AI and Ethics
https://doi.org/10.1007/s43681-024-00472-z

ORIGINAL RESEARCH

From applied ethics and ethical principles to virtue and narrative in AI practices

Paul Hayes[1] · Noel Fitzpatrick[1] · José Manuel Ferrández[2]

Received: 12 October 2023 / Accepted: 19 March 2024

Abstract
The question of how we can use ethics and ethical frameworks to avert the negative consequences of AI through guidance on human behaviour and the design of technological systems has recently been receiving increasing attention. The appropriate response to an ethics of AI has certainly been contentious. For some years the wisdom of deontology and utilitarianism in the ethics of technology has been questioned. Today, a kind of AI ethics principlism has gained a degree of widespread acceptance, yet it still invites harsh rejections in recent scholarship. In this paper, we wish to explore the contribution to an ethics of AI made by a narrative philosophy and ethics of technology inspired by the 'little ethics' of Paul Ricoeur, and virtue ethics of Alasdair MacIntyre, most recently and promisingly built upon by Wessel Reijers and Mark Coeckelbergh. The objective of this paper is to examine the extent to which a narrative and virtue based ethics (or, VPD, i.e., virtuous practice design) might be a plausible candidate for the foundation of an ethics of AI, or rather ethical AI practice. This will be achieved by exploring the ways in which this approach can respond to some of the significant faults with or critiques of applied and principles and guidelines based ethical approaches to AI ethics.

Keywords Ricoeur · Narrative · Virtues · AI · Ethics · Ethical framework · Practice

1 Introduction

Public discussion of the benefits and risks of artificial intelligence (AI) has arguably accelerated in recent months and years as the capabilities of generative AI applications have captured the public's imagination (such as large language models like ChatGPT in particular).[1] The everyday capabilities of AI applications and the potential ethical, often relatable, dilemmas they can pose (imagine the temptation of a student to plagiarize their essay using ChatGPT), are seemingly becoming more and more apparent to increasingly broader groups of people, and no longer just professionals, academics or AI enthusiasts [21, 23, 69, 100]. The question of how we can use ethics and ethical frameworks to avert the negative consequences of AI through guidance on human behaviour and the design of technological systems has also recently been receiving increasing attention [51]. Such scholarship and public debate is likely to continue intensifying as real and practical concerns about AI become more and more apparent to and understood by the general public, whose jobs and hobbies (their *practices* as such) stand to be transformed as AI applications become capable of increasingly complex and creative tasks.

The appropriate response to an ethics of AI has certainly been contentious. For some years the wisdom of deontology and utilitarianism in the ethics of technology has been questioned. Today, a kind of AI ethics principlism has gained a degree of widespread acceptance, yet it still invites harsh rejections in recent scholarship, as we will explore in what follows [33, 46, 63, 64]. In this paper, we wish to explore the

[1] Indeed, as a point of interest Google trends (https://trends.google.com/trends/explore?date=all&q=AI&hl=en) shows interest in the 'AI' search term rising exponentially beginning around late 2022.

✉ Paul Hayes
paul.hayes@tudublin.ie

Noel Fitzpatrick
Noel.fitzpatrick@tudublin.ie

José Manuel Ferrández
jm.ferrandez@upct.es

[1] TU Dublin, GradCAM, ECT Lab+, SFI ADAPT Research Centre, Dublin, Ireland

[2] Departamento de Electrónica, Tecnología de Computadores y Proyectos; ECT Lab+, Universidad Polit?cnica de Cartagena, Cartagena, Spain

Published online: 08 April 2024

contribution to an ethics of AI made by a narrative philosophy and ethics of technology inspired by the 'little ethics' of Paul Ricoeur, and virtue ethics (VE) of Alasdair MacIntyre, most recently and promisingly built upon by Wessel Reijers and Mark Coeckelbergh [75]. Reijers and Coeckelbergh add to what could be considered a growing chorus of scholars advocating for a responsible research and innovation paradigm built around a virtue ethics that champions attention to the particular, and not a nebulous constellation of ethical principles of potentially indeterminate applicability, or other applied methods of technological ethics that are divorced from the kinds of normative foundations necessary for their legitimacy.

The objective of this paper is to examine the extent to which a narrative and virtue based ethics (or, VPD – virtuous practice design) might be a plausible candidate for the foundation of an ethics of AI, or rather ethical AI practice. This will be achieved by exploring the ways in which this approach can respond to some of the significant faults with or critiques of applied and principles and guidelines based approaches to AI ethics. The paper is structured as follows. Section 2 will provide an overview of some pointed criticisms that have been applied to ethical theories more generally as well as some applied methods including value sensitive design (VSD), and more pressingly to AI ethics and AI ethics principles and guidelines. Section 3 will provide a very brief overview of Ricoeur's 'little ethics', and virtue ethics, leading to a description of advances made under the VPD based approach. Section 4 will then examine the contributions of a VPD based approach to an ethics of AI based on existing criticisms of other approaches. Section 5 will provide a brief reflection on the nature of Foucault's *dispositif* in admission that whilst VPD provides valuable tools in the quest for an ethics of AI, any such quest will always be fought against pre-existing structures and systems of power, influence and control which present a challenge to change. Section 6 concludes.

In the following paper we adopt a socio-technical view of artificial intelligence which is to say we do not understand AI merely as mathematical artefacts applied to particular problems in a vacuum, but rather as something which emerges from multiple interconnected practices including extractive ones, data science, marketing, and use cases where they can structure relations between things and people. AI forms through processes of interconnected practice, and is subsequently adopted by users in new and transforming practices—they form parts of socio-technical systems which are moved by the relations of their constituents and their background of laws, regulations, and norms [17, 24]. As Kate Crawford puts it '[a]t a fundamental level, AI is technical and social practices, institutions and infrastructures, politics and culture' and '[…] artificial intelligence is both embodied and material, made from natural resources, fuel, human labor, infrastructures, logistics, histories, and classifications' [26, p.8]. When we understand AI in this way as part of a socio-technical system, we can also understand better why it is not only important to design the AI as an artefact in an ethical and responsible way, but also to design the environments in which they are deployed to support both this ethical design and deployment within those environments.

In summary, this paper acknowledges the flaws or shortcomings of current approaches to AI ethics and the design of ethical AI systems, lays out significant existing critiques of current approaches to the ethics of technology and AI that have been raised in recent scholarship, and examines the potential of a narrative and technology ethics (VPD) to constructively address some of these challenges to the ethical design, implementation, and deployment of AI tools and systems. An ethical framework that is concerned with aiming at the good life, with and for others, and in just institutions, is a comprehensive one that asks us as ethical agents to consider what it is that is worth aiming at to be ethically accomplished individuals, how we do this in a way that respects and elevates others, and crucially, how this can be done within global systems that support and mediate ethical action. In asking us to consider the small narratives (self) to the grand ones (society), it is an ethical framework that not only asks us to narrate and interpret ourselves as ethical individuals but asks how we can build just institutions—indeed, when thinking in terms of AI, how we might build AI systems for the common good and how we might design legal, regulatory, and normative environments that support the design of the socio-technical system of AI itself.

VPD, as a development of Ricoeur's narrative philosophy and 'little ethics' is a new approach to the ethics of technology and one little experimented with thus far in real life AI use cases, to the best of our knowledge. As an emerging approach to the philosophy and ethics of technology, this approach has not yet been aimed at a multiplicity of issues in a systematic way, and to the best of our knowledge this paper represents the first effort to apply VPD in an in-depth manner to the problem of the ethical design of artificial intelligence and AI practices.[2] In this regard, this paper builds on the work presented by Reijers and Coeckelbergh in demonstrating the potential use of the approach in a specific and critical domain whilst also making links to other approaches (such as participatory design) that can complement and support it.

It should be noted that virtue ethics alone has historically enjoyed much attention in business ethics more widely [6,

[2] Although, there have been other efforts to unpack further what narrative theory means for AI in relation to responsibility and transparency for example [24, 25, 39].

13, 71, 92], with the recognised benefit of instilling practitioners with the moral knowledge and responsive aptitude (or *phronesis*) to recognise the morally salient features of situations and to respond to them with integrity, also possessing the knowledge to correctly apply ethical principles to such situations. The reader will recognise VPD as a framework significantly more complex than standard VE due to its incorporation of elements of other theories, which is also what makes it an interesting and promising candidate to support AI ethics, due to representing a more unified approach to ethics than usually seen.

2 Some limitations of current approaches to the ethics of technology and AI

Numerous critiques have been launched both against the suitability of the dominant ethical theories (consequentialism and deontology for example) and at a more applied level, the use of ethical principles or even methodological frameworks such as value sensitive design (VSD) to guide the design, implementation, and deployment (as well as general use and adoption) of AI systems or digital and novel technologies more broadly. These will be highlighted in what follows. The arguments here are not wholly or all endorsed by the present authors but will be presented to later demonstrate that the kind of narrative theory established by Paul Ricoeur and Alasdair MacIntyre (and later elaborated by Wessel Reijers and Mark Coeckelbergh for the domain of the ethics of technology) can constructively respond to at least some perceived failings or limitations of the ethics of technology. Indeed, in what follows we will also highlight the beneficial use of deontology and ethical principles to the extent that they can form overridable elements of a framework primarily rooted in virtue ethics. In the remainder of this section, some critiques of dominant ethical approaches in the ethics of technology will be overviewed.

Shannon Vallor provides brief but pointed criticism of deontology (of Kant, more specifically), and utilitarianism, questioning the applicability of the abstract and general categorical imperative to the plain of technology, where an agent's will (the universal legislator) cannot easily be informed against a backdrop of the unpredictable course followed by technological development and the use and adoption of new technologies [105]. Moral dilemmas may also arise that result in deontological rules coming into conflict, which necessitates an effective decision procedure to resolve the dilemma—rules may not be so easy to follow or choices clear, frustrated by the inherent opacity of the technological future [8, 17, 18, 105]. We cannot predict the dilemmas or challenges of tomorrow or easily devise rules that protect us against them and adequately guide our actions—this problem itself is consistent with the Collingridge dilemma whereby we are least well-placed to influence a technology when its actual impacts become clear [17]. That a strict or conservative interpretation of deontology may militate against considering impacts at all may also lead to myopic approaches to technology and technical practice design, and indeed such interpretations of deontology may also place undue emphasis on adherence to a rule without compromise and without regard for consequences—although such a dogmatic account may be an outlying one, especially when one considers developments in, for example, more intuitionist forms of deontology [8, 12, 89].

Vallor also levels similar criticism at utilitarianism, as the course of action leading to the greatest happiness (or other values, even plural, such as welfare, etc. in forms of consequentialism [17]) is incalculable in the face of the many unknown (even converging) technological and technosocial possibilities [105, pp. 7-8]. Beyond its applicability to the dynamic and highly scalable technological context, it has also been argued that utilitarianism is '[…] a framework [that] privileges the happiness and welfare of the majority, and without some refinement, can undermine the welfare of the marginalised and is indifferent to their lived experience' [40, p. 159]. For strict accounts of utilitarianism, where the happiness of each individual is measured the same as the next (where the best action is that which results in the greatest happiness for the greatest number), the needs of the few (potentially already marginalised) may quickly be overshadowed by the needs of the many in ways that do not accord with our intuitions about fairness. Prioritarian accounts of consequentialism (favouring the needs of the most disadvantaged), such as that of Derek Parfit can mitigate such dangers, but this does not answer to the difficulties of applying the framework to opaque technological futures [70].

On both accounts, Jiin-Yu Chen adds that '[t]hese theories developed abstract, timeless, and universal ways of approaching problems, at the expense of attending to their particular, concrete, and timely details' [22, p. 76]. This critique points to the core of issues with these frameworks, which is that they are arguably inadequately attuned to engaging contextual nuances of different moral situations and peoples' lived experiences across time and space.

One example one might give of the limits of these frameworks is the evolution of recommender/curation algorithms on social media which have presented numerous ethical challenges and quandaries from the perspective of design and regulation. At inception, the exact risks of designing algorithms to help connect individuals with content and other people relevant to their interests may not have been very clear. In most cases, matching people with relevant news items or blogs may have not seemed to challenge established principles or rules, and the net happiness increase from such systems may have appeared to justify their creation. As time went by, we have witnessed their capacity to create echo-chambers, filter-bubbles, polarization, and their contribution to the spread of mis- or disinformation that has undermined trust in public health information, and that has resulted in discrimination against minority groups [50, 99, 103, 108]. What may have seemed mostly innocuous at first to its original architects has evolved in a path where it has contributed to significant political and social issues, and real-world harms have occurred as a result of the misinformation and bad actors with whom people have been connected [2]. These consequences may have been rather unknown or unforeseeable before they emerged in more recent years, and difficult to plan for both for their creators, and society more largely. Now, challenges and dilemmas have arisen between facilitating and promoting access to content, and freedom of expression and to impart and receive information, and serving the interests of public health and preventing hate crimes, for example. These frameworks may also not inculcate the greatest sense of responsibility in technology developers either, when they may be able to argue that the persons primarily responsible for toxic content are not those who create the infrastructure through which in propagates, but rather those who create the content in the first place.

Value sensitive design (VSD) is an example of a methodology developed with the explicit purpose of building ethics into the development of new technologies or technology projects. VSD is a tripartite methodology consisting of conceptual, technical and empirical investigations of technical artefacts used to uncover the moral values relevant to the context of use of a technology with a view towards translating relevant norms into design requirements that support the expression of moral values in the use of the technology [36, 41, 72]. Wessel Reijers and Bert Gordijn have criticised the perceived shortcomings of VSD in suggesting their own framework (VPD) [76]. According to Reijers and Gordijn, VSD is limited by not being anchored into a specific normative theory, the arguable arbitrary selection of relevant values it entails, and its narrow concern with technological design rather than wider consideration of circumstances concerning technical practice [76]. The absence of a clear ethical (normative) framework in VSD is a significant critique—whilst this approach targets embedding some sense of 'the good' into design through its focus on values, it makes no direct commitment to a vision of the good, it is agnostic, and without such an anchor it can be difficult to justify design decisions as being 'ethical' as they are not fundamentally derived from an ethical framework [41]. It can be said that VSD provides valuable methods for investigating and implementing vital value considerations in technology design in its tripartite framework which does require thorough engagement with technical questions, as well philosophical and societal ones, in the development of new technologies. Recent scholarship in particular has demonstrated its flexibility and how it is compatible with continuous application to the design of artefacts which warrant continued reflection on their value implications—for example the concept of value change has been identified as being important to VSD, which points to useful strategies for adaptation to change, including solutions such as modularization [73]. Nevertheless, despite some scholars in VSD demonstrating concern for the design of socio-technical systems [17, 42], i.e., the whole ecosystem of artefact, person, norms, and political, economic and legal systems, it remains fair to say the approach is more narrowly concerned with and capable of translating values and norms into design of the artefact and not the system the artefact is embedded in.

Other methodologies also lack clear guidance on the use of normative theory, including ethical impact assessments, which tend not to be very prescriptive in the selection of a specific normative framework for making evaluations about the ethical impact of investigated technologies [78, 109]. Again, impact assessments tend to be narrowly focused on matters of design and use (mitigation interventions) of specific technologies and can be variable in their breadth and inclusion of diverse stakeholders in the process. The impact assessment can still be a useful tool, but must be one part of a comprehensive ethical approach and not the only part of one.

With the emergence of AI principles endorsed by international organisations such as the European Commission (e.g., *Ethics Guidelines for Trustworthy AI*), and the OECD,[3] as well as a general groundswell in academic and organisational interest in proposing and adopting different AI principles and guidelines [46], such principles are also beginning to court attention and critique, in some cases critiques extending to the whole endeavour of AI ethics itself [63, 64]. Brent Mittelstadt persuasively argues that there are four challenges to a principles based approach to AI ethics, which he compares to the otherwise successful use of a principled approach in medicine (indeed an exemplary work within this region which has clearly inspired many such AI principles is Beauchamp and Childress' *Principles of Biomedical Ethics*) [14, 63]. According to Mittlestadt the reasons a principled approach to AI ethics is fraught with challenges are (as briefly summarised) [63]:

1. **Common aims and fiduciary duty**. AI does not have a unitary goal as does medicine (patient health) as a source of solidarity. Practitioners work in competitive and profit-driven environments, are not in a formal

[3] https://oecd.ai/en/ai-principles.

profession and are not required to uphold public interests above business requirements in the private sector.

2. **Professional history and norms**. Medicine has a long history of development of norms of good behaviour and professional conduct which are variously codified throughout codes of conduct, and biomedical principlism has been informed by many years of practice, thereby emerging from a history of tradition. AI does not have a unitary culture or goals, nor has it had transformative moments in the same way, and its focus is not narrow enough to build specific best practices and particular moral duties. A large plurality of stakeholders results in pushing principles towards high levels of abstraction, and principles endorsing contested terms like 'fairness' that are not directly action guiding. Unlike medical care professionals, AI practitioners are at a distance from their (moral) patients in time and space, dealing with opaque systems and embedded in complex networks of actors. On that note it should be added that the problem of many hands may obtain, whereby assignment of responsibility becomes impossible in vast networks of agents involved in different times and places in the development and use of iterative technologies [74].
3. **Methods to translate principles into practice**. Unlike medicine, AI development does not have generations of tried and tested approaches (including case precedents) to translation of principles into working practices (developed, tested, and renewed over time by various stakeholders like accreditation boards and so forth). Norms and requirements, Mittlestadt argues, cannot be derived from mid-level principles, requiring independent justification at each stage of translation. Furthermore, the costs of integrating ethics oversight into organisations can be unattractive for profit-making entities.
4. **Legal and professional accountability**. AI practitioners are, with some exceptions (e.g., data protection law) not subject to external/legal regulatory environments that can censure them for breach of professional norms and otherwise provide the machinery of accountability for answering for wrong-doing.

A more provocative attack on AI ethics as a whole comes from Luke Munn, who somewhat mirrors arguments made by Mittlestadt, and notably Anaïs Rességuier and Rowena Rodrigues (albeit perhaps in a more pessimistic light) [64, 79]. Munn argues that AI ethics principles are essentially ambiguous or meaningless without any consensus around key terms (again such as fairness) which are subject to being interpreted by organisations to conform to pre-determined product features and business goals [64]. Furthermore, industries involved in AI are 'unethical' or 'a-ethical', marked by the values of dominant participants and a lack of ethics training of incoming professionals, and they do not consider the intersections of race, class and culture and technology and the harms they might cause, and 'downstream' ethical frameworks, situated below company culture, cannot address '[…] more fundamental inequalities and underlying social issues that shape technological development' [64]. Finally, ethics has 'failed' due to a lack of enforcement and compliance mechanisms in an industry that moves faster, and at scale, than legislation [64]. Munn proposes moving away from AI principles altogether, to expanding the field of inquiry under the framework of *AI justice*, i.e., '[…] if machine learning reflects, reproduces, and amplifies structural inequalities, then any ethical program must operate intersectionally, considering a wide array of social and political dynamics' [64]. Such a view is generally shared by scholars including Catherine D'Ignazio and Lauren F. Klein who argue from a 'data-feminist' perspective and explore paths for challenging power asymmetries (stemming from oppressive systems of power) and injustice between dominant groups and marginalised members of society as exacerbated or reified by data and digital technologies [27]. In fact, D'Ignazio and Klein list concepts that are deficient as they secure power by locating the source of problems in individuals or technical systems only, which are: 'ethics', 'bias', 'fairness', 'accountability', 'transparency', and 'understanding algorithms'. They advocate for concepts that go further by challenging 'structural power differentials' such as; 'justice', 'oppression', 'equity', 'co-liberation', 'reflexivity', and 'understanding history', 'culture', and 'context' [27, p. 60]. This transition to apparently more radical concepts and approaches, from ethics to justice, can in theory be justified by the apparent failure of the old formulations and approaches to meaningfully prevent abuses of AI and other data powered systems, whereby despite the prominence of ethical language we continue to see systemic failures that result in prejudice, discrimination and widening power differentials, and losses of privacy all reinforced by institutional evasion and indifference [4, 27, 30, 31, 41]. In Netherlands there is the relatively recent example of the (at least initial) failure of the status quo to protect people from algorithmic systems in the domain of welfare and tax where biased algorithmic systems resulted in harm to citizens [3, 17, 44]. Nevertheless, it can also be argued that the difference between AI ethics and AI justice is artificial, that the domain of ethics does accommodate, even necessitate, the more radical concepts which have admittedly as of yet failed to meaningfully materialise in many instances (we will return to this in our discussion of the *dispositif).*

Building on this and also following years of scholarship in post-phenomenology and the philosophy of technology [43, 49, 107], it is reasonable to say that it is not necessarily the things in themselves (whether that be AI or other tools) that are sources of harm, which challenges the application of principles to specific sites of action (an organisation developing an AI or other tool), when it is the complex dynamic of technologies, people, and systems and processes of their development and adoption (socio-technical systems/assemblages) that influence and bring to bear the social and ethical consequences of technological artefacts [7, 26]. Moreover, the use of technical artefacts across domains challenges the very notion of applied ethics itself when those artefacts are utilised across a spectrum of domains (business, justice and health, for example) [41, 105].

Having given a brief overview of some powerful critiques of popular current approaches to the ethics of AI (and technology more broadly), we will now proceed to introduce in more detail a framework that presents a promising alternative to some of these approaches, and one which may well answer to at least some of their deficiencies—virtuous practice design.

3 Virtuous practice and narrative and technology ethics

Two notable developments have occurred in the philosophy and ethics of technology in recent years that present, in combination, fruitful avenues for thinking about our being-in-the-world with technology and how we become human, and moreover how we do so in our movement towards the good life, with essentially ambiguous technical instruments applied in evolving and novel technical practices. The first development concerns the turn towards virtue ethics by some scholars in the ethics of technology, as an alternative to arguably more dominant theories including deontology and consequentialism [15, 22, 47, 95, 105]. Here, by virtues we refer to fixed traits of character or mind that involve dispositions to think, feel and act in particular ways appropriate to various circumstances, and where a person of practical wisdom (*phronimoi* and *phronesis*) recognises the particular morally salient features of different situations, and recognises the correct virtues applicable to different situations and thereby ways to think, feel, and act called for by them [10, 98, 102]. Such traits of character are acquired from habit, and are required in pursuing our ends in the course of our functions, the ultimate good (flourishing, *eudaimonia*, or the good life), which is 'living well and doing well' [10, 48, p. 618]. Examples of such virtues include benevolence, conscientiousness, courage, generosity, gratitude, justice, honesty, loyalty, and temperance. More recently, Shannon Vallor has proposed the following as (technomoral) virtues necessary for living well in an age of endless technosocial possibility: honesty; self-control; humility; justice; courage; empathy; care; civility; flexibility; perspective; magnanimity; and technomoral wisdom [105].

The second (and complementary) development we wish to address here is the growing interest and use of hermeneutic and narrative philosophy in the philosophy of technology. Hermeneutics has been defined as, 'an art, technique, and technology for the (correct) interpretation of cultural productions, mostly texts. [And] In the twentieth century, hermeneutics became a philosophical movement dealing with interpretation and understanding as the main features of humans' "being-in-the-world"' [88, p. 73]. More recently it has since been developed to consider 'being-in-the-world' with technology, and its uses have been variable and approaches not 'unitary' as such [88, p. 74]. This approach emphasises the contribution of technological and digital artefacts to mediation in the world in examining our processes of technological appropriation [56], as well as the very process of individuation (of becoming human) in the technosphere [32]. This approach follows the complementary tradition of post-phenomenology in examining the relations between person and thing and how these relations extend to and determine how we relate to the world around us and each-other in adopting new tools, discovering new uses for them through their affordances and thereby learning more about the world and ourselves in a continuous process of interpretation and self-interpretation that occurs in the midst of new technological evolutions and configurations [49, 56, 107]—the role of interpretation, and reinterpretation, is continuous in this process.

The gap between virtue ethics, hermeneutics and narrative was arguably most famously closed by Paul Ricoeur in his development of narrative philosophy and his development of a 'little ethics', and Alasdair MacIntyre in his thorough treatment of the virtues with reference to narrative in his seminal title *After Virtue* [58, 81–84]. For Ricoeur, the ethical intention was aiming at the good life, with and for others, and in just institutions [84]. He proposed an Aristotelian *eudaimonist* philosophy privileging the idea of ethics (as opposed to morality alone), but placing it within the (*prima facie*) boundaries of deontological norms that act as (overridable) constraints to the ethical intention, and within the wider framework of political institutions and practices that exist to adjudicate competing claims and oversee justice. Hermeneutically speaking, actions are readable as text, and one who interprets these actions (their own) is self-interpreting against the background of the aim of the good life and their particular decisions and choices, and this kind of interpretation of the ethical self becomes self-esteem (the 'reflexive moment of the wish for the "good life"') [84, pp. 179–180, 192]. Figuring intimately into the narratives of human lives, and how we self-interpret, are practices—practices being socially established and co-operative activities, which are coherently structured around their constitutive rules, and that we engage in to achieve goods internal to those practices (as the musician pursues beautiful music) through standards of excellence, and which rely upon the virtues to be achieved [58]. These practices are engaged in by persons who link them to their life plans (of being a pianist, for example), and both together feed into narrative unity of life, which is the basis for the aim of the good life [58, 84]. The triadic structure of the ethical intention is useful as it promotes both reflexivity and attentiveness to the other (as solicitude)—it focuses on the ethical construction of self in concert with other in practices that promote public goods, where competing claims are resolved at the institutional level. It is attentive to the very particular details of ethical situations in being tied to ideas of narrative, that is, the story being told, and the composition of practice (which consists of the small units of basic actions directed by constitutive rules towards goods which themselves ought to be evaluated). The theory in being *eudaimonist* owes much to and accommodates the virtues and is therefore concerned with building ethically competent individuals–but also in looking to the moral norm as a constraint to action it acknowledges the importance of the duties, obligations, and principles of deontology although with the clause that apparent duties must yield where they may harm other persons. Moreover, by including just institutions in the triadic framework, elements of political philosophy and a general attention to the construction of just structures of governance make this a framework that extends beyond reflection on individual action to one where it becomes apparent that the ultimate aim and obligation of each individual is to contribute toward the development and maintenance of structures of justice that benefit all and fairly adjudicate all (competing) claims. Here we have a theory, which while fundamentally *eudaimonist*, arguably incorporates some of the best elements of a number of theories and approaches into a unified and comprehensive one. Moreover, this framework can incorporate concern for consequence (*eudaimonist* theory is teleological in nature). When the moral norm yields, arguably it does so because an action might tend towards the harm of another as a consequence—the theory is open to a range of considerations and also solutions in that it is also concerned with dialogue, debate, and governance.

More recently, the narrative philosophy and ethics of Ricoeur and MacIntyre have been applied to the domain of technology, prominently by Wessel Reijers and Mark Coeckelbergh over a series of recently published works, which also build upon the tradition of post-phenomenology in the philosophy of technology [75–77]. Following an earlier suggestion by David Kaplan that Ricouer's (in particular) philosophy could bring much to the philosophy and ethics of technology, they have built upon a hermeneutic framework incorporating Ricoeur's 'little ethics' (and the multi-staged process following the aforementioned three parts of the ethical intention) and MacIntyre's approach to virtue ethics, applied to technology which they recognise as being text-like itself in its encounters with humans in different contexts, and which can have a role as co-narrator in human lives through its configurative potential [53, 75]. In their framework, Reijers and Coeckelbergh apply Ricoeur's concept of mimesis to understand how encounters with technology are prefigured, configuring, and refiguring, or, the nature of technological mediation and the extent of transformation in understanding of the world or self-understanding technologies can contribute to, and importantly transformations of technical practices. For them, the narrative mode makes technical practices intelligible. Moreover, the authors build on Ricoeurian concepts of textuality, literacy, temporality, and distancing in order to determine the configurative potential or reality of technology vis-à-vis technical practices including how technological tools impact orderings of events and actions, who can use and is affected by them, the degree from which they abstract from the physical world and so forth [75]. It is through narrative investigations that technical practices are studied. These investigations are conducted with a view to understanding the ethical (and even hermeneutic and phenomenological) implications of new and evolving technical practices. Such investigations are set out on to support the design of technical practices that ideally nourish and do not impede the virtues (or rather technology in practice extends rather than precludes human virtuous capacities), as well as help identify relevant norms, codes and other interventions to regulate practices that are effected by new and emerging technologies—that is, this process is undertaken to support *virtuous practice design* [75, 76].

Like VSD, the VPD framework is tripartite and iterative, being conducted through three phases drawn from the narrative philosophy of Paul Ricoeur [81–84], which are:

- *Phase 1: Description < > Interpretation*: The purpose of this first phase is acquiring an understanding of the full network of technical practices in which a given technology is embedded [75, p. 156];
- *Phase 2: Interpretation < > evaluation*: The second phase consists of gaining an understanding of the technical practices with regards to 'puzzles' they raise in relation to life plans, standards of excellence, and the narrative unity of life [75, p. 156] and is informed by prior stages;
- *Phase 3: Evaluation < > prescription*: This final phase consists of evaluating the technical practices in relation to the ethical intention [75, p. 156]. The purpose of this

> is prescription towards a stable for-the-sake-of-which of technical practice that cultivates virtues [75, p. 174].

Unlike VSD, VPD commits to narrative theory to provide the basis for a hermeneutic understanding of technical practice and mediations, as well as to an ethical theory rooted in virtue ethics but in direct dialogue too with deontology as well as incorporating and legitimising existing AI principles (on the condition that they can be embraced by the community of AI developers[4] and practitioners and can be overridden where they conflict with the ethical intention). Now, design decisions and evaluations of states of affairs make direct reference to normative theory, focus on virtues instead and not only potentially an arbitrary list of values. The framework also extends more fruitfully its concern from only the ethical design of artefacts to entire practices implicated by the artefact, to ultimately directing the attention of AI developers and practitioners to political engagement in negotiating the norms and boundaries of their practice [75]. What's more, VPD need not be considered a complete alternative to VSD but a logical progression, thereby it can incorporate VSD's methods where necessary, and even evaluate value implications where they correspond with standards of excellence of practices and the internal goods they pursue—we will briefly return to this in what follows.

Before that, for a brief and more practical point of comparison, we can imagine both frameworks being applied to the development of a crime hotspot algorithm that produces crime risk scores for different administrative units in a city. VSD might undercover relevant values to this context (policing and data science) and endeavour to engage multiple relevant stakeholders, eliciting information to support the ethical design of the algorithm that minimizes bias, privacy intrusion and interference with autonomy through technical solutions. A VPD based approach may be yet more collaborative, endeavouring to capture changes to how the police do their job because of the algorithm, to what extent it co-narrates their practice or impedes or encourages the cultivation and exercise of virtues (their ability to be fair and beneficent to civilians). VPD would also examine narratives of persons from within patrolled districts to examine how such technologies may affect their day to day lives and examine histories of, for example, racial profiling. The VPD approach may result in the need for the same technical measures being implemented in the artefact, but also points towards collaboration between community, police, and AI developer to shape deployment and operational rules of the AI tool so that it can enable police to pursue the internal goods of the practice (crime safety) whilst utilising their virtues to deal fairly with citizens and the community or to attempt another non-technical strategy where this is not likely (for example, working with social services interventions in high-risk areas rather than increasing police patrols). Reality will always be more complex, yet the example points towards the extended horizon of responsibility placed on the shoulders of those who adhere to the VPD framework.

In the following, we will examine how features of this framework may address some of the weaknesses of other approaches in the ethics of technology. We refer to this overall narrative framework built around narrative philosophy and Ricoeur's 'little ethics' as virtuous practice design (VPD), based on how it was referred to upon its earliest apparent inception [76]. The present authors are unaware of it being referred to as this in more recent iterations [75], but will retain the initial VPD terminology as it succinctly conveys the intent of the framework explored—the design of virtuous technical practices. Finally, before proceeding we must acknowledge some foreseeable weaknesses of this approach. It would appear that the kind of narrative investigation of technical practices suggested by Reijers and Coeckelbergh is a truly significant and in-depth, multi-step and iterative undertaking that would likely be resource intensive and a challenge to execute for small and medium sized businesses. Another potential issue is that we live in a global, plural world and this ethical framework emerged in the Global North primarily through a series of Western thinkers therefore one might argue that it may not be suitable for translation across cultural contexts, or that its use might be at odds with decoloniality. The first point is significant, but there is a cost to responsible innovation that is necessary to bear in order to build a sustainable socio-technical system and functional society. We will revisit this in Sect. 4.5. The second point is enduring in any discussion of ethics, however we can say, at least, an approach built on VE is firstly one that promotes an attentive disposition towards the needs of others in their diversity (the virtues of care, flexibility, perspective and so forth), and secondly Vallor has done much to demonstrate that the virtues are culturally robust with similar traditions identified in Confucianism and Buddhism [105].

4 The advantages of virtues and narrative for supporting virtuous practice design of AI systems

From Sect. 2, a cross-section of some of the various challenges to a successful ethics of technology and AI can be broadly aggregated as:

1. Inadequate or inappropriate ethical frameworks;

[4] Understood as the technical profession of design, development and deployment of AI solutions [104].

2. Ahistorical and ambiguous principles detached from practice;
3. Ethics does not challenge fundamental inequalities at the societal level that are perpetuated at the level of practice;
4. A lack of enforcement and compliance mechanisms;
5. Resourcing ethics (and the will to do so).

There is no panacea that can address all of these categories of challenges and their nuances, yet virtue ethics, particularly when combined with narrative theory and carried towards a methodology of virtuous practice design, provides constructive responses to some of these weighty challenges and promises a potentially fruitful approach to the ethics of technology and AI. Moreover, a VPD approach can help transform responsible research and innovation (RRI) without necessarily upending it in its entirety. VPD provides useful conceptual and methodological tools that can bolster technical practice without strictly dismissing what may have worked or what can work about what came before it. Attention to the moral norm, for example, demonstrates that principles have their place in acting as overridable constraints when informed properly by their context. It represents additional tools that can be used to bolster responsible innovation and technical practice, tools that can perhaps supplement and save some existing approaches rather than replacing them entirely. The following five subsections will provide some responses to each of the outlined challenges, demonstrating how VPD can contribute constructively to resolving some of these difficult and nuanced issues.

Going forward, the AI developer is the focus of analysis here, as AI development is a major site where the implementation of ethical principles has been argued to fail—though note that VPD necessitates wider engagement (including by the various technical practitioners not just designing but *using* a tool). The following will argue how VPD can support virtuous practice and virtuous practice design by AI developers primarily, but acknowledges that virtuous practice design must be observed by all stakeholders of significant influence (policy-makers, service providers and of course users of AI tools etc.).

4.1 Inadequate or inappropriate ethical frameworks

The first challenge here relates to the use of ethical frameworks that are problematic in themselves (e.g., utilitarianism favouring happiness of the majority), or of their limited applicability against the massive scalability and unpredictability of technology and its course of evolution in terms of design and adoption which challenges their conceptual and analytical resources. Firstly, Vallor proposes VE as a viable way forward of addressing the technological challenges of the 21st century [105]. VE permits a flexibility not usually characteristic of rival theories, and is concerned with context and particularity, and cultivating skills in moral agents to recognise morally salient features in novel situations and responding to them with style and as particular situations demand (through the virtue of *phronesis*). The responsivity of VE, and the idea of cultivating virtues and practical wisdom as a mediating virtue, may reduce uncertainty of action stemming from the ambiguous calculus of utilitarianism or the lack of concrete guidance provided by the categorical imperative in a dynamic technological landscape, or at least better support moral agents in responding to emerging and changing demands. Moral expertise is reflected in, but not drawn from, fixed principles [105], which is to say that principles can emerge from observations of patterns of right action, but right action does not (necessarily) emerge from principles. VE allows the virtuous agent to meet techno-social convergence (perhaps the deployment of algorithms in novel circumstances) with the skills to recognise right action in the face of ethical uncertainty (such as by reaching out to affected stakeholders and re-designing any elements of a system that might be conducive to unfairness).

The VE enriched narrative approach suggested by Reijers and Coeckelbergh based on MacIntyre and Ricoeur's work further strengthens the case for a virtue ethics based approach to the ethics of technology and AI, by focusing on timely, concrete and particular details [22, 58, 75, 84]. This approach requires scrutiny of practices and technological configuration occurring within those practices through interpretation of textuality, literacy, temporality, and distancing, examination of goods internal to those practices as well as the ideals, life plans and standards of excellence that all connect them to the idea of the good life. Such an approach (ideally) cultivates a detailed understanding of stakeholders in a technical practice, why they engage in it (or how they are affected by it), the relational connections between tools and their users, makers and governors (as defined by the constitutive rules of the practice), standards of excellence, and inquiry and reflection on standards of excellence and virtues necessary to secure internal goods within the particular context of a practice. By considering technology within practice, ethical inquiry becomes bound to specific, rather than general, contexts of technological use cases and ethical recommendations are attached to ethically relevant features of these particular cases—for instance, abstract questions of the application of fairness may become more concrete as we investigate a credit rating algorithm and its surrounding narratives and reveal the need to take specific action (synthetic data), or design the practice it is embedded in to enable ample human intervention to allow individual appeals.

Moreover, VE extends the field of consideration in ethics beyond reason (or universal rationality) to relational, embodied, and importantly emotional considerations that may be eschewed in Kantian thinking, and includes '[…] emotional and social intelligence: keen awareness of the motivations, feelings, beliefs, and desires of others […] [105, pp.25–26]. This approach properly acknowledges legitimate human capacities (feeling and emotion) which have ethical salience (emotions have been said to be felt value judgments [68, 86]) which are valid and even necessary in ethical deliberation, but are not regarded as legitimate to ethical deliberation in some (albeit not all [101]) forms of deontological (or rationalist) thinking [59]. In fact, some argue that values are grounded in emotions, and emotions are the currency of values [59]. This acknowledgement of emotion as a valid form of knowledge (ethical knowledge in this case) collapses what has been argued to be a false (and alienating) binary between reason and emotion [27] and creates opportunities for inclusive participation [96] and expanding the field of what we might consider *phronimoi.* Through skilful and necessary application of empathy and compassion, and feeling with others, achieved through listening to others' stories, we can better understand and respond to sources of harm that otherwise might go ignored or overlooked.

It is important not to overlook the role of emotion in moral judgment and its relation with reason itself is often one of close connection (for example, as argued by May and Kumar, feelings can facilitate inference), with some arguing that it can be difficult to pry them apart and indeed we may need emotions in order to have practical rationality [59, 87]. The relationship between reason and emotion is not a clear cut one, nor one of stability in practice, with some arguing that whilst emotion aids reason, it can also corrupt it (but likewise, a sufficiently clinical approach to reasoning may push back what might be considered valid emotions, e.g., fighting the grief of the passing of a loved one by reasoning that death is natural and inevitable) [59]. There arise situations where we can simply feel something is wrong whilst reason that it complies with principles (where we might say in some cases the moral norm should yield to solicitude), vice versa, which creates difficult inconsistencies [59]. There have been historical examples of situations where principles and rules have given way to solicitude, where empathy and compassion have resulted in revision of principles and rules which were understood as unethical upon reflection and in dialogue with emotion, for example, improvements LGBTQI + rights in many countries [59]. Reason and emotion are arguably both then, when in dialogue, necessary for appropriate moral judgments but their interactions should respond to the objective features of particular cases, and this process should be one which corresponds with reflective equilibrium and the balancing of the general and the particular [59]. The clash of the affective with the rule is similar also to the clash of a rule with a rule and the processes required or prescribed by VPD understand and respond to such clashes to promote objective decision-making, including through an *ethics of argumentation*, which can carry tradition and convention to considered conviction and reflective equilibrium [84]. In situations where time for dialogue is available—diverse stakeholders can be brought into discussions where reason and emotion are in tension.

What alternative ethical frameworks may not so easily support is the empathic feeling with others that can direct a developer's attention to sources of harm through emotional revelation of shared humanity potentially via the narrative form as a medium. The AI developer who comes into contact with the story of someone who has experienced some injustice, perhaps a spurious welfare investigation into the wellbeing of their children due to their limited financial resources, may be more inclined to reflect on their training and input data and who is being flagged by their system and why, and deliberate on whether the tool's outputs are appropriate. Such an impetus may arise upon the realisation, for instance, that parenting on limited resources is not bad parenting, illuminated by the perception of the pain and humiliation of the other. Both oral and written narratives of the experiences of others can lead to these emotional revelations.[5]

We can expand further on VPD's implementation of deontology and the background to it. Ricoeur does not dismiss deontology, which represents the moral norm, and is a constraint on actions towards the ethical intention of living well, with and for others, in just institutions. Ricoeur is aware of the limitations of deontology, but so is he aware of the necessity of universalizable rules that articulate the ethical intention, in order to stipulate some formal boundaries of acceptable action, and which recognise human autonomy and human plurality [84]. The ethical intention and the moral norm can come into tension, and after reflection, it is the moral norm that must yield to the ethical intention, properly evaluated in light of deliberation and considered conviction. Ricoeur's use of deontology allows for the creation of formal rules, which whilst not based on observed patterns of right action or derived from exemplary behaviour from *phronimoi*, act as an initial and reasonable bulwark against evil action and are overridable where they obstruct right action due to unforeseen or previously unforeseeable reasons. Deontology in this case supplements his centrally *eudaimonistic* theory, and does not supplant it nor the important resources it brings (along with a proper

[5] See Virginia Eubanks' excellent Automating Inequality for an in-depth examination of how algorithmic systems contribute to the further marginalisation of people in poverty [30].

accord of emotional capacities by way of critical solicitude). It is therefore arguably somewhat resistant to technological uncertainty inasmuch as it does not attach to monolithic or intractable rules. In the practice of AI development, where the aspirations of virtue come into tension with deontological norms (or those norms conflict), there must be a process of reflective equilibrium and reflection on considered conviction that may necessitate periods of stakeholder consultation.

On the topic of inadequate or inappropriate theoretical frameworks, it was pointed out in Sect. 2 that some otherwise useful methodologies lack any anchoring normative framework or at least fail to endorse one (i.e., VSD and ethical impact assessments). Reijers and Gordijn provide a good defence of the selection of VE as a grounding normative theory in the ethical design of technical practices (as opposed to the VSD) approach [76]. They argue that VPD avoids (the debatable) arbitrariness of VSD's value selection due to being supported by a normative theory (VE), and furthermore that this approach lends itself to uncovering (with stakeholder engagement) life plans, standards of excellence, and virtues with regards to technical practice (rather than simply technical artefacts, it, as highlighted throughout here, considers entire practices, human development and regulation involving technologies) all grounded in an Aristotelian philosophical anthropology [76]. Furthermore, elements of VPD itself incorporate traits of the ethical impact assessment and can credibly ground them in VE much the same way, through stakeholder engagement and identification of life plans, virtues and standards of excellence that might be affected by changes in technical practice, albeit within a more comprehensive framework and not a reductive one. The description < > interpretation phases, and interpretation < > evaluation phases and the methods entailed cohere with the course of risk and impact identification, and the evaluation < > prescription phase coheres with the outcome of actionable recommendations in an impact assessment. Arguably, an approach to the ethics of technology (and by extension, AI) as outlined by Reijers and Coeckelbergh already incorporates an ethical impact assessment itself, even if none of its stages or phases are explicitly referred to as such [75].

It might further be argued, as we suggested earlier, that there is a complementarity between VSD and VPD and that the methods of VSD need not be overlooked. More radically, it might even be a defensible position to look to VPD as an evolution of rather than an alternative to VSD—it still considers values in its own way by incorporating considerations of the *eudaimonistic* goods that we pursue in practice and the standards of excellence that can be applied to evaluate how we reach those goods within those practices.

These goods and how we reach towards them constitute the ends and movement towards good life. We value them, and they are obtained through and sustain the virtues. The key difference between the frameworks are, as stated, that VPD endorses the virtues (without ignoring deontology and so forth) and extends its field of concern more outwards to all practices implicated by technological artefacts at study as well as questions more significantly beyond the remit of artefactual design. Both VPD and VSD ultimately consider values in design, for instance the privacy and autonomy implications of a facial recognition system—however VPD will additionally and more explicitly consider the further implications of such systems for the (specifically prescribed) virtues of the users in their design and deployment, which may prompt an AI developer to dialogue seriously with intended end-users and regulatory bodies about appropriate uses of such systems above and beyond technical measures that can be designed into such systems (e.g., data expiry of recorded footage). Some such examples will be explored again in the following subsection.

There are reasons to continue accepting the advances of VSD with reference to VPD (for one example, modular design in relation to value change, as stated, remains a relevant finding and proposal [73]). VSD scholars have been probing and experimenting with system design in novel ways—they may find that it is worth bridging the gap into VPD [61]. Now having explored some advantages of the VPD framework as an ethical framework as opposed to other choices, let us turn now to how VPD might credibly address some problems with more principles based approaches to an ethics of AI and technology.

4.2 Ahistorical and ambiguous principles detached from practice

This challenge is somewhat like the previous one, however it allows us to address the proliferation and adoption of AI principles in particular, and not general ethical frameworks such as consequentialism and deontology, and their place in the ethics of technology. Principles have worked in professions such as medicine due to a long and storied history and tradition of building an ethical culture and responsibility within the profession, complete with rigorous accountability mechanisms, case histories and precedents, arguably leading to more concrete understandings of the relevance of biomedical principles to concrete situations [63]. By contrast, the competing interests and plurality of actors in AI, a practice (or practices) of convergence inasmuch as it is domain agnostic with regards to its multifarious applications, without a singular ethical culture, can lead to abstract principles or understandings of principles surrounding contested terms (like fairness) and such principles do not have direct action guiding content, nor can they easily be translated into practice [63, 64].

The practices of AI developers understood broadly (AI can be said to be composed of multiple practices such as data science and software engineering) are relatively recent and are evolving at great pace. Whilst much effort by way of creation of principles and standards[6] is being made to instil AI practices with some ethical direction, such efforts are fragmented and are not necessarily or always emerging from a history of tradition of the practices themselves (even if arguably they represent the first significant steps in building an ethical tradition), which are also fragmented and have themselves a limited or variable history of tradition. It might be more fruitful at this time to observe the practices of AI development as being at the juncture of the practices to which they attach. Considered in this light, we can demonstrate the use of a VPD approach that subordinates principles to engagement with the particular, and a focus on the ethical intention. To this extent, it might be emphasized that virtues themselves are not only supported or hampered in the design of technical (and AI) practices, but support the very process of VPD, i.e., an AI tool developer will have to exercise at least perspective, flexibility, and civility in engaging in social, other-regarding processes of stakeholder engagement and contemplating the ethical intention.

Those developing technological tools for particular (or multiple) use cases in distinct other practices (like professions) should strive to identify as, minimally, guests to those other practices their work informs and is informed by. Whilst AI development is a distinct practice defined by the pursuit of internal goods, broadly speaking, such as efficiency, or knowledge, such internal goods only become meaningful in the context for which they are intended. An AI tool developed to diagnose different illnesses can bring about efficiency in medical diagnosis, and such an efficiency gain might be a good resulting from the efforts of AI developers, but such efficiency only becomes meaningful within the context of another practice, that of medical care (the for-the-sake-of-which of the AI tool), a practice it may transform (or refigure). Here, we have a good opportunity to expand on the example of the medical care scenario and probe how applying VPD to it may look. The roles of AI developers and medical practitioners form interweaving narratives in pursuit of the good life—the story of the data scientist in the wider AI practice interweaves with that of the medical technician using an AI tool for medical diagnosis. The *telos* of one practice, AI development, represents a movement towards a further *telos* on the way to the good life (medical care)—the technical and medical care practice are inextricably conjoined by the AI tool. Phase 1 description < > interpretation should reveal the relevant narratives, stakeholders, their relationships, and the mediating influence of the technology and how it shapes the practice of medical care.

The AI developer, whose work is now conjoined with medical practice, should be attentive to prefigured narrative surrounding medical practice, obtain an understanding of relevant life plans of stakeholders, its culture of ethics and existing constitutive rules, standards of excellence (e.g., correct diagnosis of an illness) that their work will have potentially transformative implications for, as well as existing law and regulations applied to the field of medicine with implications for their work (data protection legislation), and how they relate to their own ideals, and life plans (i.e., phase 2 interpretation < > evaluation). The AI developer then is required to be attentive to the context of their intended use case, and should engage with a broad array of stakeholders (patients, doctors, medical technicians, policy officials, marginalised communities who may be overlooked in medical care settings) in order to properly understand this context and in order to reveal their unique needs, and possible boundaries of action. In unravelling these narratives, the ethical needs and requirements of the use case start to become apparent—they are not given *a priori* by principles, but made concrete by the morally salient features of the case as revealed by characters (stakeholders) in the story. The AI developer should also be sensitive to the constitutive rules of their own practice, as well as standards of excellence (even if they may be nascent).

At this point overall, a narrative investigation may have revealed acceptable false positive/false negative rates for illness prediction, historically overlooked communities in data sets, cultural objections to machine deployment, standards of excellence concerning patient confidentiality, factors relating to medical codes of conduct, whether the use of a new machine creates distance between doctor and patient or threatens the autonomy of either, whether the new tool is appropriately accessible to medical technicians—all of which will figure into Phase 3 evaluation < > prescription. Phase 3 requires AI developer to cultivate self-regarding virtues including humility, courage, perspective, self-control and magnanimity for self-esteem, to be fostered through mentorship, as well as taking on a norm (such as ethical codes), given to themselves [75]. The normative ideal of non-discrimination might make the AI developer establish procedures for mitigating data bias. Similarly, with regard to the normative ideal of confidentiality in medical practice, the AI developer could interpret a norm of non-maleficence to use anonymous datasets in their work, or use best practice to develop secure systems that protect patient data on site if necessary.

[6] For more on standards, see for example https://www.iso.org/sectors/it-technologies/ai.

An AI developer should also consider other-regarding virtues including care, friendship and honesty [75]. Firstly, the AI developers should be concerned with building a culture of cultivation of these virtues internally, but again, in coordination with medical practitioners, can devise ways to prevent their artefact from impeding medical practitioners' opportunities to exercise these virtues (in this regard, explainable AI might for instance help medical technicians and specialists to interpret and honestly communicate information about the tool's decisions and limitations). Next, with reference to Kant's second formulation of the categorical imperative (the moral norm), the AI tool developer should consider taking and specifying the norm of respect for others (for example in a code of conduct) both in their interactions with each-other internally as colleagues (treating one and other with dignity and respecting autonomy) and making design decisions based on the observed constitutive rules and standards of excellence in medical practice that appropriately respect others, including for example respecting the autonomy of doctors (e.g., ensuring a system does not convey predictive decisions in a coercive way). VPD then should attempt to curtail disruption of the ethical foundations of practices by aiming to integrate the new technical aspect of a practice within existing ethical cultures to the greatest extent possible, in dialogue with the practitioners (of medicine, in this case). There are tensions involved in this process, and these will require further research and experimentation in use case settings.

An AI tool developer can consider critical solicitude if they can foresee any conflict between the ethical intention and the norm evident from the intended use case. Generally speaking however, an AI tool developer should maintain continuous feedback from vulnerable populations or their advocates to devise strategies to mitigate harms to vulnerable people, as should medical practitioners—we will discuss this topic again briefly in the following subsection. It might be again that insufficient data about a given population may result in inaccurate illness diagnosis, and therefore the designer should consult with those populations in order to ascertain a way forward. Further elements of the framework will be illustrated in the following subsections, however for now it should be evident that a VPD approach can go some way towards overcoming the weaknesses of principles in AI ethics, and it can do so in fact without even dispensing with them entirely—a VPD approach requires the kind of immersion in practice that can render the abstract more concrete. A VPD approach mobilises all stakeholders in a technical practice to the extent that the use case practice subsumes the design/development phase of the AI practice. AI developers arguably do not have a long history of tradition in terms of standards of excellence and generic principles may not provide much action guidance, however with attention to the context (to the relevant narratives and the protagonists in those narratives), AI tool developers can adjust their decisions and design choices to the needs of the stakeholders involved; from vulnerable populations who can use their own voices in helping to conceptualise terms like fairness, to medical practitioners (for example) who have long established standards of excellence from which AI tool developers can learn, and can then design their tools to minimally disrupt adherence to those standards and the cultivation and exercise of virtue in those practices.

Singular ethical cultures can arguably be fostered over time with the help of appropriate exemplars and feedback from relevant stakeholders, however given the tendency of attachment to technical tools to many different practices, AI developers will also need to appreciate and respect the cultures that they enter into as people providing intended solutions in domains that are not necessarily their own. The moral norm does not preclude the adoption of principles, however with the understanding such principles are subordinate to the ethical intention and merely act as boundaries that are overridable—ultimately it is from studying particular contexts and engaging with stakeholders that they are supplied with useful substance. These principles do not exist in a vacuum, and are bolstered by processes of extensive stakeholder engagement and narrative interpretation. Translation of principles into practice is based on studying particular contexts, and with due reflection on existing norms of a particular use case, as well as the life plans, ideals, and internal goods associated with the practice of the use case.

It bears emphasis before continuing that principles do possess an inherent value in themselves by providing starting points for ethical discussion across any domain of ethics. Moral expertise is reflected in, but not drawn from, fixed principles [105], which is to say that principles can emerge from observations of patterns of right and virtuous action, and whilst right action does not (necessarily) emerge from principles, they can be indicative of necessary discussion and the right actions which have been historically observed in certain contexts. As argued, in the practice of AI development patterns of virtuous action [105] are under constant negotiation and movement but can continue to inform and crystallise into useful principles, the content of which and applicability to particular cases can also be supplied by ongoing reflection and stakeholder engagement and examination of relevant pre-figured narratives.

Having now proposed, with some illustrations, how a VPD based approach can potentially go above and beyond the requirements of a principles based approach, and how it can provide more concrete guidance to AI tool developers from a level of immersion and understanding of linked practices, we will next move on to the question of inequality and systems of oppression.

4.3 Ethics does not challenge fundamental inequalities at the societal level that are perpetuated at the level of practice

Catherine D'Ignazio and Lauren F. Klein, in describing power, argue that:

> We use the term *power* to describe the current configuration of structural privilege and structural oppression, in which some groups experience unearned advantages—because various systems have been designed by people like them and work for people [like] them—and other groups experience systematic disadvantages—because those same systems were not designed by them or with people like them in mind [27, p. 24].

Systems entrench the privilege of those that they benefit, their designers, across different domains of power (structural, hegemonic, disciplinary, interpersonal) at the expense of minorities, for example, whose needs either go unconsidered or sometimes by contrast, who become victimised and scrutinised through these domains of power [27]. Similarly Kate Crawford argues that AI systems are embedded across different domains of power (social, political, cultural and economic), and that they are shaped by humans and institutions that determine what such systems should do and how they want it to be done [26]. Ultimately, such unequal power relations continue to manifest into oppressive practices supported by technological mediation as dominant groups reify such forms of domination in developing algorithms (unintentionally or not) (such as in law enforcement [80]) that '[…] reproduce, optimize, and amplify existing structural inequalities' [26, p. 211]. A well-known example of the kind of real-world impact of this was the COMPAS recidivism prediction tool used throughout the judicial process in parts of America, about which a ProPublica report argued that Black defendants were unfairly and disproportionately predicted to re-offend at higher rates than White defendants [9].

Many authors argue that ethics itself is inadequate to challenge such structural inequality and oppression; self-regulation and centralisation of efforts in systems dominated by particular groups continue to reify existing power structures, and moreover decisions about regulation and ethical imperatives are dominated by the Global North, and for this reason discussion should instead focus on justice, or AI justice, and not ethics per se [26, 27, 64]. Again recall D'Ignazio and Klein's endorsement of concepts that go further by challenging 'structural power differentials' such as; justice, oppression, equity, co-liberation, reflexivity, and 'understanding history, culture, and context' [27, p. 60]. The VPD approach (as outlined to some extent here already) also integrates these concepts or at least their essence for the most part—they are either virtues, right actions stemming from virtues, or goods internal to practice, and are each necessary for the ethical aim to live well, with and for others, in just institutions. Whilst ethics itself as a term is open to appropriation by bad or dishonest actors set on maintaining existing oppressive power structures—the fact remains that the terms above remain very much intrinsic features of robust ethical discourse and reflection on achieving the good life for all. VPD therefore also endorses the above concepts, signalling that discussion of ethics and justice are not mutually exclusive—ethics is reflection on the movement towards the good life, which cannot be without justice, and freedom from oppression.

Particular aspects of VPD that bolster concern and justice for others are the 'living well', 'with and for others' and 'in just institutions' elements of Ricoeur's triadic structure of the ethical intention. This requires the design of technical practices that cultivate virtues including care, friendship, and honesty. Care is an important virtue to highlight in relation to this—it is a virtue which has been embraced by feminist scholars and also exists as its own ethical framework in the form of the ethics of care [45, 67]. Some scholars of which have even investigated the relationship of care with Ricoeur's theory [20, 66, 106]. It is an ethic or virtue of care that supports a relational understanding of humanity and action, acknowledging the interdependence of people, and the importance of right feeling for others and acting from that right feeling [94, 105]. Care is defined by Vallor as '[…] *a skillful, attentive, responsible, and emotionally responsive disposition to personally meet the needs of those with whom we share our technosocial environment'* [105, p. 138]. A caring practitioner must be aware of and listen to, and meet the needs of those with whom they engage in a practice (especially those moral patients who are subject to actions of a given technical practice). As argued by Reijers and Coeckelbergh, such a virtue was reconceptualised as solicitude by Ricoeur, which is where the self recognises themselves as another among others, one who cannot esteem others unless they first esteem themselves, and that solicitude carries self-esteem towards justice [75, 84, 85]. Critical solicitude is an integral and important feature of VPD, turning attention towards norms that can harm vulnerable persons, and ensuring that the other, recognised as another self, becomes the proper locus of attention when considering the design of technical practice. Care and solicitude call upon practitioners to seek expert guidance, *phronimoi*, which whilst Reijers and Coeckelbergh suggest, for example, expert advocates, solicitude does not preclude the participation and testimony of members of vulnerable communities who are experts in their own lived experiences [65, 75]. Additionally, the need for such expertise and representation of diverse lived experience demonstrates the importance of inclusion, insofar as possible, of these *phronimoi* in professional roles in practice (including teams that design or use AI systems, and leadership roles within respective organisations of practice).

VPD (following Ricoeur) also explicitly requires consideration of political practices and justice through the 'just institutions' element of the ethical intention. The equality element of this highlights the importance of the virtues of justice, civility, and flexibility [75]. Justice, as understood by Vallor (and similarly to Ricoeur) is [105, p. 128]:

> [*the*][…] *reliable disposition to seek a fair and equitable distribution of the benefits and risks of emerging technologies* […][*and*] a *characteristic concern for how emerging technologies impact the basic rights, dignity, or welfare of individuals and groups.*

VPD calls for civic education for practitioners in these themes and issues, in order to '[instil][…] practitioners with political virtues relevant for establishing the for-the-sake-of-which of the technical practices they engage in' [75, p. 183]. Properly virtuous practitioners (from AI tool designers to policy makers) will consider the justice and equality ramifications (after listening to those stakeholders affected), of particular AI tools. For instance, observance of the virtue of justice would reveal to AI developers and policy-makers that the development and use of an AI tool that allocated disproportionately high crime hotspot scores to neighbourhoods consisting of large minority populations would be an unacceptable and unjust risk to those communities. It might be that the level of police scrutiny encouraged by such a tool could lead to unjust interferences in the neighbourhood's residents' rights—therefore the development of the tool should be revised or abandoned. Such actions and considerations are realised in *the rule of justice*, which requires attention to legal regulation (non-discrimination and equality are well entrenched in many constitutions and in international human rights law), and renders practitioners responsible for contributing to legal and regulatory development where dangerous gaps are identified (as well enjoining them not exploit those gaps themselves) [75].

The next relevant concept to address here is the sense of justice, which '[…] refers to the extent to which practitioners are capable of arbitrating between competing forms of domination, between the governed and the governing and between groups in civil society, in establishing the for-the-sake-of-which of a technical practice' [75, p. 184]. This entails providing mechanisms to challenge domination in technical practice, through design decisions in practices—Reijers and Coeckelbergh provide the example of flagging content online, an important and timely example in an era marked by the proliferation of hate speech on social media services that is harmful to minorities. The sense of justice in technical practice can viewed as involving multiple levels, from the purely design level, to legal and regulatory levels and should provide for democratic-decision making procedures for arbitration, which are themselves open to critique [75].

Again, stakeholder engagement and narrative interpretation are key to understanding pre-figured and configured narratives. This presents the opportunity to listen to the stories of a range of stakeholders, to understand the lived experiences of others, understand how they are helped or harmed by particular or proposed technical practices, and empathise with those experiences and respond accordingly. The concepts at play in a VPD based approach, when authentically adhered to, would arguably tend towards dismantling unjust power structures that dominate vulnerable groups and subject them to harm, particularly in light of solicitude requiring a proper other-regarding attitude and reliance on *phronimoi* from different backgrounds with different lived experiences and expertise; civic education for practitioners to help them understand the relevant virtues and contemporary intersectional issues; and through the sense of justice, open and democratic processes of decision-making and arbitration that are themselves open to critique and that can lead to appropriate law and regulation that protect the rights of vulnerable groups and individuals.

It is on this topic that we can point to the utility of participatory design (PD) and endorse it as a complementary if not constitutive methodology for ensuring representation of a plurality of communities in the design of artefacts, practices, and ideally even futures [52]. As a set of practices, PD facilitates all kinds of activities to support end user participation in the design process of artefacts, notably including narrative methods (including design fictions) which enable participants to convey their narrative histories and aspirations for the future, in stories that are value-laden, and indicative of ways of life including visions of the good life [57]. PD processes can bring the narratives of potentially marginalised others to the attention of AI developers, and through their narratives and dialogue with them, help AI developers work towards practical implementation of ideas of justice into artefacts and technical practices through processes of co-design. Such processes can help with the creation of tools that work for a plurality of groups who can better have their identity recognised by AI systems which will then be less likely to exclude or press them for conformance in ways that are harmful and alienating to their identity and personal narratives [16]. One can imagine the development of more culturally attuned and less homogeneous media recommender systems, for example, with the active participation of diverse stakeholders in the development of such tools.

Recently in PD, there has been a shift among some scholars to decolonial design practices (or away from 'oppressive' systems of design which have been argued to be the mainstream) that embrace non-western epistemologies (through 'radical openness') and different ways of knowing and doing that aim towards pluriversal futures and away from universalist design principles, and de-linking from dominant systems and hegemonic ideas that inadequately acknowledge situated knowledges and the contextual circumstances of marginalised communities [5, 29, 52]. An interesting example of decolonial design practices can be found in [52], where Kambunga et al. engaged Namibian youth (born frees) in special 'safe spaces' where they could dialogue with peers, relay their narratives and ultimately work with students to design artistic installations that conveyed their perspectives and hopes for the future (and promoted decolonial narratives) in a society apparently experiencing some division between young and old arising from a colonial legacy. In this case, the past was explicitly engaged with in an effort by participants to design futures themselves [52]. Participants contributed to design with their stories and the richness of their histories—they are experts in their lived experiences [5, 65]. Decolonial practices are set against certain types of standardization that reject nuance and plurality and would seek to impose the universal upon the plural, and are set against capitalist market logics [5]. By bringing together designer (AI developer) and marginal and plural communities as co-designers, both groups can work together towards understanding different visions of the good life–which is after all not relative but intersubjective–and how the good life can be reached and what virtues can help us reach it, as well as how they can be interculturally understood [11, 84]. Decolonial design practices ensure that a narrative investigation as understood by VPD includes plural narratives, without which ethical design is not possible, and by engaging in decolonial practice grand hegemonic narratives can be challenged. It is at this point of collecting, in collaboration with others, a plurality of historical and speculative narratives that we can begin contemplating alternate narratives of desirable futures that can help us understand how to build technologies that help us reach them [25, 39].

Now, having proposed with some illustrations the fruitful possibilities for a VPD approach to address elements of injustice and systemic oppression, we will move on to the question of the problem of a lack of enforcement and compliance mechanisms in AI ethics, and what VPD has to contribute to this.

4.4 A lack of enforcement and compliance mechanisms

A lack of compliance mechanisms and the unregulated/self-regulated nature of AI practices has been a problematic issue in the ethics of technology and AI [26, 64, 79]. It can be fairly argued that there is little incentive to meaningfully comply with ethical principles when private interests (profit) drive much decision-making in an industry often centred on speed and disruption. VPD addresses this problem with two notable prongs. The first is that VPD mandates the development of ethical cultures within practices (even where they are arguably not present to a significant degree), instilling professionals with the virtues necessary to design and conduct their practices ethically, and to use humility to accept their limits, self-control to choose appropriate goods, care and responsibility towards others, and courage to challenge unjust actions and decisions taking place within their practice. They strive for standards of excellence. This is done through mentorship and civic education. The virtues necessary to realise the ethical intention also move through the sieve of the norm, including self-given codes of conduct. VPD imagines technical practices where practitioners think, feel, want and act rightly based on keen attention to ethical issues, and such virtuous agents are formed through training and education. VPD does not present ambiguous or vague principles that a technical practitioner can wilfully interpret to suit their own ends, but requires virtuous agents to study the (wider) contexts of their practice. VPD mandates slow, reflexive [93] and attentive innovation, and expects it from virtuous agents even in the absence of regulation or law. To this extent, it does suggest a level of self-regulation in the absence of external or legal regulation. Nevertheless, as for the second prong, VPD fully endorses the movement towards external regulation, through participatory processes of engagement with just institutions as described in the preceding sub-section. Virtuous practitioners accept the responsibility for their influence on their practice, should be attentive to legal or regulatory deficiencies and should participate with others in building enforceable norms by way of law and regulation that could result in new accountability mechanisms for their practices [75].

Ensuring virtues are exercised against market pressures in capitalist economies is surely a practical challenge, but one which must be managed by sufficient human resource structures, incentives, and penalties for malicious or negligent disregard for ethical standards and outcomes.

VPD promotes programmes of education and training for virtuous agents to support their reflexivity and perspective, and their moral attention towards, ideally, *phronesis*, so that they can act independently (whilst involving the vulnerable other) in the absence of regulation, but endorses through political practice, the development of law and regulation that can subject practitioners to censure and define and codify specific boundaries and obligations for practitioners. The virtuous AI developer practitioner is, as a responsibility, expected to use their industry insights and experience to lobby for necessary changes at the policy level—activities which could be executed through different workshops, civil society activity, and multimedia campaigns as well as submissions of policy recommendations directly to national and regional lawmakers. There are exemplars, in the fashion of VE, who have engaged in some such activities and may provide fitting models for what AI developers may hope to achieve [62].

It might be noted that in the European Union the AI Act was recently approved by EU Parliament, and it will soon enforce requirements on AI developers that correspond to a tiered risk-based approach depending on the capabilities of the AI system in question.[7] This will not operate at a global level but will eventually demonstrate the capacity for success or failure of such initiatives. In the meantime, VPD promotes the regulatory gap being filled by adherence to the moral norm (including codes of ethics etc.) as adopted by practitioners, as bolstered by and subordinate to plural overriding considerations subject to dialogue and debate with stakeholders in the absence of consensus, whilst ideally AI developers lead in collaborative lobby efforts for effective regulation. Participatory and decolonial design efforts may even help to shape such lobbying efforts by carrying the voices of marginalised or overlooked (or ignored) groups.

To summarise, in practice, an organisation adhering to a VPD framework is one which would build a human resources and professional development machinery that supports the ethical development of professionals through a combination of programmes including mentorship, seminars and workshops, for example. Moral agency is fostered and cultivated within employees, who would still be bound by the organisational and international codes of conduct and regulation, albeit ideally they would be of a certain moral skill to excel at understanding existing rules and operating beyond their at times limited guidance, or in their absence. An AI practitioner with true *phronesis*, over the course of their involvement in the development of an AI tool, may for example be better able to identify whose participation has been overlooked in the process of development and how they should be included in it—a practical knowledge which may not always be reflected in guidelines and rules in specific circumstances. This *phronimoi* then will also ideally have the knowledge and will to recognise deficiencies in regulatory environments and lobby for changes that can benefit the whole industry. What this is to say is that change cannot emerge solely from a virtuous practitioner alone, nor can any spontaneous changes in human nature be presumed, but change proceeds from the mutual feedback of virtuous agents, and the moral norms and law which maintain a role in establishing important boundaries of action, yet it is up to the efforts of individuals to champion the creation or refinement of effective and applicable norms.

We will now address one final challenge to an ethics of AI and how VPD can respond to that, which is the challenge of resourcing ethics and the will to do it.

4.5 Resourcing ethics (and the will to do so)

A final challenge addressed here is the cost of implementing and deploying ethics related processes in an organisation, as well as, in some way relatedly, the will to even do so. As argued by Mittelstadt, the costs of ethics can be unattractive to AI companies, and indeed similarly the costs of compliance with regulation can be implausible for smaller businesses [38, 63]. These (companies) are often primarily answerable to shareholders. Costs can be a disincentive and, again, ethics processes can reduce capacity for fast and innovative disruption. VPD makes no practical considerations towards costs, which it may even incur through education and training processes, and the regulatory compliance that it effectively champions–this is an undeniable reality. From a societal perspective, such costs would be worth it even at the risk of slowing technological progress, as the revenue generated by a rapidly developed AI tool is meaningless in comparison to the harm and violence it could plausibly cause to human beings if designed, developed and deployed without due care. This is not to say solutions may not be available, especially those in the interest of fairness of competition between AI organisations (with added costs, smaller entrepreneurs may fade away and thereby secure the power of monopolistic entities). Larger organisations should theoretically not only be able to weather the costs of devising ethics departments and processes, but may also be able to pay a tax (for example) that could be redistributed to small or medium sized entities in the practice of AI. Whilst it is the structures of the capitalist, market-based economy that create disincentives for the costs of engaging in ethics and regulatory compliance (at least to the extent that compliance is not always marketable), the same such structures and logic of supply and demand it might be argued would create (and arguably have been creating[8]) firms that can provide external, third-party ethics services to support and advocate for the development of ethical AI systems and practices. Through processes of taxation and redistribution, it might be plausible for smaller businesses to engage

[7] https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.

[8] See for example organisations such as Trilateral Research that provide ethics and other support services (https://trilateralresearch.com/).

third-party ethics support through the disbursement of state aid, or even the creation of state funded ethics compliance organisations where there is insufficient market demand for such services. As we have seen, VPD places responsibility on those involved in the design of AI systems and practices to be advocates for change and the onus is on responsible practitioners to recognise and lobby for viable paths forward for supporting sustainable AI ethics mechanisms and to be a part of processes of dialogue with policy-makers to secure this. This issue is one that requires further (interdisciplinary) research but as we can see there are potential paths forward even outside of radical change—nevertheless, VPD commits practitioners to being a part of influencing the environment in which they operate for its betterment.

In practice then, what one might see, for example, from an organisation or coalition of organisations that subscribe to a VPD based framework is also involvement in the design and advocacy of fair and reasonable policies that focus on the redistribution of some wealth from large corporations to smaller ones so that they can remain competitive in a business more stringently concerned with the application of ethics at every level. It is through the explicit engagement with the political that this framework suggests that may lead to more politically engaged practitioners that lobby for the responsible governance of their practice.

The will to cultivate an ethics approach, as described here, within organisations can also be difficult to instil and may not always come endogenously. On this point we can refer back to the previous subsection on regulation and the ultimate need for it—if the will to adopt ethical practices is not present, or being fostered, specific regulatory requirements may be needed to mandate specific ethics processes within those organisations, and both processes of internal change and external pressure are endorsed by VPD.

5 *Dispositifs* of AI practices

A central concept at play in the work of Michel Foucault and one which bears reflection here due to its relevance to the foregoing is that of the *dispositif*, or apparatus. Most succinctly, what an apparatus is has been put by Giorgio Agamben as '[…] a set of practices and mechanisms (both linguistic and nonlinguistic, juridicial, technical, and military) that aim to face an urgent need and to obtain an effect that is more or less immediate' [1, p. 8]. For his part, Foucault himself elaborated on the *dispositif* in more detail:

> What I'm trying to pick out with this term is, firstly, a thoroughly heterogeneous ensemble consisting of discourses, institutions, architectural forms, regulatory decisions, laws, administrative measures, scientific statements, philosophical, moral and philanthropic propositions—in short, the said as much as the unsaid. Such are the elements of the apparatus. The apparatus itself is the system of relations that can be established between these elements [34, p. 194].

Serving a strategic nature, the apparatus represents a manipulation of forces towards its ends, it is '[…] strategies of relations of forces supporting, and supported by, types of knowledge' [34, p. 196]. The *dispositif* and related power relations are an intrinsic element of human becoming according to Foucault, it creates subjectivities, it creates subject—that is, those who are subject (objects of self-knowledge) and who are subjugated in political processes in the dual process of subjectification [37, 60]. The *dispositif* creates possibilities for *action and understanding* within its moving lines of force around relations of power and knowledge in particular domains [19, 28]—that is to say it effects the constellations of relations people have with the world in the process of becoming, as Don Ihde may say, or their individuation as Bernard Stiegler and Gilbert Simondon may have argued [49, 91, 97].

AI practices in their various forms (those incorporated in the design and development of AI and in their uses) are a part of apparatuses themselves, and combined with the traditions of different domains of apparatuses or *dispositifs* and the heterogeneous elements contained within those *dispositifs* which capture and subjectify (see [1]) persons and shape the fields of human possibility in terms of action and knowledge. AI systems too, fall within the *dispositif* and both emerge from and reinforce it (consider again algorithms feeding into the logic and execution of criminal justice systems, or even recommender systems telling us what we should watch or purchase next and why we should do so).

Here, we have assessed and examined the ways by which a VPD framework can fruitfully address ethical problems in the design and use of AI, or more specifically technical practices of AI. However, the existing *dispositifs* of AI practice, those which AI practices support and are supported by, may be so entrenched that they prejudice or limit possibilities for virtuous action by subjectifying humans in particular ways that are oriented around the maintenance of the *dispositif* and not in favour of the good life per se. The *dispositifs* arguably curtail the possibility of human agency, and structure the conditions of what Aristotle called *hexis*, the basis of human virtue as human disposition formed by habit and feeling [54]. The apparatus, the *dispositif*, may irresistibly limit or pre-determine the possibilities for moral development and action through complexes of various physical, legal, technological and cultural barriers.

So *dispositifs* may be so entrenched and destructive that they do not permit the possibility of meaningful ethical action or outcomes—the *dispositif* moulds the human disposition towards virtue or vice, or we might say that apparatus constrains *hexis*. Certainly, the expansion of *dispositifs* of security (Foucault noted that the security apparatuses had the tendency to expand, they were 'centrifugal' [35, p. 45]) is cause for some pessimism when one considers how it has been buoyed by and has spurred particular deleterious AI practices that include unfair and discriminatory profiling and threats to privacy [41]. The multiple levels of a 'little ethics' inspired VPD framework may help to reform the *dispositif*, and indeed a courageous and creative global citizenry may well be up to the task of reworking or resisting the *dispositif*, if we understand that the *dispositif* is something that can be meaningfully resisted on its own terms, and not something that requires more abstract and radical rethinking [19, 28, 34, 37]. A 'little ethics' based framework requires or implies a mobilisation of political effort, expansive dialogue and reform in the name of justice, and these are the tools carried by the VPD framework examined here. The question will persist as to whether those tools provided by VPD are radical enough to challenge lumbering, enduring and ancient *dispositifs* constructed under exploitative political economies that privilege few and harm many, or whether such tools are doomed to failure by operating within existing *dispositifs* where deeper critical reflection and systemic change may require yet a more radical approach to rethinking ethics in the 21st century [55]. Fundamentally, however, the *dispositif* is a challenge to the efficacy of VPD and any other ethical framework which controls the logic and incentives and disincentives to action. Arguably, the *dispositif* as we know it emerges from hegemonic narratives that coalesce into a free-market capitalist narrative that sets the costs and rewards for all action within the ethical sphere and influences global perspectives on the good life. Ultimately, perhaps we must operate within the structures of this narrative and change it from within the confines of industry (likely the only option for most people with any power to affect change) or attempt to dismantle it from outside whilst seeking alternative grand narratives that will cause a pivot of the *dispositif* into something more sustainable and compatible with a pluriverse [90]. Some might argue that VPD is more compatible with the former approach and trusts in slow but progressive change from within by presuming that human centres of power within practice can shape their practices towards ethical ends, but this is not to say that it does not itself permit the creation or advancement of radical practices opposed to the *dispositif* and hegemonic narratives. What can be argued with more certainty is that VPD does align with practices of decolonial design in its endorsing the attentive pursuit of plural narratives for ethical practice design, and in seeking out such narratives consistently and amplifying the voices of persons who are often overlooked and oppressed, a process of redistribution of power through expanded narratives may be reasonably plausible and can begin to challenge if not eventually subvert the grand hegemonic ones [5, 42, 52].

Within the *dispositif* all forms of ethics are aspirational and arguably the *dispositif* must bend for ethics to truly blossom—yet even within the logic of the *disposiif* as it is there remains a space where ethics fit, the marketable logic of acceptability, i.e. safe and robust systems that conform with ethical precepts are ones more likely to be accepted by society, and therefore bought into and sold. So long as there is space for this fit, even under a cynical logic, this space may provide opportunity for genuine ethical growth and forces of positive change.

Ultimately, the critiques of ethics to which VPD can fruitfully respond are serious. VPD is a promising approach, yet it warrants further reflection and testing going forward, as the problems it seeks to address are systemic and not minor. Nevertheless, it has an advantageous start in recognising that the field of ethics is not limited to individual human action, abstract and disconnected from milieux of rich context and history, but that a true and meaningful ethics is also one that exists in a space of other-regarding dialogue and within the purview of political action.

6 Conclusion

A virtuous practice design based framework incorporates a comprehensive ethical system ('little ethics') conceived by Paul Ricoeur that focuses on the cultivation of an ethical self that develops and pursues their life plans with due regard for the others with whom they share their environment and with whom their personal narratives intertwine into something more global, global and mediated by (just) institutional relations. Critical to the ethical intention and instrumental to its success are the virtues, those dispositions which help us to grow as ethical selves, to develop self-esteem and self-respect through our practices—that help us to think, act, and feel rightly in our pursuits and our relations with the other. Coupled with a narrative philosophy, it asks us to consider how new technical objects change our practices, as well as our understanding of ourselves and the world. A narrative investigation asks us to consider the ethical implications of new practices for those who are affected by them, and how they change their lives, and moreover how we might sustain the virtues within evolving and new technical practices. Unlike other ethical approaches, VPD asks us to immerse ourselves in the particular—to know the context of a practice incorporating new technological and AI tools down to basic actions and contemplate how these practices can be designed, fundamentally, with the ethical intention in mind, i.e., to live well, with and for others, in just institutions. What it asks of us, as AI developers and technical practitioners or global citizens more generally is rather demanding and serious—to be ethical individuals who are other-regarding and politically engaged if not activist. It arguably provides a more meaningful framework than others.

Ultimately, the VPD framework provides a suite of complex and useful resources to utilise in an ethics of artificial intelligence. A strength of VPD is its use of virtue ethics as its normative framework, ensuring that those who subscribe to it in AI practices pay attention to the particular, ethically salient features of situations of moral import. AI practices should be designed to support virtuous action, and moreover the design of AI should take place within environments that support virtuous action. The framework presented may not even necessarily be radically disruptive to the current status quo. It does not outright dispute the usefulness of principles and guidelines, as so-called moral norms, so long as those are accepted in and by a community of practice and are subordinate to the ethical action stemming from *phronetic* deliberation and even democratic discussion. It does not reject ethical principles and guidelines, it merely asks for more epistemic action from practitioners, both in terms of, essentially paying attention to context, as well as knowing themselves and the other. Such epistemic action supports the cultivation and exercise of the virtues, it can concretely support ethical action.

An additional strength of the VPD approach is its focus on narrative and a methodology of narrative investigation that emphasises a move from description to prescription—practices, including AI practices, are mapped and studied for the very actions and goals that define them in their variable instantiations, as well as the protagonists who appear in these stories, moral patients, who are ever at risk of being ignored by key decision-makers in the design and development of new technologies and their deployments across many meaningful contexts.

Whether or not a VPD approach provides tools which are radical enough to challenge existing *dispostifs* remains to be seen and requires further investigation. Certainly, useful resources are provided, and importantly resources which apply to individual action, to inter-personal interactions and even right up to grander regulatory decisions—the approach recognises that meaningful and lasting action is achieved only in multi-sectoral and even political dimensions, but whether it is enough to meet the *dispositif* essentially on the terms of the *dispositif* or to rethink ethics, politics, and the political economy anew is a deeper, worthy, and significant conversation that should not be overlooked in considering a necessary ethics of AI in uncertain times.

There remains future work to be done on this framework (and indeed, the others too) in the domain of business which will require pilot studies to determine how to best optimise ethical frameworks and methodologies to fast-paced business environments. The systematic implementation of narrative investigations and impact measurement of this process is another important topic of related research that should be undertaken in future such studies. Such pilot studies would also benefit from examining comparative and complementary approaches, including the specific methods of participatory (and decolonial) design and experimentation with tried and tested methods from value sensitive design.

Acknowledgements This research was partly funded by the ADAPT Centre which is funded under the SFI Research Centres Programme (Grant 13/RC/2106_P2) and is co-funded under the European Regional Development Funds. This research was partly funded by EC funded H2020 SwafS Project EUt+ EXTRAS (#101035812). This research was partly funded by the EC funded HE MSCA SE project EpisTeaM (#101129655). The Authors wish to express their gratitude for the significant and insightful guidance provided by the anonymous reviewers.

Funding Open Access funding provided by the IReL Consortium

Declarations

Competing interests Authors have no conflicts of interest or competing interests to declare.

References

1. Agamben, G.: What Is an Apparatus? and Other Essays, 1st edition. Stanford University Press, Stanford, Calif (2009)
2. Akpan, N.: The very real consequences of fake news stories and why your brain can't ignore them. In: PBS NewsHour. (2016). https://www.pbs.org/newshour/science/real-consequences-fake-news-stories-brain-cant-ignore. Accessed 8 Mar 2024
3. Algorithm Watch: How Dutch activists got an invasive fraud detection algorithm banned. In: AlgorithmWatch. (2020). https://algorithmwatch.org/en/syri-netherlands-algorithm/. Accessed 6 Feb 2024
4. Ali, S.J., Christin, A., Smart, A., Katila, R.: Walking the Walk of AI Ethics in Technology Companies. (2023). https://hai.stanford.edu/sites/default/files/2023-12/Policy-Brief-AI-Ethics_0.pdf. Accessed 6 Feb 2024
5. Alvarado Garcia, A., Maestre, J.F., Barcham, M., Iriarte, M., Wong-Villacres, M., Lemus, O.A., Dudani, P., Reynolds-Cuéllar, P., Wang, R., Cerratto Pargman, T.: Decolonial Pathways:

Our Manifesto for a Decolonizing Agenda in HCI Research and Design. In: Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, pp 1–9 (2021)

6. Alzola, M.: Corporate roles and Virtues. In: Sison, A.J.G., Beabout, G.R., Ferrero, I. (eds.) Handbook of Virtue Ethics in Business and Management, pp. 47–56. Springer Netherlands, Dordrecht (2017)
7. Ananny, M., Crawford, K.: Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. New. Media Soc. **20**, 973–989 (2018). https://doi.org/10.1177/1461444816676645
8. Anderson, S.L., Anderson, M.: AI and ethics. AI Ethics. **1**, 27–31 (2021). https://doi.org/10.1007/s43681-020-00003-6
9. Angwin, J., Larson, J., Matu, S., Kirchner, L.: Machine Bias. In: ProPublica. (2016). https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. Accessed 19 Oct 2018
10. Aristotle: The Nature of Virtue. In: Shafer-Landau R (ed) Ethical Theory: An Anthology, 2nd Edition, 2nd edition. Wiley-Blackwell, Chichester, West Sussex; Malden, MA, pp 615–629 (2012)
11. Atkins, K. (nd) Paul Ricouer (ed.): https://iep.utm.edu/ricoeur/. Accessed 24 Nov 2022
12. Audi, R.: The Good in the Right: A Theory of Intuition and Intrinsic Value. Princeton University Press, Princeton, NJ (2005)
13. Audi, R.: Virtue Ethics as a resource in business. Bus. Ethics Q. **22**, 273–291 (2012)
14. Beauchamp, T.L., Childress, J.F.: Principles of Biomedical Ethics, Seventh Edition. Oxford University Press, Oxford, New York (2013)
15. Bergen, J.P., Robaey, Z.: Designing in Times of uncertainty: What Virtue Ethics can bring to Engineering Ethics in the twenty-First Century. In: Dennis, M.J., Ishmaev, G., Umbrello, S., van den Hoven, J. (eds.) Values for a Post-Pandemic Future, pp. 163–183. Springer International Publishing, Cham (2022)
16. Buddemeyer, A., Nwogu, J., Solyst, J., Walker, E., Nkrumah, T., Ogan, A., Hatley, L., Stewart, A.: Unwritten Magic: Participatory Design of AI Dialogue to Empower Marginalized Voices. In: Proceedings of the 2022 ACM Conference on Information Technology for Social Good. ACM, Limassol Cyprus, pp 366–372 (2022)
17. Buijsman, S., Klenk, M., van den Hoven, J. (forthcoming), Smuha, N. (eds.): Cambridge Handbook on the Law, Ethics and Policy of AI. Cambridge University Press
18. Burton, E., Goldsmith, J., Mattei, N., Siler, C., Swiatek, S.-J.: Computing and Technology Ethics: Engaging through Science Fiction. MIT Press, Cambridge, Massachusetts (2023)
19. Callewaert, S.: Foucault's Concept of Dispositif. Prakt Grunde 29–52 (2017)
20. Carney, E.: Depending on practice: Paul Ricoeur and the Ethics of Care. Ateliers Léthique Ethics Forum. **10**, 29–48 (2015). https://doi.org/10.7202/1037650ar
21. Chen, B.X.: How to Use ChatGPT and Still Be a Good Person. In: N. Y. Times. (2022). https://www.nytimes.com/2022/12/21/technology/personaltech/how-to-use-chatgpt-ethically.html. Accessed 13 March 2024
22. Chen, J.-Y.: Virtue and the scientist: Using Virtue Ethics to Examine Science's ethical and Moral challenges. Sci. Eng. Ethics. **21**, 75–94 (2015). https://doi.org/10.1007/s11948-014-9522-3
23. Cheong, C.: ChatGPT helped me renovate my kitchen. Here's how it saves me time on everyday tasks outside of work. In: Bus. Insid. (2024). https://www.businessinsider.com/how-use-chatgpt-daily-life-work-save-time-2024-3. Accessed 8 Mar 2024
24. Coeckelbergh, M.: Time Machines: Artificial Intelligence, process, and narrative. Philos. Technol. **34**, 1623–1638 (2021). https://doi.org/10.1007/s13347-021-00479-y
25. Coeckelbergh, M.: Narrative responsibility and artificial intelligence. AI Soc. doi. (2021). https://doi.org/10.1007/s00146-021-01375-x
26. Crawford, K.: Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press, New Haven, London (2022)
27. D`ignazio, C., Klein, L.F.: Data Feminism. MIT Press, Cambridge, MA (2020)
28. Deleuze, G.: Dispositif(Apparatus). In: Nale, J., Lawlor, L. (eds.) The Cambridge Foucault Lexicon, pp. 126–132. Cambridge University Press, Cambridge (2014)
29. Escobar, A.: Designs for the Pluriverse: Radical Interdependence, Autonomy, and the Making of Worlds. Duke University Press, Durham, NC (2018)
30. Eubanks, V.: Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's, New York, NY (2018)
31. Ferguson, A.G.: The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement. NYU, New York, NY (2017)
32. Fitzpatrick, N., Kelleher, J.: On the exactitude of Big Data: La Bêtise and Artificial Intelligence. La. Deluziana. (2018). https://doi.org/10.21427/dfw8-m918
33. Floridi, L., Cowls, J.: A Unified Framework of five principles for AI in Society. In: Machine Learning and the City, pp. 535–545. John Wiley & Sons, Ltd, Hoboken, NJ (2022)
34. Foucault, M.: Power/Knowledge: Selected Interviews and Other Writings, 1972–1977. Random House USA Inc, New York, NY (1980)
35. Foucault, M., Ewald, F., Fontana, A., Davidson, A.I.: Security, Territory, Population: Lectures at the Collège De France 1977–1978, First Edition. Picador, New York, NY (2009)
36. Friedman, B., Kahn, P.H., Borning, A., Huldtgren, A.: Value Sensitive Design and Information systems. Early Engagem. New. Technol. Open. Lab. 55–95 (2013). https://doi.org/10.1007/978-94-007-7844-3_4
37. Frost, T.: The Dispositif between Foucault and Agamben. Law Cult. Humanit. **15**, 151–171 (2019). https://doi.org/10.1177/1743872115571697
38. Hacker, P.: Comments on the Final Trilogue Version of the AI Act. (2024). https://www.europeannewschool.eu/images/chairs/hacker/Comments%20on%20the%20AI%20Act.pdf. Accessed 13 March 2024
39. Hayes, P., Fitzpatrick, N.: Narrativity and responsible and transparent ai practices. AI Soc. doi. (2024). https://doi.org/10.1007/s00146-024-01881-8
40. Hayes, P., Jackson, D.: Care ethics and the responsible management of power and privacy in digitally enhanced disaster response. J. Inf. Commun. Ethics Soc. **18**, 157–174 (2020). https://doi.org/10.1108/JICES-02-2019-0020
41. Hayes, P., van de Poel, I., Steen, M.: Algorithms and values in justice and security. AI Soc. **35**, 533–555 (2020). https://doi.org/10.1007/s00146-019-00932-9
42. Hayes, P., van de Poel, I., Steen, M.: Moral transparency of and concerning algorithmic tools. AI Ethics. **3**, 585–600 (2023). https://doi.org/10.1007/s43681-022-00190-4
43. Heidegger, M.: The Question Concerning Technology: and Other Essays, Reissue Edition. Harper Perennial, New York; London; Toronto (2013)
44. Heikkla, M.: Dutch scandal serves as a warning for Europe over risks of using algorithms. In: POLITICO. (2022). https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/. Accessed 6 Feb 2024
45. Held, V.: The Ethics of Care: Personal, Political, and Global: Personal, Political, Global, New Ed Edition. Oxford University Press, New York; Oxford (2007)

46. Hickok, M.: Lessons learned from AI ethics principles for future actions. AI Ethics. **1**, 41–47 (2021). https://doi.org/10.1007/s43681-020-00008-1
47. Howard, D.: Virtue in Cyberconflict. In: Floridi, L., Taddeo, M. (eds.) The Ethics of Information Warfare, pp. 155–168. Springer International Publishing, Cham (2014)
48. Hursthouse, R.: Normative Virtue Ethics. In: Shafer-Landau R (ed) Ethical Theory: An Anthology, 2nd Edition, 2nd edition. Wiley-Blackwell, Chichester, West Sussex; Malden, MA, pp 645–652 (2012)
49. Ihde, D.: Technology and the Lifeworld: from Garden to Earth. Indiana University Press, Bloomington (1990)
50. Interian, R., Marzo, G., Mendoza, R., Ribeiro, I. CC: Network polarization, filter bubbles, and echo chambers: An annotated review of measures and reduction methods. Int. Trans. Oper. Res. **30**, 3122–3158 (2023). https://doi.org/10.1111/itor.13224
51. Jobin, A., Ienca, M., Vayena, E.: The global landscape of AI ethics guidelines. Nat. Mach. Intell. **1**, 389–399 (2019). https://doi.org/10.1038/s42256-019-0088-2
52. Kambunga, A.P., Smith, R.C., Winschiers-Theophilus, H., Otto, T.: Decolonial design practices: Creating safe spaces for plural voices on contested pasts, presents, and futures. Des. Stud. **86**, 101170 (2023). https://doi.org/10.1016/j.destud.2023.101170
53. Kaplan, D.M.: Paul Ricoeur and the philosophy of Technology. J. Fr. Francoph Philos. **16**, 42–56 (2006). https://doi.org/10.5195/jffp.2006.182
54. Kraut, R.: Aristotle's Ethics. In: Zalta, E.N., Nodelman, U. (eds.) The Stanford Encyclopedia of Philosophy, Fall 2022. Metaphysics Research Lab, Stanford University (2022)
55. Krzykawski, M., Lindberg, S.: Ēthos and Technology. In: Stiegler, B. (ed.) Bifurcate: There is no Alternative, pp. 195–219. Open Humanities, London (2021)
56. Kudina, O.: Alexa, who am I? Voice Assistants and Hermeneutic Lemniscate as the technologically mediated sense-making. Hum. Stud. **44**, 233–253 (2021). https://doi.org/10.1007/s10746-021-09572-9
57. Liao, Q.V., Muller, M.: Enabling Value Sensitive AI Systems through Participatory Design Fictions. (2019). http://arxiv.org/abs/1912.07381. Accessed 1 Feb 2024
58. MacIntyre, A.: After Virtue, Reprint Edition. Bloomsbury Academic, London (2013)
59. May, J., Kumar, V.: Moral reasoning and emotion. In: Jones, K., Timmons, M., Zimmerman, A. (eds.) Routledge Handbook on Moral Epistemology, pp. 139–156. Routledge, New York, NY; Abingdon, Oxon (2018)
60. May, T.: Subjectification. In: Nale, J., Lawlor, L. (eds.) The Cambridge Foucault Lexicon, pp. 496–502. Cambridge University Press, Cambridge (2014)
61. Melnyk, A., Edmonds, B., Ghorbani, A., van de Poel, I.: Editorial: Modelling values in Social, Technical, and Ecological systems. J. Artif. Soc. Soc. Simul. **27** (2024). https://doi.org/10.18564/jasss.5361
62. Melton, M.: The top 12 people in artificial-intelligence policy, ethics, and research. In: Bus. Insid. (2023). https://www.businessinsider.com/ai-100-top-12-people-policy-ethics-and-research-2023-11. Accessed 7 Feb 2024
63. Mittelstadt, B.: Principles alone cannot guarantee ethical AI. Nat. Mach. Intell. **1**, 501–507 (2019). https://doi.org/10.1038/s42256-019-0114-4
64. Munn, L.: The uselessness of AI ethics. AI Ethics doi. (2022). https://doi.org/10.1007/s43681-022-00209-w
65. Nascimento, F.: Technologies, narratives, and practical wisdom. Études Ricoeuriennes Ricoeur Stud. **10**, 21–35 (2019). https://doi.org/10.5195/errs.2019.481
66. van Nistelrooij, I., Schaafsma, P., Tronto, J.C.: Ricoeur and the ethics of care. Med. Health Care Philos. **17**, 485–491 (2014). https://doi.org/10.1007/s11019-014-9595-4
67. Noddings, N.: An Ethic of Caring. In: Shafer-Landau R (ed) Ethical Theory: An Anthology, 2nd Revised edition edition. John Wiley & Sons, Chichester, West Sussex; Malden, MA, pp 699–712 (2012)
68. Nussbaum, M.C.: Upheavals of Thought: The Intelligence of Emotions, 1st edition. Cambridge University Press, Cambridge (2003)
69. Oritz, S.: 6 ways ChatGPT can make your everyday life easier. In: ZDNET. (2024). https://www.zdnet.com/article/5-ways-chatgpt-can-save-you-time-in-the-new-year/. Accessed 8 Mar 2024
70. Parfit, D.: Equality and Priority. Ratio. **10**, 202–221 (1997). https://doi.org/10.1111/1467-9329.00041
71. Pinto-Garay, J.: Virtue Ethics in Business: Scale and Scope. In: Business Ethics, pp. 67–86. Emerald Publishing Limited (2019)
72. van de Poel, I.: Translating values into design requirements. In: Michelfelder, D.P., McCarthy, N., Goldberg, D.E. (eds.) Philosophy and Engineering: Reflections on Practice, Principles and Process, pp. 253–266. Springer Netherlands, Dordrecht (2013)
73. van de Poel, I.: Design for value change. Ethics Inf. Technol. **23**, 27–31 (2021). https://doi.org/10.1007/s10676-018-9461-9
74. van de Poel, I., Royakkers, L., Zwart, S.D.: Moral Responsibility and the Problem of Many Hands, 1 edition. Routledge, New York (2015)
75. Reijers, W., Coeckelbergh, M.: Narrative and Technology Ethics, 1st ed. 2020 edition. Palgrave Macmillan, Basingstoke (2020)
76. Reijers, W., Gordijn, B.: Moving from value sensitive design to virtuous practice design. J. Inf. Commun. Ethics Soc. **17**, 196–209 (2019). https://doi.org/10.1108/JICES-10-2018-0080
77. Reijers, W., Romele, A., Coeckelbergh, M.: Interpreting Technology: Ricoeur on Questions Concerning Ethics and Philosophy of Technology. Rowman & Littlefield, Lanham (2021)
78. Reijers, W., Wright, D., Brey, P., Weber, K., Rodrigues, R., O'Sullivan, D., Gordijn, B.: Methods for Practising Ethics in Research and Innovation: A literature review, critical analysis and recommendations. Sci. Eng. Ethics. **24**, 1437–1481 (2018). https://doi.org/10.1007/s11948-017-9961-8
79. Rességuier, A., Rodrigues, R.: AI ethics should not remain toothless! A call to bring back the teeth of ethics. Big Data Soc. **7**, 2053951720942541 (2020). https://doi.org/10.1177/2053951720942541
80. Richardson, R., Schultz, J., Crawford, K.: Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice. Social Science Research Network, Rochester, NY (2019)
81. Ricoeur, P.: Time and Narrative, Volume 1: v. 1, New Edition. University of Chicago Press, Chicago, IL (1990)
82. Ricoeur, P.: Time and Narrative, Volume 2, New Edition. University of Chicago Press, Chicago, IL (1990)
83. Ricoeur, P.: Time and Narrative, Volume 3: v. 3, New Edition. University of Chicago Press, Chicago, IL (1990)
84. Ricoeur, P.: Oneself as another. University of Chicago Press, Chicago, IL (1995)
85. Ricoeur, P.: Reflections on the Just. ReadHowYouWant, Chicago, IL; London (2011)
86. Roeser, S.: Moral Emotions and Intuitions. Palgrave Macmillan UK, Basingstoke, New York (2011)
87. Roeser, S.: Risk, Technology, and Moral Emotions, 1 Edition. Routledge, New York (2017)
88. Romele, A., Severo, M., Furia, P.: Digital hermeneutics: From interpreting with machines to interpretational machines. AI Soc. **35**, 73–86 (2020). https://doi.org/10.1007/s00146-018-0856-2
89. Ross, D.: The Right And The Good, 2 edition. Oxford University Press, U.S.A., Oxford (2003)

90. Sætra, H.S., Coeckelbergh, M., Danaher, J.: The AI ethicist's dilemma: Fighting Big Tech by supporting big tech. AI Ethics. **2**, 15–27 (2022). https://doi.org/10.1007/s43681-021-00123-7
91. Simondon, G.: On the Mode of Existence of Technical Objects. Univ Of Minnesota, Minneapolis (2017)
92. Solomon, R.C.: Business Ethics and Virtue. In: A Companion to Business Ethics, pp. 30–37. Wiley, Ltd (1999)
93. Steen, M.: Slow Innovation: The need for reflexivity in responsible Innovation (RI). J. Responsible Innov. **8**, 254–260 (2021). https://doi.org/10.1080/23299460.2021.1904346
94. Steen, M.: Ethics for People Who Work in Tech, 1st edition. Chapman and Hall/CRC, Boca Raton, FL; Abingdon, Oxon (2022)
95. Steen, M., Sand, M., Van de Poel, I.: Virtue Ethics for responsible Innovation. Bus. Prof. Ethics J. (2021). https://doi.org/10.5840/bpej2021319108
96. Steinert, S., Roeser, S.: Emotions, values and technology: Illuminating the blind spots. J. Responsible Innov. **7**, 298–319 (2020). https://doi.org/10.1080/23299460.2020.1738024
97. Stiegler, B.: Technics and Time, 1: The Fault of Epimetheus, 1st edition. Stanford University Press, Stanford, Calif (1998)
98. Swanton, C.: A Virtue Ethical Account of Right Action. In: Shafer-Landau R (ed) Ethical Theory: An Anthology, 2nd Edition, 2nd edition. Wiley-Blackwell, Chichester, West Sussex; Malden, MA, pp 664–675 (2012)
99. Szakacs, J., Bognar, E.: The impact of disinformation campaigns about migrants and minority groups in the EU. (2021). https://www.europarl.europa.eu/thinktank/en/document/EXPO_IDA(2021)653641. Accessed 6 Feb 2024
100. Teal, M.: The Ethics of College Students Using ChatGPT. In: Ethics Policy. (2023). https://ethicspolicy.unc.edu/news/2023/04/17/the-ethics-of-college-students-using-chatgpt/. Accessed 8 Mar 2024
101. Timmons, M.: Toward a Sentimentalist Deontology. (2007). https://doi.org/10.7551/mitpress/7504.003.0021
102. Timmons, M.: Moral Theory: An Introduction, 2 edition. Rowman & Littlefield Publishers, Lanham, Md (2012)
103. Tommasel, A., Menczer, F.: Do Recommender Systems Make Social Media More Susceptible to Misinformation Spreaders? In: Proceedings of the 16th ACM Conference on Recommender Systems. Association for Computing Machinery, New York, NY, USA, pp 550–555 (2022)
104. Tuama, D.: What is an AI Developer & How to Become One? In: Code Inst. IE. (2023). https://codeinstitute.net/ie/blog/ai-developer/. Accessed 7 Feb 2024
105. Vallor, S.: Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting, Reprint Edition. Oxford University Press, New York, NY (2018)
106. Van Stichel, E.: Love and Justice's dialectical relationship: Ricoeur's contribution on the relationship between care and justice within care ethics. Med. Health Care Philos. **17**, 499–508 (2014). https://doi.org/10.1007/s11019-013-9536-7
107. Verbeek, P.P.: Toward a theory of Technological Mediation A Program for Postphenomenological Research. In: Friis, J.K.B.O., Crease, R.C. (eds.) Technoscience and Postphenomenology: The Manhattan Papers, pp. 189–204. Lexington Books, London (2016)
108. World Health Organisation: Infodemics and misinformation negatively affect people's health behaviours, new WHO review finds. (2022). https://www.who.int/europe/news/item/01-09-2022-infodemics-and-misinformation-negatively-affect-people-s-health-behaviours--new-who-review-finds. Accessed 8 Mar 2024
109. Wright, D.: A framework for the ethical impact assessment of information technology. Ethics Inf. Technol. **13**, 199–226 (2011). https://doi.org/10.1007/s10676-010-9242-6