <*lifts hands in surrender*>

*ML = Machine Learning; DS = Data Science; RO = Robotics; AI = Artificial Intelligence; DL = Deep Learning; RS = Research Science. Participant IDs were assigned using a random number generator

in this category said their exposure was limited to compliance topics like research ethics, data security, and privacy training. Sixteen (16) said they received ethics training at work, but again the training was limited to research integrity ethics or compliance about data security and privacy. One (1) was not asked the question. The remaining eleven (11) said that they had received no ethics training at all. Eight (8) respondents gave multiple answers to this question.

While most participants in this study claimed to have limited or no exposure to ethics training, they all ($n = 38$; 2 not asked) agreed they ought to be exposed to it. Yet, a close evaluation of the transcripts reveal that the labels "ethics" and "training" were doing something developers felt was important, but it was also something they did not like. The following verbatims are illustrative:

> I'm smiling because these are the types of classes that you would have […] sloughed off. You're taking these advanced math and computer science courses and then one day they tell you, now you need to sit in a room and learn about ethics, and a lot of the people that I was in classes with would have blown off that class – 2353_MachineLearning.
>
> Training is something I think has negative connotations for a lot of people. […] For the compliance trainings, I don't think it sticks and I don't think it engenders a

> passion for it in the people who go through the training – 73057_DataScience.

The prevailing preference among all participants was for ethics to be woven into daily practice, ideally with the assistance of domain experts, rather than as separate training courses. For example:

> Whatever ethical considerations that come to bear should be reflected in the day to day. It definitely shouldn't be an ongoing series of trainings the whole year, because that leads to people considering things to be irrelevant – 7591_DataScience.
>
> In an ideal world, it would be a collaboration. Ideally, that collaboration would start very early in the game. More often, it comes in quite late, when the product is almost ready – 7433_DeepLearning.

4.4 Barriers to ethical wisdom

Although we did not specifically ask about barriers to gaining and applying ethical wisdom, participants nonetheless would often speak about what was impeding their own ethical development. Three barriers were most prevalent. First was speed-to-market, or an "innovation first, ethics second" orientation that prioritized technology creation before considering ethical implications. For example, participant 1855_Robotics advocated to "focus initially on advancing the science, and once it does something well, then we can re-engineer it to make it more ethical." But the pressure of moving quickly also hindered ethical considerations, as participant 2032_DataScience pointed out that "there is no time to think of ethics […] the only thing people think of is, 'Are we abiding by [relevant law], yes or no?' If it's a yes, then it's a mad race."

The second barrier was *limited perspectives*, whether technical, structural, or ideological in nature. On the technical side, developers admitted that just getting a system to work at all took up a great deal of time. "It's difficult to apply ethics when you're struggling for six hours to get an [application package running]," said participant 13479_MachineLearning, for example. Similarly, participant 2760_Robotics said that developers do think of themselves has having ethical agency, but "the vast majority of our headspace is taken up by just getting anything to work, any of the time."

More structurally, developers noted that the way in which their training happens does not often leave room for ethical reflection. Participant 4807_ArtificialIntelligence, for example, noted that "trying to be someone who is actively thinking about ethics while plying their trade as computer scientist is extra work. It requires you to be more aware and to engage another muscle in thinking than you normally would given your training from undergrad." This was echoed by participant 2353_MachineLearning, who said, "There's no effort to try to figure out who you are as a person and to see what the history of your usage of this product is. That's not really something that is prevalent in my industry."

Ideologically, there can be an over-reliance on the purported liberating powers of data and automation. This is revealed in, for example, participant 3841_DataScience's observations that some developers could be "completely blinded" by the belief that they are "replicating the truth" that exists inherently the data. This barrier also showed up in what participant 7433_DeepLearning called a "protective-defensive type of attitude" in which some developers believe "it is for the greater good that the model be out there […] and anything that's slowing that down is like, 'Is it ethical to keep this internally much longer when it could be out in the world to save lives?'".

The most common barrier to ethical development, however, was *misaligned incentives*. For many AI developers, incentives are heavily weighted in favor of corporate reputation and profit. For example, participant 7671_ArtificialIntelligence noted that often the "only perspective that really mattered" was whether a given decision could make the company look bad in the press. Paradoxically, one participant also lamented that bolstering corporate reputation as a competitive edge often meant overstating what an AI system was capable of:

> I think most people who haven't drunk their own Kool-Aid or aren't mad on power are realistic about [AI] and a little bit bewildered by what's happening and are trying to find ways to make the conversation more sane. But you can't, because you just don't get picked up [in the press] if you say sane things – 2760_Robotics.

5 Discussion

Aristotle argued that practical wisdom is a combination of ethical will and ethical skill [14]. Previous research has shown that AI developers have a great deal of ethical agency, as well as the will and license to exercise this agency in the service of what they think is right [30, 32]. What they lack is domain-specific ethical skill. To be clear, we are not suggesting that AI developers are unethical, only that their education and training does not adequately prepare them to manage the ethical import of their work. As this study has shown, most developers are aware of the ethical territories they are being asked to navigate and of the specific moral quandaries that arise in their line of work. The ethical issues they spoke about correspond with those frequently discussed in the scholarly literature on AI (see, for example, [37–40]). Yet, the resources available to help them to respond in an ethically sound manner are limited and inconsistent.

In the absence of concrete rules or robust guidance, developers most often seek help from each other. Moreover, this is where they would prefer that ethics conversations occur. This would not be a problem if there were skilled and wise help available within the community, but this research suggests that this is rarely the case. Still, developers are open to accepting guidance from subject matter experts. They prefer this guidance to be an embedded part of their practice, which is largely in keeping with how they gained their technical expertise – as hands-on learning.

This research also shows significant barriers are working against the development of practical wisdom in the AI developer community. These include the near sacred status the industry gives to innovation, the myopia that can occur in technical practice, few provisions for reflection and dialogue, and incentive structures that value profits and reputation above all else. Collectively, the effect of these barriers may be why cynical stereotypes about developers persist. As Frey [41] rightly states, "Character emerges from surroundings and issues back into them."

There may be a helpful distinction to be made here between "character" understood as a person's moral temperament and what prominent virtue ethicist Alasdair MacIntyre [42] called *characters*, or occupational roles that become "the moral representatives of their culture because of the way in which moral and metaphysical ideas and theories assume through them an embodied existence in the social world" (pp. 32–33). MacIntyre argues that what defined, for example, the moral culture of Victorian England was in part the *characters* of the Explorer and the Engineer. Individuals filled these roles, yes, but in doing so they became the physical manifestation of the moral and metaphysical beliefs the English had about themselves at the time. It may likewise be true that the morality of the current age is, in part, embodied by the *character* of the AI Developer.

For better or worse, AI developers are presently one of the leading moral representatives of a techno-optimist and techno-determinist culture. The automated systems they build come with an almost mythological promise to accelerate human well-being, ushering in, at last, a future of health and leisure for everyone. The *character* of the AI Developer is an embodiment of this moral ideal: eternally young, brilliant, and confident, with vast wealth and prestige at their disposal. This myth persists even as AI systems fail to deliver on their existential promises. It cannot be lost on us that humans continue to do the monotonous, often traumatic, and low-paying labor of data cleaning and classifying so that the machines can now write poetry and create art [43].

Individual developers need not like or embrace this *character* to nonetheless be subject to the cultural beliefs they are embodying. As MacIntyre notes, a priest who has lost his faith is still giving physical manifestation to the beliefs of Catholic Christianity when he officiates mass (pp 33–34). We find no evidence in this research that the move-fast-and-break-things tech bro culture is the ideal developers look up to. And there is certainly no evidence that developers ought to live this way to be a good developer and live a morally sound professional life. If anything, this *character* poses the most significant hinderance to clarifying either *what it is good to do* or *who it is good to be* as an AI developer.

6 Limitations

This paper relies on qualitative findings from semi-structured interviews with developers recruited worldwide. As noted in previous literature using this data set [32], certain limitations must be noted. First, the sample size of 40 is small for a global study, even if saturation suggests our findings are representative. Second, we relied on developers who voluntarily agreed to be interviewed about ethics in their field, which means that our findings may be subject to selection bias. More research should be conducted within geographic regions, industries, and organizations to better reveal how virtue ethics can be elevated in AI developers' practice.

7 Recommendations and conclusion

Who AI developers are while they are developing automated systems cannot be divorced from their circumstances or the people around them. The end of "ethical AI" is not theirs alone to achieve. It requires a culture of ethical awareness and reflection, lived by everyone involved in the development, deployment, use, and maintenance of automated systems. We return, therefore, to the problem of many hands.

As we mentioned in Sect. 2.0, dozens or hundreds of developers may each be doing something morally negligible, yet the sum of all their actions carries heavy moral weight. Moreover, automated systems are in an almost constant state of being inherited by other developers, who regularly update the systems and change their behavior. Under these circumstances, locating responsibility becomes much more difficult. Indeed, a recent study by Widder and Nafus [44] concluded that the modular nature of software development isolates developers from the impact of their work, which encourages them to think of themselves simply as part of the supply chain rather than as active agents engaged in work "where a deep collaborative relationship might develop." We postulate that legal liability for harms caused by automated systems will (appropriately) continue to fall to the institutions deploying them. Yet, the fact that we cannot locate moral responsibility on any one developer is not a good enough reason to leave the developer community out of our discussions about their ethical responsibility or accountability.

If we are to succeed at ethical AI, we cannot afford to ignore the ethical agency or ethical development of the practitioners we task with building it [32]. This means developers cannot simply be "magicians" hired to construct the technical systems that will ultimately save us. They must be engaged and fully visible as members of the community of beneficiaries the system is purported to help. If they are not working in silos, but as co-creators of systems with clear ends for the community to which they also belong, "many hands" may become a help rather than a problem.

We make three recommendations. First, there is a real danger of stifling practical wisdom if business and academia reduce ethics to compliance. If the goal of automation is efficiency, then as ethics scales in an organization it too will be subject to the efficiency gains promised by automation. It will become "e-learning" modules or annual compliance trainings. Virtue ethics in these circumstances ceases to be an embodied practice; it does not "engender passion" in the subject, as one participant in this study said. Compliance can be effective in motivating people to abide by the rules, but wisdom is about navigating spaces where the rules do not apply in a simple way. Organizations, governments, and academic institutions that are building or deploying automated systems need to cultivate an environment that normalizes ethics discussions with developers. They can do this, in part, by embedding subject matter experts (ethicists, sociologists, psychologists, etc.) and other rights-holders into the lab environment. Ethics involves critically reflective practices in a communal context. Ethics cannot simply be a box that gets checked on a legal form.

This, then, leads us to our other two recommendations, one concerning competencies, and the other the AI developer community. Educators of AI developers, whether in formal university settings, online academies, or elsewhere, have a responsibility to accelerate the process of guiding learners out of non-wisdom towards wisdom. This is an acute issue because some evidence indicates that engineering education makes students *less* sensitive to ethical issues [45]. Developers completing these programs need to emerge more ethically competent than they currently are. Education programs minimally need to integrate ethical reasoning into all computer science courses. This should not be a single course on ethics in the curriculum. Rather, every module offered in these programs should include practice-based education in which the developers engage with complicated ethical questions. There are already some studies attesting to the efficacy of this approach [46, 47]. Additionally, some universities and other institutions are experimenting with extensive integration of ethics into engineering education [48, 48]. Some have also established institutes that provide resources and education to other universities, developers in industry, and to organizations (for example, [50]).

Finally, there is the community of AI developers. At present, AI developers do not constitute an organized profession (in the normative sense of that term) so much as a body of experts working in industries that have different vested interests. There may be cause to formally professionalize and license the practice, as others have suggested [4, 51–53]. Professionalization could offer important protections against the whistleblower's distress we discussed earlier in this paper. It could also allow a unified developer community to act as a check against other powerful interests. But this will require that developers organize and agree to have their professional status governed by a licensing board. The full implications of this are beyond the scope of this paper and should be more fully studied.

The body politic will continue to demand that automated systems are fair, responsible, just, and so on. Developers of these systems must therefore become more competent in fairness, responsibility, and justice if they are going to succeed in building systems that reflect those virtues. It is therefore essential to complement the work underway to achieve 'ethical AI' with the cultivation of virtues in developer education and occupational practice.

References

1. Littman, M.L., et al.: Gathering Strength, Gathering Storms: The One Hundred year Study on Artificial Intelligence (AI100) 2021 Study Panel Report. Stanford University (2021)
2. Acemoglu, D.: Harms of AI. National Bureau of Economic Research (2021)

3. Ryan, M., Stahl, B.C.: Artificial intelligence ethics guidelines for developers and users: Clarifying their content and normative implications. J. Inform. Communication Ethics Soc. **19**(1), 61–86 (2021)
4. Mittelstadt, B.: Principles alone cannot guarantee ethical AI. Nat. Mach. Intell. **1**(11), 501–507 (2019)
5. Munn, L.: *The uselessness of AI ethics* AI and Ethics, : p. 1–9. (2022)
6. Garg, P., Villasenor, J., Foggo, V.: *Fairness metrics: A comparative analysis*. in *IEEE International Conference on Big Data (Big Data)*. 2020. IEEE. (2020)
7. Mitcham, C.: A philosophical inadequacy of engineering. Monist. **92**(3), 339–356 (2009)
8. Meyer, M., Rego, A.: *Measuring practical wisdom: Exploring the value of Aristotle's phronesis for business and leadership* Handbook of practical wisdom in business and management, : p. 1–18. (2020)
9. Swartwood, J.: Can we measure practical wisdom? J. Moral. Educ. **49**(1), 71–97 (2020)
10. Laas, K., Davis, M., Hildt, E.: Codes of Ethics and Ethical Guidelines. Springer (2022)
11. Gotterbarn, D., et al.: *ACM code of ethics and professional conduct* (2018)
12. Pugh, E.W.: *Creating the ieee code of ethics*. in *IEEE Conference on the History of Technical Societies*. 2009. IEEE. (2009)
13. Hursthouse, R., Pettigrove, G.: Virtue Ethics in Stanford Encyclopedia of Philosophy. Metaphysics Research Lab (2007)
14. Aristotle, D., Ross, Brown, L.: The Nicomachean Ethics. Oxford University Press, USA (2009)
15. Keown, D.: The Nature of Buddhist Ethics. springer (2016)
16. Harris, C.E.: The good engineer: Giving virtue its due in engineering ethics. Sci Eng. Ethics. **14**, 153–164 (2008)
17. Hagendorff, T.: *AI ethics and its pitfalls: not living up to its own standards?* AI and Ethics, : p. 1–8. (2022)
18. Hagendorff, T.: A virtue-based framework to support putting AI ethics into practice. Philos. Technol. **35**(3), 55 (2022)
19. Barford, L.: *Contemporary virtue ethics and the engineers of autonomous systems*. in *2019 IEEE International Symposium on Technology and Society (ISTAS)*. IEEE. (2019)
20. Vallor, S.: Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting. Oxford University Press (2016)
21. Floridi, L.: The Cambridge Handbook of Information and Computer Ethics. Cambridge University Press: Cambridge; (2010)
22. Johnson, D.G.: *Computer Ethics*. Third Edition ed. New Jersey: Prentice-Hall. 240. (2001)
23. Fleddermann, C.B.: Engineering Ethics. Erlangga (2004)
24. Thompson, C.: Coders: The Making of a new Tribe and the Remaking of the World. Penguin, New York (2019)
25. Miller, M.E.: *SF 'Tech Bro'Writes Open Letter to Mayor:'I Shouldn't Have to See the Pain, Struggle, and Despair of Homeless People,''Riff Raff.'* Washington Post, (2016)
26. Goldstaub, T.: The Dangers of tech-bro AI. MIT Technology Review, November/December, np (2017)
27. Broome, T.H. Jr., Peirce, J.: The heroic engineer. J. Eng. Educ. **86**(1), 51–55 (1997)
28. Basart, J.M., Serra, M.: Engineering ethics beyond engineers' ethics. Sci Eng. Ethics. **19**, 179–187 (2013)
29. Nissenbaum, H.: Accountability in a computerized society. Sci Eng. Ethics. **2**(1), 25–42 (1996)
30. Orr, W., Davis, J.L.: Attributions of ethical responsibility by Artificial Intelligence practitioners. Communication Soc. **23**(5), 719–735 (2020). Information
31. Cobern, W.W.: Constructivism J. Educational Psychol. Consultation. **4**(1), 105–112 (1993)
32. Griffin, T.A., B.P. Green, and J.V.M. Welie, The ethical agency of AI developers. AI and Ethics, 2023.
33. Shaub, M.K., Finn, D.W., Munter, P.: The effects of auditors' ethical orientation on commitment and ethical sensitivity. Behav. Res. Account. **5**(1), 145–169 (1993)
34. Volker, J.M.: Counseling Experience, Moral Judgment, Awareness of Consequences and Moral Sensitivity in Counseling Practice. University of Minnesota (1984)
35. Bebeau, M.J., Rest, J.R., Yamoor, C.M.: Measuring dental students' ethical sensitivity. J. Dent. Educ. **49**(4), 225–235 (1985)
36. Welie, J.V. and J.T. Rule, Overcoming isolationism. Moral competencies, virtues and the importance of connectedness. Justice in oral health care. Ethic. Educ. Perspect. 97–125 (2006)
37. Jobin, A., Ienca, M., Vayena, E.: The global landscape of AI ethics guidelines. Nat. Mach. Intell. **1**(9), 389–399 (2019)
38. Floridi, L., et al.: AI4People—An ethical Framework for a good AI society: Opportunities, risks, principles, and recommendations. Mind. Mach. **28**(4), 689–707 (2018)
39. Zhou, J., Chen, F.: AI Ethics: From Principles to Practice, vol. 38, pp. 2693–2703. AI & SOCIETY (2023). 6
40. Kamila, M.K., Jasrotia, S.S.: Ethical issues in the development of artificial intelligence: Recognizing the risks. Int. J. Ethics Syst., (2023)
41. Frey, W.J.: Teaching virtue: Pedagogical implications of moral psychology. Sci Eng. Ethics. **16**, 611–628 (2010)
42. MacIntyre, A.: After Virtue. A&C Black (2013)
43. Hao, K.: and A.P. Hernandez *How the AI Industry Profits from Catastrophe*. MIT Technology Review (2022)
44. Widder, D.G., Nafus, D.: Dislocated accountabilities in the AI supply chain: Modularity and developers' notions of responsibility. Big Data Soc. **10**(1), 20539517231177620 (2023)
45. Cech, E.A.: Culture of disengagement in engineering education? Sci. Technol. Hum. Values. **39**(1), 42–72 (2014)
46. O'Sullivan, D., et al.: *Ethics4eu: Designing New Curricula For Computer Science Ethics Education: Case Studies For Ai Ethics* (2023)
47. Vakkuri, V., Kemell, K.-K.: Implementing AI Ethics in Practice: An Empirical Evaluation of the RESOLVEDD Strategy. Springer International Publishing, Cham (2019)
48. Univeristy, S.C.: *Engineering and the Good Life*. ; Available from: (2024). https://www.scu.edu/engineering/academic-programs/engineering-and-the-good-life/
49. Gaudet, M.J.: *A discernment model for teaching tech ethics*. in *Annual Meeting for the Society of Christian Ethics*. Virtual Meeting. (2021)
50. Green, B.P., et al.: A University Applied Ethics Center: The Markkula Center for Applied Ethics at Santa Clara University. J. Moral Theol. **9**(Special Issue 2), 209–228 (2020)
51. Green, B.P. Are science, technology, and engineering now the most important subjects for ethics? Our need to respond. In 2014 IEEE International Symposium on Ethics in Science, Technology and Engineering. 2014.
52. Green, B.P. Emerging technologies, catastrophic risks, and ethics: three strategies for reducing risk. In 2016 IEEE International Symposium on Ethics in Engineering, Science and Technology (ETHICS). 2016.
53. Strümke, I., Slavkovik, M., Madai, V.I.: The social dilemma in artificial intelligence development and why we have to solve it. AI Ethics. **2**(4), 655–665 (2022)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.