Journal of Responsible Technology 22 (2025) 100115

Contents lists available at ScienceDirect

Journal of Responsible Technology

journal homepage: www.sciencedirect.com/journal/journal-of-responsible-technology

Research Article

# Reflexivity and AI start-ups: A collective virtue for dynamic teams

Marco Innocenti *

*University of Milan and Paris 1 Pantheon-Sorbonne University, France*

ARTICLE INFO

*Keywords:*
Responsible innovation
Ai ethics
Virtue ethics
Swarm intelligence
Organization studies

ABSTRACT

This paper investigates the ethical challenges faced by AI-driven start-ups, where the rapid pace of innovation and limited resources often preclude team members from fully understanding the product under development or its societal implications. We propose the concept of "swarm moral reflexivity", where ethical reflection emerges collectively from the interactions of individuals focused on their specific tasks. Drawing on Swarm Intelligence theories and Alasdair MacIntyre's framework of moral deliberation, this approach enables teams to engage with ethical issues through daily encounters with conflicting responsibilities, rather than relying on top-down value systems or comprehensive ethical oversight. Our model suggests that decentralised, collective moral awareness can effectively support Responsible Innovation in AI start-ups, ensuring that ethical concerns are recognised and addressed throughout the development process, even in fast-paced and resource-constrained environments.

## 1. Introduction

Software engineering presents unique ethical challenges compared to other professional fields like medicine or business. As Zuber et al. (2022) explain, the actions and decisions made by software engineers are directly translated into technical requirements that may be replicated millions of times across products. Additionally, because software engineering lacks the institutional sanctions and professional commitment to public welfare seen in other fields, development teams should be equipped to raise and face ethical concerns autonomously. On the other hand, values are often taken as the "key starting point to enable normative deliberation in software engineering" (Zuber et al., 2022). They form the basis of the Codes of Conduct of these professions, presenting a set of self-evident values to be pursued. Bezuidenhout and Ratti (2020) call this approach "macroethics", describing it as a form of data and AI ethics that prioritise abstract principles such as "identity and subjectivity, social exclusion and inequality, politics and democracy […]". Besides their "high degree of abstraction" (Zuber et al., 2024), the limitations of this approach include the inability to take into account cases in which these values come into conflict (e.g. transparency and privacy), and the difficulty of identifying "normative desirable products" from an "endless list of values" specific to each type of technology (Zuber et al., 2022). Even when these principles are translated into specific 'rules' to follow in an attempt to counteract their abstract nature, the result is an approach to ethics that emphasises "mindless compliance" over the cultivation of autonomous ethical agents (Bezuidenhout & Ratti, 2020). Moreover, this top-down and "substantialist" orientation towards values does not allow for the responsiveness to value change necessary when socially disruptive technological innovations are developed (cf. Boenink & Kudina, 2020), as in the case of Artificial Intelligence and its applications.

Hence, there is a need to find a 'procedural' alternative for this normative deliberation in work teams that develop neural networks or implement AI in existing products, thus supporting what Hagendorff (2022) calls a "practical turn" in AI ethics. In the face of the ethical concerns of these technologies (see Mittelstadt et al., 2016), a challenge for AI ethics is to provide tools and methodologies to put into practice the principles by which these concerns are to be countered. In other words, AI ethics should "ensure that there is a 'how' for every 'what' at each stage of the Machine Learning pipeline" (Morley et al., 2020). More specifically, in the context of R&D teams, AI ethics should see to it that practitioners are able and willing to "continuously evaluate each new design project from its inception, throughout the development process, its deployment, operations, and its maintenance" (Zuber et al., 2024). According to Ratti and Graves (2021), "we need to cultivate ethics as a form of reflection, rather than as an effort to compile lists of rules". Thus, their concept of "microethics" is framed as "a form of ethical training and exercise in which the goal for a data scientist is to learn how to identify the ethical relevance of his/her day-to-day activities" (Ratti & Graves, 2021). While Bezuidenhout & Ratti, 2021 apply this model to teaching data science virtue ethics, Ratti and Graves (2021) present a proposal for developing the relevant "moral abilities" of virtuous data

* Corresponding author: 6 Rue Guenot, Paris 75011, France.
*E-mail address:* marco.innocenti@unimi.it.

https://doi.org/10.1016/j.jrt.2025.100115

Available online 14 March 2025

scientists within their professional environment. In advocating for the teaching of ethics "in the same manner as technical skills," they also explore how to facilitate ethical reflection within the working processes that underpin this role on a daily basis. Concerning start-up environments, which will be central in this article, we should notice how this discourse aligns with Ryan et al.'s (2024) indications for applying Responsible Innovation in this context, where the importance of "proactive", "positive duties", namely practices promoting responsibility in the team, are privileged. These practices, promoting an "*ethos*" that professionals can recognise as their own (Zuber et al., 2022), include, for instance, raising awareness of the ethical implications of their technologies and promoting activities where they can be weighted (Ryan et al., 2024). This "micro-virtue ethics" perspective, as Bezuidenhout & Ratti, 2021 call it, differs significantly from the "preventive" or "negative" approaches that characterise deontological and consequentialist professional ethics (Schmidt, 2014).

According to Gogoll et al. (2021), "an approach built on an ethical deliberation of the software engineer" may be a viable path to enable them to develop "ethically sound" software. For example, the Value Sensitive Design (VSD) methodology, which integrates stakeholder values into the design process, can effectively align with the Agile project management approach without compromising its 'agility'. Directions for doing so include defining a clear hierarchy among values (Winkler, 2020) or adding the VSD value representation tools to the Agile workflow (Umbrello & Gambelin, 2023) to iteratively analyse the relevant values throughout the R&D process (see Umbrello & van de Poel, 2021 for an application of this analysis to AI technologies). The goal is to make design processes more responsive to the ever-changing ethical concerns of stakeholders, thereby allowing for a more flexible and participatory solution to moral decision-making in software development. Without relying exclusively on the VSD framework, Zuber et al. (2022) proposed implementing "techno-ethical deliberation" directly within the Agile methodology, emphasising the latter's practicality in promoting autonomous moral reasoning in the members and the pooling of these reflections, pouring them into product design.

In this paper, we will continue this investigation into the work environment and the organisational characteristics required for this ethical deliberation to take place even in the absence of a dedicated ethicist (as advocated by McLennan et al., 2020). We will consider the thriving context of AI start-ups, which often lack the resources and interest to hire this figure. In particular, we will focus on a precise issue, namely the viability of moral reflexivity (defined in terms of a 'procedural' Responsible Innovation) in a context where there is neither the time nor the possibility for the individual members of these teams to have an overall picture of the product being developed and its production process. This analysis thus differs from that of Zuber et al. (2024) in that we do not intend to identify the starting point of ethical judgement in broad questions such as "Is a world desirable where everyone has access to this technology?" but in the discrepancies that the individual may feel between her role responsibility and her other societal responsibilities in their daily, practical work.

The aforementioned lack of time and predictability in AI start-ups can be attributed both to the frenetic context of the development and application of artificial intelligence and the experimental nature of many of their innovations (§2). In §3, we will compare this issue with the literature devoted to 'collective virtues', suggesting that moral reflexivity can emerge in a collective through the interaction between agents. Drawing on the 'Swarm Intelligence' model, we will then propose a 'swarm moral reflexivity'. It would describe the reflexivity on the goods pursued by the whole production process emerging from a group of agents who consider that only from the perspective of their specific tasks. Through the work of Alasdair MacIntyre, we will see in §4 how moral reflexivity (duly distinguished from 'practical reasoning') can rise even without adopting an overall contemplation of one's own actions. Instead, it can emerge from the conflict between the 'goods' or responsibilities pursued in our everyday life. In §5, finally, we will present a coherent model for embedding this type of moral reflexivity within a team of developers so that members are receptive to moral conflicts that affect their work and report (or signal) them to colleagues, who, in turn, would be active in evaluating and communicating them to others.

## 2. Responsible innovation and AI start-ups

According to Brem et al. (2023, p. 774), in recent years, we have witnessed exponential growth in the number of start-ups "that build their value proposition around an advance in AI research (e.g., new deep learning technique, new visual recognition or computer vision algorithm, or a novel combination of existing technologies)". This AI-related boom in start-up investments counterpoints the general crisis that hit this form of enterprises over the same period, i.e., a "broader start-up downturn" (Griffith, 2024). As noted by Tekic and Füller (2023), this trend is stimulated by the rise of investments by venture capitalists, but also by tech giants such as "Apple, Google, Microsoft, Facebook, and Amazon". Alongside these are "[c]ompanies from the automotive, biotech, retail, banking, and even oil and gas sectors", also seeking partnerships or acquisitions. This is due to the highly uncertain nature of innovations related to this type of technology in the early adoption phase, which "asks for experimentation, flexibility, and agility - favouring start-ups and their way of operation". Conversely, partnering with advanced technology firms (such as IBM or Microsoft) or specialised consultancies offers companies early access to new technologies and valuable implementation expertise (Tekic & Füller, 2023).

Despite their prominence in today's innovation landscape, the RI ('Responsible Innovation') paradigm finds it hard to apply to these types of companies, characterised by a fluid organisation and a regime of instability (Burnell et al., 2023; Deuten & Rip, 2000). As Ryan et al. (2023) point out, these characteristics may require different strategies for "responsible behaviour" than those originally conceived for the academic research environment and large organisations. It should also be mentioned that the intrinsic unpredictability of start-up development often leads to disregarding ethical concerns. This is especially true when these practices are at odds with the pressure from investors to produce results and achieve a successful market impact (Ustek-Spilda et al., 2019) or involve costs that are difficult to justify in this type of organisation (Hayes et al., 2024). Furthermore, add to this the feverish climate of AI-driven innovation, which is taking hold in all sectors at a frenetic pace. This type of development has been compared to the advent of electricity or the Internet (Füller et al., 2022), but experts also note that its 'adoption curve' is much narrower. According to Cooper (2024, p. 63), the 'Adoption Window' of these groundbreaking technologies has become increasingly narrower over time, "from 85 years [for] the 1st to 28 years for the 3rd Industrial Revolution to today's 4th Industrial Revolution. The trend line shows AI's Adoption Window will be only 13–15 years".

Whether creating a new AI software or adapting an AI to an IoT product, the race for this type of innovation seems diametrically opposed to the 'Slow Innovation' advocated by Marc Steen (2021, p. 258) as a model for a 'human-centric' approach to the design of the AI in the European Union, and "an alternative to the US' company-driven AI or China's state-driven AI". This Human-Centred AI approach is the attempt to 'keep the human in the loop' of the development process of artificial intelligence, be it at the level of the individual, society (Rahwan, 2018), or organisations (Herrmann & Pfeiffer, 2023). This is further expressed in the image, used by some researchers, of a symbiosis of the human and the machine's computational capabilities, both on an individual level (Horvatić & Lipic, 2021) and in a collective, social sphere (Xu et al., 2023). In more concrete terms, some HCAI ('Human-Centered AI') theorists have emphasised the need for a focus on the societal impact of AI (Qian et al., 2024), on its explainability (Riedl, 2019), or on its user-oriented functionality (Bingley et al., 2023), whereas Auernhammer (2020) advocates a "comprehensive or pan-disciplinary design approach". Whichever approach is preferred, we

believe, going back to Steen (2021), that the starting point for its successful application is the introduction of a reflexive attitude to the innovation process, which, as we will see in the following paragraph, corresponds to one of the elements of the procedural approach to Responsible Innovation.

### 2.1. Reflexivity in RI

Of the various formulations of Responsible Innovation, we focus here on the one proposed by Stilgoe et al. (2013). This has been defined as a "procedural" approach by Boenink and Kudina (2020) since it relies on a process of assessing, confronting, and discussing the different values of the various shareholders and stakeholders, as well as their expression in the innovation. In this framework, RI is defined as the integration of four key features in the innovation process, namely 'anticipation', 'reflexivity', 'inclusion', and 'responsiveness' (it is also called 'AIRR framework'). These four terms, designed for governing innovation processes, have been discussed for their suitability for the private sector by Lubberink et al. (2017). As it might be clear, our paper limits itself to analysing the 'reflexivity' dimension, defined as the scrutiny of "commitments and assumptions" and "value systems and theories" that guide the innovation process. That implies assessing their appropriateness with the specific situation and challenges, as well as their effects on the outcome (Lubberink et al., 2017, p. 1571; Stilgoe et al., 2013). This is a dimension often neglected in the professional sphere, although the dedication to objectivity and thoughtful adherence to professional standards are highly significant in both science and engineering (Martin, 2002, pp. 547, 548). Moreover, according to the review of 'reflexivity' in business organisations traced by Lubberink et al. (2017), although factors like profit and legislation remain the key drivers for the introduction of sustainability considerations in innovation contexts, this does not exclude the possibility of "conflicting options" for which a "personal ethics" remains indispensable. As we read in this article, therefore, "innovators should not only live up to their *role responsibility* but also their *wider moral responsibilities*" (our italics).

This 'reflexivity' on the values endorsed by a team's activity would be part of the professional virtues pledged by Harris (2008, p. 158) and, more specifically, one of the "non-technical excellences" that "include not only such traditional virtues as honesty and professional integrity, but also virtues like concern for the environment and a sensitivity to the social implications of technology". According to him, these virtues should belong to engineers and designers, especially those developing new technologies to be released into society with potentially disruptive effects. The question we will be investigating throughout this paper, as we mentioned in §1, is whether and how this virtue can be attributed to an R&D team even when, individually, no one (or only a few) of its members can dedicate themselves to this kind of activity, given the complexity of the task and the lack of time and resources. To understand how to cope with this question, which we deem highly relevant in the context of hi-tech start-ups, we will first collocate this discourse in the field of collective virtues and then try to delineate the characteristics of an organisation (or a group) suitable for cultivating high-level moral reflexivity as a team.

Except for those points where we will clearly refer to an individual attitude, we will use the noun 'reflexivity' (and the adjective 'reflexive') instead of 'reflectivity' (and 'reflective') due to the former's greater suitability for describing self-referential attitudes in both human and non-human entities. While 'reflectivity' implies reflexivity, it conveys a deeper human attitude of self-analysis, which we will explore further in the context of MacIntyre's thought. However, when we examine the way theorists of the AIRR framework have positioned it as a cornerstone of Responsible Innovation (RI), it becomes clear that 'reflexivity' cannot be limited to mere self-referentiality. Instead, it has been conceptualised as encompassing the critical examination of "commitments and assumptions", as well as the "value systems and theories" that guide the innovation processes (Lubberink et al., 2017).

## 3. Collective virtues

In 2005, Miranda Fricker inaugurated the literature on collective virtues by discussing the social ontology of Margaret Gilbert (1989). As in the former's article, we should begin this section by defining the nature of the group we refer to or, rather, the angle from which a given group is considered. As Fricker does, we might understand a group as a collective, namely "the group considered as non-reducible to its component individuals" (nor to a "summative account" of them). For clarity, the collective under discussion corresponds to what French (1984) designates as a "conglomerate collectivity", as opposed to "aggregate collectivities". The latter refers to groups formed by circumstance (such as individuals waiting at a bus stop) or by shared characteristics, as in French's example of 'white American racists'. In such cases, moral responsibility can only be ascribed to individual members rather than to the group as a whole (French, 1984, pp. 12, 13). Conglomerate collectivities, by contrast, not only bring individuals together but also assign them defined roles, establish "standards of conduct" and 'of excellence', and implement "design procedures by which courses of concerted actions can be, though not necessarily are, chosen" (French, 1984, pp. 13, 14). Within a business context, this kind of collective is often referred to as a "corporate actor" (Gowri, 2007) or a "corporate agent" (Björnsson & Hess, 2017). While these terms are particularly suited to this field, and French's vocabulary may provide an alternative for clarity, the term 'collective' will be used for the sake of simplicity.

That said, there is little point in reconstructing the discourse that various authors have made regarding the possibility that some virtues may be "*genuinely* collective virtues, or *irreducible* collective virtues" (Byerly & Byerly, 2016), i.e. virtues attributable to a group that do not simply correspond to the virtues of its members. Instead, it will be appropriate to define the position of this article in this context, thus understanding which aspects of this debate are most relevant to this investigation. First of all, we do not intend to theoretically investigate the collective virtues in general, as did Fricker's (2005) article mentioned above, Cordell's (2017) response, and the more recent proposal of a 'theory of a Collective Virtue' advanced by Baddorf and McKay (2024). Nonetheless, it is necessary to briefly explain why virtues can be ascribed to collective entities and why this may be a reasonable position to adopt.

Sandin (2007) identifies four main arguments supporting this claim, later categorised by Sandin (2012) as the "linguistic argument", the "stability argument", the "capability argument", and the "pragmatic argument". The first argument, which highlights how we commonly speak "as if some types of collectives can display traits of character", does not in itself establish that collectives possess virtues. However, it offers "circumstantial evidence" suggesting that the concept of collective virtues may be meaningful. The 'stability argument' brings this discussion into dialogue with the 'situationist critique' of virtue ethics, as well-organised collectives can demonstrate patterns of behaviour that are more consistent over time than those of individuals. Beyond a more structured decision-making process, making them able to respond to varying circumstances, some collectives also possess "larger information-processing capabilities" than individuals, including in ethical deliberation, as defended by the 'capability argument'. Lastly, the 'pragmatic argument' maintains that "it is efficient to ascribe virtues to collectives", since attributing virtues and fostering a corporate culture that encourages virtuous conduct can influence the attitudes and actions of individual members (Sandin, 2007). Sandin's (2007; 2012) argument, then, focuses on assessing the reasonableness and advantages of ascribing collective virtues in professional contexts such as agriculture and the military, as well as determining which virtues are relevant in these fields. Since these considerations fall outside the scope of this discussion, we prefer to follow Astola (2022) or Astola et al. (2022) in studying the possibility that a specific virtue, such as creativity and responsibility-as-virtue, can be attributed to a group. In our case, the

virtue in question is a 'global' moral reflexivity, turned toward the start-up's final product and overall production process and their effects on society and the environment. We ask whether a group can possess this regardless of the existence of a similar virtue in any member of the collective.

Expressed in these terms, the problem differs distinctly from Fricker's (2005) argument. She argues that a collective can be virtuous by possessing a small number of virtuous persons who lead a majority of "passengers" who allow themselves to be led, or even if the totality of the group members are of the latter type, provided the collective has some "explicit, comprehensive and effective procedures in place" that support virtuous intentions. As in this second case, in our hypothesis we posit that none of the members of the collective (at least *qua* members) has the kind of (high-level) moral reflectivity required of the collective but instead possesses a more restricted reflectivity directed at specific aspects that may vary according to the member's role and sensitivity. Therefore, *ex hypothesi*, none of the team members have a moral reflectivity that addresses the general overview of the production of a certain technological innovation to be introduced into the environment and society and its possible consequences.

### 3.1. The 'Swarm intelligence' model

We wonder, therefore, what organisational form, if any, would support the "emergence" of this virtue at a "macro level" from "micro-level components and processes" out of which emergent phenomena arise (Goldstein, 1999, p. 49). An analogy can be derived from the theories on the so-called 'Swarm Intelligence', a concept drawn from the behaviour of decentralised and self-organised systems, such as those detectable in social insects. In the case of the *Apis mellifera*, for example, a colony can show highly complex behaviour without individual insects having "any representation, any map or explicit knowledge of the global structure they produce. A single insect is not able to assess a global situation, to centralize information about the state of its entire colony" (Garnier et al., 2007, p. 7). According to Seeley (2002), an individual honeybee's cognitive and communicative abilities may be high yet insufficient to effectively oversee the functioning of the entire colony or account for the full range of behaviours observed at the colony level. In most situations, a single honeybee cannot independently devise an efficient solution to a colony-wide problem, whereas the colony can easily find a solution. As is evident, in our analysis, the ability to find 'smart' solutions as a group will be replaced by the ability to reflect on the final goods pursued by the research and development process as a whole. To stress this again, these supposed limitations in the members do not stem from their deficiencies as individuals outside the group, but from the lack of time and resources they may experience within a start-up team in the fast-paced field of AI-driven innovation - which, on the other hand, adds complexity to the evaluation of the possible outcomes of one's action.

As noted by Consiglio (2019), this model allows for the emergence of collective behaviour that does not require a "collective intentionality" (CI) in the terms defined by Searle (2006), according to which each agent in CI should feel part of a joint project. At the same time, however, the members' (limited) reflective disposition may not repair our argument from the criticism advanced by Fricker (2005) towards those hypothetical virtuous collectives guided by a so-called 'invisible hand', i.e. in which the vices of the members compensate each other giving a 'virtuous' outcome. In her view, this positive outcome is insufficient to consider the collective virtuous, given the absence of good motives or skills attributable to an agent. On this point, we appeal to Baddorf and McKay's (2024) argument and their 'imitationist' account of collective virtue, according to which collectives can be deemed virtuous provided that they "possess reasonably broad and stable dispositions to functionally and intentionally imitate individual virtues and vices". That means these collective virtues should "involve intentional states with the same (or very similar) intentional content" of individual ones, which could be the case of the reflexive collective we are describing. Indeed, the intentional content of a reflexive collective is, in essence, the disposition to self-examine the values manifested by its actions and endorsed through its desired outcomes. This disposition is functionally akin to the intentional content of a reflective individual.

Following this formulation, the two authors agree with Fricker that if the compensation of vices occurs by pure chance, the collective is not virtuous (there being no 'stable' disposition), whereas this is different if it happens by 'design', as the authors suggest being the case in the formation of juries in the United States. In §5, we will consider the model of 'swarm intelligence' as a possible way to design this type of 'stable' interaction in a collective, by modelling a flowchart for ethical decision-making in a team. First, however, we will mention an occasion where collective reflexivity has already been discussed in the literature concerning collective virtues. Although, in this case, such reflexivity is seen as a basis for virtue and not as virtue itself, this will help us to better define the issue we are dealing with.

### 3.2. Collective reflexivity

After presenting their theory on collective virtues, Baddorf and McKay (2024) confront an objection put forward by Cordell (2017). The latter highlights how virtues require "advanced capacities for reflection (e.g., on the value of the virtue in question)" (Baddorf & McKay, 2024, p. 417) and how, therefore, collectives lacking this kind of reflexive attitude would rather have desirable features or qualities, and not fully-fledged virtues. The authors discuss this possible objection by suggesting that collectives may indeed be capable of this reflection, or at least "of capacities that are functionally and intentionally similar enough to their individual counterparts to suffice for genuinely agential virtue" (p. 419). They do not pursue this hypothesis to its end, limiting themselves to noting what they call a "promising avenue", which consists of the possibility that "some individuals [the few virtuous 'leaders' in Fricker's proposal] could reflect on the value of a collective virtue in such a way that their reflection instantiated collective reflection" (p. 419). This way, collectives could "functionally and intentionally imitate" the reflection on virtue through the reflective capacities that we can attribute to certain members of the collective. Nevertheless, this proposal does not entirely address the issue we are analysing in this article, since this approach presupposes, much like Fricker's proposal, that at least some members of the team possess a high level of reflection.

Nevertheless, their proposal opens to different directions of enquiry, among which we choose two, which will support our argument. The first question that might emerge from engaging with the argument discussed by Baddorf and McKay concerns whether the individual reflection should consist of a moment of detachment in which one has an overall view of the process at hand or whether opportunities for moral reasoning embedded in the reality of everyday alternatives suffice. On the other hand, we can also ask in which way the individual reflection can give rise to the collective one, supposing that Fricker's model of virtuous leaders and "passengers" is not the only viable. To answer the first question, we turn to the virtues and moral reflection as understood by Alasdair MacIntyre. As we shall see, in fact, such a perspective makes it possible to conceive of a type of moral reflexivity that is grounded in specific obligations rather than being directly oriented towards a broader perspective. Then, we will address the other half in §5, where we will look closer at the possibility of applying swarm theory in start-up management to sustain this collective virtue.

## 4. MacIntyre's moral reflectivity

According to MacIntyre, individuals can only answer the question "What am I to do?", from which all practical reasoning stems (MacIntyre, 1988, p. 24), if they are also able to answer the prior question "Of what story or stories do I find myself a part?" (MacIntyre, 2007, p. 216). These questions are crucial for a reflective individual,

since being a "reflective moral agent" (MacIntyre, 2016, p. 74) means questioning whether one has sufficiently good reasons supporting their desires (2016, p. 8), and this implies understanding what the "goods that furnish us with reasons for desiring as we do" may be (2016, p. 9). Identifying these goods involves rendering one's desires intelligible to others by situating them within broader narratives that are shared and understood collectively. These goods can correspond to the standards of excellence corresponding to "particular roles or functions within the context of particular kinds of practice", e.g. in our professional life, as well as to external goods of these practices, such as personal wealth or overall well-being. However, a "good human being" should not limit herself to acting within the confines of these roles but should also act and judge "*qua* human being" (MacIntyre, 1999a, p. 65).

For this reason, individuals must be able to "reflect upon their role-playing in ways that are not dictated by those same roles". This does not require adopting an external, absolute standpoint to evaluate these standards of excellence and goods. Instead, this "critical practical enquiry" takes shape through the capacity to assess a given social practice by considering the goods that are central to another practice to which one also belongs. MacIntyre, therefore, highlights the importance of two virtues: integrity and constancy. A person of integrity cannot adopt entirely different personas across social contexts, having set "inflexible limits to one's adaptability" within the roles they assume. Similarly, constancy constrains the "flexibility" of character, enabling individuals to remain committed to the same goods *over time*, without being pulled in different directions by shifting social circumstances. As we shall later examine, modern individuals often lack this kind of inflexibility, as their identity and moral obligations are no longer rooted in the different social practices they engage in Instead, "individuals with these two virtues will learn not only how to occupy some determinate set of roles within their social order, but also how to think of their goods and of their character independently of the requirements of those roles" (MacIntyre, 1999b). This perspective requires us to view our lives as comprehensive systems in which the various goods and obligations we abide by are interconnected. Accordingly, when one pursues a particular good, this should also contribute to attaining all other goods. In other words, "that final good must stand to all other goods in this way: that, if and insofar as other goods are given a *due* place in an agent's life, that agent is directed toward the achievement of this, her or his final good, *and* vice versa" (2016, p. 53, my italics). The emphases within this quotation underscore the circular nature of the argument, where the pursuit of the final good necessitates ordering particular goods correctly, and the 'due' order is defined by its capacity to facilitate the pursuit of the final good. We may thus interpret this final good as the holistic system resulting from the proper integration of various goods, such that "in acting for the sake of achieving some particular good, we also act for the sake of achieving our final end" (2016, p. 229).

MacIntyre's view, which links human flourishing to the coherent integration of multiple goals within a broader framework, differs from Aristotle's and remains open to the possibility of "tragic conflict" in the life of a virtuous person. Even with well-developed practical wisdom, situations may arise where the pursuit of one good cannot be reconciled with the attainment of another (2007, p. 201). On the other hand, this allows moral reflection to emerge from the tension between two shared narratives, and the corresponding goods, that shape an individual's actions. Using a term coined by Korsgaard (1996) and adopted by Fricker (2005), this consists of a friction between two or more "practical identities", i.e. the possibilities of acting (and deciding what to do) *qua* members of two different groups, for whom two different actions would be the most appropriate. This kind of reflection on what goods to pursue is not brought about by "theoretical reflection, but in everyday shared activities and the evaluation of alternatives that those activities impose" (MacIntyre, 1999a, p. 136). This also means that such moral reflection occurs in weighing equally valid goods that arise in daily practice rather than in the contemplation of goods of a higher order. At the same time, this reflection does not necessitate understanding our 'real' self, which constitutes our identity as human and rational beings (see Korsgaard (1996)). It is sufficient, instead, to compare our practical identities with one another.

## 5. Reflexivity as a collective virtue in start-ups

Through MacIntyre's texts, we have examined how moral reflexivity can arise without implying a global view of the meaning and context of one's actions. Based on these considerations, we ask what kind of collective, or dynamics within a collective, can guarantee it the virtue of a moral reflexivity (or "wisdom", adopting the term used by Vallor, 2016) that is commensurate with the complexity of the issues raised by the introduction of a technological innovation into our society and ecosystem. Whatever the ideal characteristics of this collective, however, "it must be emphasised that those deliberations require structural embedding, but cannot be replaced by it, i.e., what is required is a process that enables and fosters ethical thinking" (Zuber et al., 2022, p. 6). Having said that, a collective in which a high-level quality of the group can emerge from low-level qualities of the members is the one described by the theories about Swarm Intelligence (as already discussed in §3), of which we have not, however, defined either criteria or conditions for it to be suitable for the work context under consideration.

When we examine this particular 'sphere of action', i.e. the workplace, it becomes evident that authors who draw from MacIntyre's writings in their analysis of business ethics often neglect this specific aspect of his thought. Instead, they tend to focus on the distinction between internal and external goods, highlighting the significance of the former in promoting virtuous behaviour and meaningful work (see Beadle & Knight, 2012; Moore, 2005, 2022). Others focus on the role of management and institutions, which they view as inherently in conflict with the pursuit of the internal goods within a given practice (see, for instance, Brewer, 1997; Burns, 2022; Moore, 2002, 2008; Sinnicks, 2018). Since these issues are not central to the objectives of this research, we will refrain from exploring them in further detail. We will focus, instead, on what emerges from MacIntyre's writings as a significant obstacle to a reflective attitude of the individual, namely what he calls the "compartmentalised lives" characteristic of modernity, in which people move abruptly from one social context to another (MacIntyre, 2016, p. 202). In these terms, the problem lies in the absence of contact (and, thereby, of conflict) between the different 'social roles' we play, on whose intersection moral reflection is nourished. Under these circumstances, "issues which open up conflicts […] had hitherto been partly or wholly concealed from view" (2016, p. 203). In a working context, this type of compartmentalisation can take place on two levels: the individual worker, who does not connect her tasks with responsibilities that may affect other aspects of their life, and the level of communication between colleagues. Compartmentalisation on the first level prevents the moral conflict between contradictory responsibilities that triggers ethical judgement of one's actions. When such partitioning occurs on the second level, it affects the sharing of concerns - when they arise - and the activation of similar reflections in the other members of the collective. To be sure, this second aspect may also concern start-ups since a frequent dialogue between different technical competencies does not imply an exact counterpart on the level of moral issues.

Translated positively, this de-compartmentalisation should give rise to "shared deliberate reasoning", for which MacIntyre defines three conditions to be found in community members (MacIntyre, 1999a, p. 136). These conditions complement a wider set of intellectual and moral virtues that underpin shared moral reasoning (1999a, p. 97). They comprise three key aspects: 1) possessing adequate self-knowledge; 2) the ability to critically detach oneself from personal desires; and 3) recognising one's dependence on others. The first two correspond to the initial level of compartmentalisation discussed earlier, while the third aligns with the second level. When examined in relation to the "three types of preconditions" necessary for moral agency within social

structures (MacIntyre, 1999b), it becomes clear that sustaining these capacities requires an appropriate social environment. Such an environment can "sustain the relevant kind of understanding of the self, the relevant kind of critical discourse and reflection, and the relevant kind of accountability". Given the considerable overlap of the preconditions with the previously mentioned conditions, the analysis of the former will be incorporated into the articulation of the latter.

When examining the first condition, self-knowledge, it emerges through both "self-examination" and "accountability to those particular others who have reason to look to us to help in meeting their needs" (1999a, p. 95). This accountability entails recognising one's social roles, which in turn shapes an understanding of one's moral responsibilities. Moral conflicts may arise when the demands of these roles are weighed against other perceived obligations, stimulating critical reflection. As noted in the previous section, the issue is not solely about identifying the best course of action in pursuit of a specific goal. Rather, once a course of action is recognised as desirable or necessary, the question remains whether it can be reconciled with other moral obligations the individual considers essential. Self-knowledge, therefore, requires an awareness of the good one seeks through one's actions, as well as an understanding of how the pursuit of that good informs and justifies certain desires. Although this process may seem intrinsically subjective, the social environment plays a fundamental role in shaping self-knowledge. Indeed, being held accountable for one's actions nurtures the habit of examining oneself through the perspectives of others, allowing individuals to understand themselves as they are understood by those around them (1999a, p. 148).

Turning to the second condition, i.e., the ability to detach themselves from their own desires, this should be understood as an individual's capacity to critically distance herself not only from the actions typically regarded as the most efficient within her position but also from the expectations attached to her "role responsibility". Individuals are required to "understand their moral identity as to some degree distinct from and independent of their social role" (1999b). By adopting a perspective that, though not absolute, enables them to recognise and challenge the biases shaping their role, they would gain the ability to re-examine familiar issues from a different standpoint (2016, p. 112). Another aspect of this capacity lies in the ability to conceive of alternative possibilities, both in terms of courses of action and potential futures, that might emerge in response to a given situation (1999a, p. 76). Notably, the feasibility of these alternatives depends on a combination of self-knowledge and engagement with others (1999a, p. 95).

Coherently with what just said, the third condition emphasises that the pursuit of individual goods is always situated within the framework of a community, which collectively determines the common goods. MacIntyre articulates this idea, stating: "We should […] make it our aim that, so far as possible, each individual contributes to the shared deliberations of the community in her or his voice" (1999a, p. 140). The success of "shared rational deliberation" is central to the development of rational agents (2016, p. 57), and this can only be achieved when all relevant individuals recognise their accountability to others, extending beyond the boundaries of their "role performances" (1999b). This also means that sharing ethical concerns has the potential to dismantle "knowledge silos" (Zuber et al., 2022), revealing whether these concerns are magnified or diminished through interaction with other elements of the product development process. Furthermore, it encourages collaboration with technical specialists from other fields to explore solutions, thereby expanding the scope of moral reflection (and the search for a practical solution) across the team's various areas of expertise.

As we see in *Diagram 1*, after member A has shared their concerns, every other member of the community (if like-minded) will then evaluate A's desirable action (required by A's role) from the perspective of their own social responsibilities (unrelated to technical efficiency and product rentability). Then, if they confirm A's moral concerns (even if their reasons may be different from A's), they could assess them from their role in the production system. There might be the case that, from this point of view, they find a technological solution to propose to the team. Otherwise, they should contribute to the spread of A's moral concerns in the collective, engaging other members in this reflection. Therefore, we propose that a start-up that encourages these three features in its members and in their interactions may result in collective moral reflexivity that "functionally and intentionally imitates" (in the words of Baddorf & McKay, 2024) an individual reflexive virtue. To summarise, this collective virtue would emerge in the network of individuals who process their possible, efficiency-driven actions from the perspective of their other social roles. If they find any moral conflict, they point it out to others, who will examine the issue from their own perspective to reinforce (or not) this signal - thus shaping a kind of 'swarm moral reflexivity'.

- Advantages and disadvantages of the Swarm Intelligence model

Returning to the Swarm Intelligence model, we find similar dynamics in beehive colonies, for instance, when they choose where to nest. On these occasions, the individual scouts, having explored a possible site, begin to recruit other honeybees through the bee dance. This dance indicates to others the direction and distance of the site, but also its 'value' as perceived by the individual honeybee, which will dance longer if it 'believes' it is a valuable site to relocate the colony to. However, as Seeley and Visscher (2004) reveal, "a dance follower does not blindly imitate the dancer that she has followed. Instead, she leaves the swarm cluster, examines the advertised site, and only if she too judges that it is a worthy site does she produce a dance for it". Having said this, it is evident how the judgement of the individual honeybee

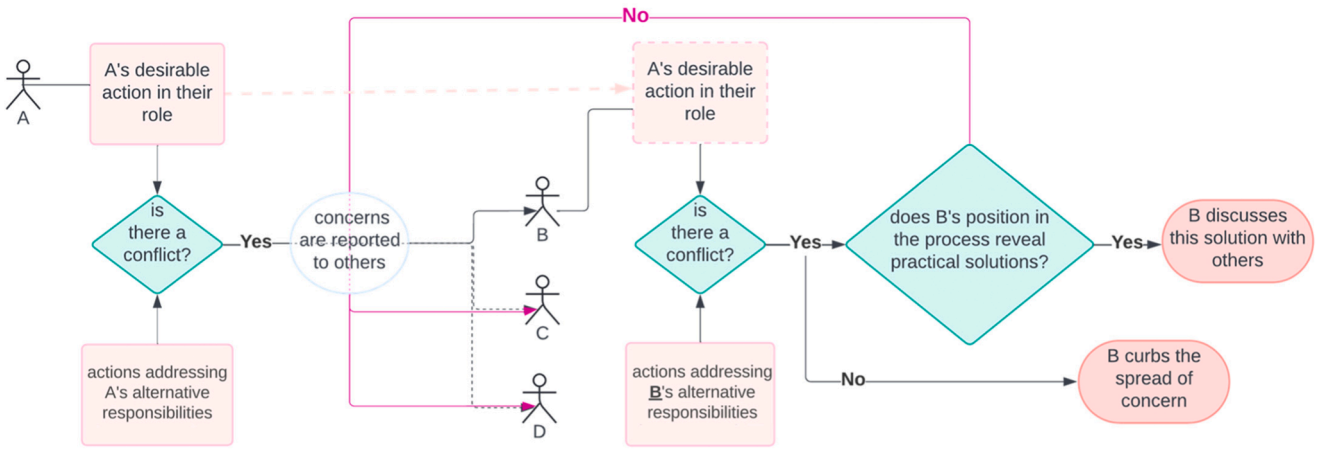


**Diagram 1.** The 'Swarm moral reflexivity' in action.

(and of the swarm as a whole) concerns the appropriateness of the site vis-à-vis a predetermined end, and does not question the validity of this end as the sole end. Adopting the classic Weberian distinction between "instrumentally rational" and "value-rational" behaviour, the bee and the swarm remain in the sphere of the former, while MacIntyre's ethics encourages (also) the latter attitude. We think, however, that we have made it sufficiently clear on which aspect the analogy with swarm behaviour rests, and that further exploration of the differences between human and honeybee cognition would be out of place in this article.

Besides the analogies, a 'swarm moral reflexivity' could also have the advantages listed by Harvey (2018, pp. 119, 120) about emergent behaviour in a 'swarm intelligence': simplicity (in the task required to the "individual agents"), robustness, flexibility, scalability, autonomy, and parallelism. The major limitation concerning the "time required to reach a solution" (p. 121), due to the consultation of the different members and their expertise, can be offset by the persistence of this process and its applicability in the background of the day-to-day concrete practices of product development. Moreover, these dynamics would not be only suitable for the "emergent networks" proposed by Goldstein (1999, pp. 66, 67) as a possible adaptation of the concept of 'emergence' in organisations. According to him, this would result from a self-organised, participative structure involving "intra- and intergroup dynamics". Instead, a 'swarm' moral reasoning could coexist with a hierarchical, imposed structure ensuring stability and goal-directedness to the processes internal to the organisation while allowing critical voices to arise that direct or put into question specific activities, producing a concrete result on the overall outcome.

That said, one additional criticism may become apparent. It may be that a negative outcome of the production process results from a series of functional steps that, although examined with detachment by self-knowing individuals, do not pass a 'critical threshold' of conflict between efficiency and the extra-work responsibilities they acknowledge. This is exemplified by Hess (2014, p. 248) when she speculates about an 'ACME' corporation in which all members are committed to environmental protection. In ACME's decentralised decision-making, however, after a sequence of "piecemeal modifications and others during implementation - each innocuous and rational enough within its limited sphere - the new production line results in a continuing discharge that pollutes a local river". In this example, the narrow scope of the individual components prevents their virtue from translating into a collective one - with a disastrous consequence. In the face of this possible criticism, we object that, in the case presented, the distribution of tasks does not cover the totality of aspects of the product and production chain that the collective, as a whole, should be able to control. In fact, we believe that a significant problem in the hypothetical ACME is that none of the members of the collective is in the position, in their daily tasks, to make a decision or (value) assessment regarding the corporation's discharges.

## 6. Conclusion

In this article, we have proposed three aspects that a start-up company can encourage internally so that the collective virtue of moral reflexivity can emerge in the interaction between its members. These aspects involve the worker's critical detachment from the most efficient solution to the assigned task, the maintenance of a viewpoint oriented towards personal social responsibilities (possibly encouraging their assumption through an appropriate corporate culture), and the readiness to communicate any moral conflicts between these two elements to other members. All this could make the collective virtue of moral reflexivity possible without requiring individual members to assume an overall moral outlook on their (future) technology and their R&D and production system. We have illustrated how this attitude can trigger a diffusion of moral evaluations of one's work in a dynamic being analogous to that in swarms (or in a colony of social insects). In fact, if we consider the internal decision-making process in honeybee colonies, already mentioned in §5, we see how "[t]here is no need for any individual to possess a global view of the alternatives, nor for any mechanism to tally and compare 'votes' for them. The 'smarts' of a swarm derives from a combination of many individuals working in parallel, each one making sophisticated assessments of nest-site properties, and a group process of feedback in recruitment" (Seeley & Visscher, 2004). Needless to say, however, just as this decision-making process is not enough for the displacement of the bee colony to a new nest, collective reflexivity in a team does not suffice for the ethical improvement of the emerging technology. Other stages of technical deliberation, such as the discussion of technological solutions eventually found by other members (see *Diagram 1*), are instead indispensable to the process of developing responsible innovation.

## Data availability

We do not analyse or generate any datasets, because our work proceeds within a theoretical approach. One can obtain the relevant materials from the references below.

## CRediT authorship contribution statement

**Marco Innocenti:** Writing – review & editing, Writing – original draft.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References


Astola, M. (2022). Collective responsibility should be treated as a virtue. *Royal Institute of Philosophy Supplements, 92*, 27–44. https://doi.org/10.1017/S1358246122000133

Astola, M., Bombaerts, G., Spahn, A., & Royakkers, L. (2022). Can creativity be a collective virtue? Insights for the ethics of innovation. *Journal of Business Ethics, 179* (3), 907–918. https://doi.org/10.1007/s10551-021-04833-0

Auernhammer, J. (2020). Human-centered AI: The role of Human-centered design research in the development of AI. In *DRS Biennial Conference Series*.

Baddorf, M., & McKay, N. (2024). A theory of collective virtue. *Journal of Ethics and Social Philosophy, 27*(3). https://doi.org/10.26556/jesp.v27i3.2005

Beadle, R., & Knight, K. (2012). Virtue and meaningful work. *Business Ethics Quarterly, 22* (2), 433–450.

Bezuidenhout, L., & Ratti, E. (2021). What does it mean to embed ethics in data science? An integrative approach based on microethics and virtues. *AI & Society, 36*(3), 939–953. https://doi.org/10.1007/s00146-020-01112-w

Bingley, W. J., Curtis, C., Lockey, S., Bialkowski, A., Gillespie, N., Haslam, S. A., et al. (2023). Where is the human in human-centered AI? Insights from developer priorities and user experiences. *Computers in Human Behavior, 141*, Article 107617. https://doi.org/10.1016/j.chb.2022.107617

Björnsson, G., & Hess, K. (2017). Corporate crocodile tears? On the reactive attitudes of Corporate agents. *Philosophy and Phenomenological Research, 94*(2), 273–298. https://doi.org/10.1111/phpr.12260

Boenink, M., & Kudina, O. (2020). Values in responsible research and innovation: From entities to practices. *Journal of Responsible Innovation, 7*(3), 450–470. https://doi.org/10.1080/23299460.2020.1806451

Brem, A., Giones, F., & Werle, M. (2023). The AI Digital revolution in innovation: A conceptual framework of artificial intelligence technologies for the management of innovation. *IEEE Transactions on Engineering Management, 70*(2), 770–776. https://doi.org/10.1109/TEM.2021.3109983

Brewer, K. B. (1997). Management as a practice: A response to Alasdair MACINTYRE. *Journal of Business Ethics, 16*(8), 825–833. https://doi.org/10.1023/A:1017997200200

Burnell, D., Neubert, E., & Fisher, G. (2023). Venture tales: Practical storytelling strategies underpinning entrepreneurial narratives. *Business Horizons, 66*(3), 325–346. https://doi.org/10.1016/j.bushor.2023.01.003

Burns, T. (2022). Alasdair MacIntyre on the division of goods and "the corrupting power of institutions. *Frontiers in Sociology, 7*, Article 986184. https://doi.org/10.3389/fsoc.2022.986184

Byerly, T. R., & Byerly, M. (2016). Collective virtue. *The Journal of Value Inquiry, 50*(1), 33–50. https://doi.org/10.1007/s10790-015-9484-y

Consiglio, F. (2019). Collective intelligence and social ontology. *Bridging The Divide Between Human And Animal Collective Cognition Through Stigmergy And Peircean Semiotics, 11*.

Cooper, R. G. (2024). The AI transformation of product innovation. *Industrial Marketing Management, 119*, 62–74. https://doi.org/10.1016/j.indmarman.2024.03.008

Cordell, S. (2017). Group virtues: No great leap forward with collectivism. *Res Publica, 23* (1), 43–59. https://doi.org/10.1007/s11158-015-9317-7

Deuten, J. J., & Rip, A. (2000). Narrative infrastructure in product creation processes. *Organization, 7*(1), 69–93. https://doi.org/10.1177/135050840071005

Fricker, M. (2005). Can there Be institutional virtues? *Oxford Studies in Epistemology, 3*, 235–252.

Füller, J., Hutter, K., Wahl, J., Bilgram, V., & Tekic, Z. (2022). How AI revolutionizes innovation management – Perceptions and implementation preferences of AI-based innovators. *Technological Forecasting and Social Change, 178*, Article 121598. https://doi.org/10.1016/j.techfore.2022.121598

Garnier, S., Gautrais, J., & Theraulaz, G. (2007). The biological principles of swarm intelligence. *Swarm Intelligence, 1*(1), 3–31. https://doi.org/10.1007/s11721-007-0004-y

Gilbert, M. (1989). *On social facts*. Princeton University Press. https://doi.org/10.2307/j.ctv10vm20z

Gogoll, J., Zuber, N., Kacianka, S., Greger, T., Pretschner, A., & Nida-Rümelin, J. (2021). Ethics in the software development process: From codes of conduct to ethical deliberation. *Philosophy & Technology, 34*(4), 1085–1108. https://doi.org/10.1007/s13347-021-00451-w

Goldstein, J. (1999). Emergence as a construct: History and issues. *Emergence, 1*(1), 49–72. https://doi.org/10.1207/s15327000em0101_4

Gowri, A. (2007). On corporate virtue. *Journal of Business Ethics, 70*(4), 391–400. https://doi.org/10.1007/s10551-006-9117-2

Griffith, E. (2024). Investors pour $27.1 billion into A.I. Start-ups, defying a downturn. *The New York Times*. https://www.nytimes.com/2024/07/03/technology/ai-start ups-funding.html.

Hagendorff, T. (2022). A virtue-based framework to support putting AI ethics into practice. *Philosophy & Technology, 35*(3), 55. https://doi.org/10.1007/s13347-022-00553-z

Harris, C. E. (2008). The good engineer: Giving virtue its due in engineering ethics. *Science and Engineering Ethics, 14*(2), 153–164. https://doi.org/10.1007/s11948-008-9068-3

Harvey, J. (2018). The blessing and curse of emergence in swarm intelligence systems. Eds.. In H. A. Abbass, J. Scholz, & D. J. Reid (Eds.), *Foundations of trusted autonomy* (pp. 117–124). Springer International Publishing. https://doi.org/10.1007/978-3-319-64816-3_6

Hayes, P., Fitzpatrick, N., & Ferrández, J. M. (2024). From applied ethics and ethical principles to virtue and narrative in AI practices. *AI and Ethics*. https://doi.org/10.1007/s43681-024-00472-z

Herrmann, T., & Pfeiffer, S. (2023). Keeping the organization in the loop: A socio-technical extension of human-centered artificial intelligence. *AI & Society, 38*(4), 1523–1542. https://doi.org/10.1007/s00146-022-01391-5

Hess, K. M. (2014). The free will of corporations (and other collectives). *Philosophical Studies, 168*(1), 241–260. https://doi.org/10.1007/s11098-013-0128-4

Horvatić, D., & Lipic, T. (2021). Human-Centric AI: The symbiosis of Human and artificial intelligence. *Entropy, 23*(3). https://doi.org/10.3390/e23030332. Article 3.

Korsgaard, C. M. (1996). *The Sources of Normativity*. New York: Cambridge.

Lubberink, R., Blok, V., Van Ophem, J., & Omta, O. (2017). Lessons for responsible innovation in the business context: A systematic literature review of responsible, social and sustainable innovation practices. *Sustainability, 9*(5). https://doi.org/10.3390/su9050721. Article 5.

MacIntyre, A. C. (1988). *Whose justice? Which rationality*? University of Notre Dame Press.

MacIntyre, A. C. (1999a). *Dependent rational animals*. Carus Publishing Company.

MacIntyre, A. C. (1999b). Social structures and their threats to moral agency. *Philosophy, 74*(289), 311–329.

MacIntyre, A. C. (2007). *After virtue: A study in moral theory*. University of Notre Dame Press, 3rd ed.

MacIntyre, A. C. (2016). *Ethics in the conflicts of modernity: An essay on desire, practical reasoning, and narrative*. Cambridge University Press. https://doi.org/10.1017/9781316816967

Martin, M. W. (2002). Personal meaning and ethics in engineering. *Science and Engineering Ethics, 8*(4), 545–560. https://doi.org/10.1007/s11948-002-0008-3

McLennan, S., Fiske, A., Celi, L. A., Müller, R., Harder, J., Ritt, K., et al. (2020). An embedded ethics approach for AI development. *Nature Machine Intelligence, 2*(9), 488–490. https://doi.org/10.1038/s42256-020-0214-1.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society, 3*(2), Article 2053951716679679. https://doi.org/10.1177/2053951716679679

Moore, G. (2008). *Re*-imagining the morality of management: A modern virtue ethics approach. *Business Ethics Quarterly, 18*(4). https://doi.org/10.5840/beq200818435

Moore, G. (2002). On the implications of the practice–Institution distinction: Macintyre and the application of modern virtue ethics to business. *Business Ethics Quarterly, 12* (1), 19–32. https://doi.org/10.2307/3857646

Moore, G. (2005). Humanizing business: A modern virtue ethics approach. *Business Ethics Quarterly, 15*. https://doi.org/10.2307/3857679

Moore, G. (2022). A macintyrean virtue ethics perspective on humanizing business. Eds.. In M. Dion, R. E. Freeman, & S. D. Dmytriyev (Eds.), *Humanizing business: 53. Humanizing business* (pp. 33–42). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-72204-3_3

Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics, 26*(4), 2141–2168. https://doi.org/10.1007/s11948-019-00165-5

Qian, Y., Siau, K., & Nah, F. (2024). Societal impacts of artificial intelligence: Ethical, legal, and governance issues. *Societal Impacts, 3*. https://doi.org/10.1016/j.socimp.2024.100040

Rahwan, I. (2018). Society-in-the-loop: Programming the algorithmic social contract. *Ethics and Information Technology, 20*(1), 5–14. https://doi.org/10.1007/s10676-017-9430-8

Ratti, E., & Graves, M. (2021). Cultivating moral attention: A virtue-oriented approach to responsible data science in healthcare. *Philosophy & Technology, 34*(4), 1819–1846. https://doi.org/10.1007/s13347-021-00490-3

Riedl, M. O. (2019). Human-centered artificial intelligence and machine learning. *Human Behavior and Emerging Technologies, 1*(1), 33–36. https://doi.org/10.1002/hbe2.117

Ryan, M., Popa, E. O., Blok, V., Declich, A., Berliri, M., Alfonsi, A., et al. (2023). A model of social responsibility for start-ups: Developing a cross-fertilisation of responsible innovation, the lean start-up approach, and the quadruple helix approach. *Journal of Responsible Innovation, 10*(1), Article 2264615. https://doi.org/10.1080/23299460.2023.2264615

Ryan, M., Popa, E. O., Blok, V., Declich, A., Berliri, M., Alfonsi, A., et al. (2024). Start doing the right thing: Indicators for socially responsible Start-ups and investors. *Journal of Responsible Technology, 20*. https://doi.org/10.1016/j.jrt.2024.100094

Sandin, P. (2012). A collective virtue approach to agricultural ethics. Eds.. In T. Potthast, & S. Meisch (Eds.), *Climate change and sustainable development: Ethical perspectives on land use and food production* (pp. 151–153). Wageningen: Academic Publishers. https://doi.org/10.3920/978-90-8686-753-0_20

Sandin, P. (2007). Collective military virtues. *Journal of Military Ethics*. https://doi.org/10.1080/15027570701755505

Schmidt, J. A. (2014). Changing the Paradigm for Engineering Ethics. *Science and Engineering Ethics, 20*(4), 985–1010. https://doi.org/10.1007/s11948-013-9491-y.

Searle, J. R. (2006). Social ontology: Some basic principles. *Anthropological Theory, 6*(1), 12–29. https://doi.org/10.1177/1463499606061731

Seeley, T. D. (2002). When is self-organization used in biological systems? *The Biological Bulletin, 202*(3), 314–318. https://doi.org/10.2307/1543484

Seeley, T. D., & Visscher, P. K. (2004). Group decision making in nest-site selection by honey bees. *Apidologie, 35*(2), 101–116. https://doi.org/10.1051/apido:2004004

Sinnicks, M. (2018). Leadership after virtue: MacIntyre's critique of management reconsidered. *Journal of Business Ethics, 147*(4), 735–746. https://doi.org/10.1007/s10551-016-3381-6

Steen, M. (2021). Slow innovation: The need for reflexivity in responsible innovation (RI). *Journal of Responsible Innovation, 8*(2), 254–260. https://doi.org/10.1080/23299460.2021.1904346

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. *Research Policy, 42*(9), 1568–1580. https://doi.org/10.1016/j.respol.2013.05.008

Tekic, Z., & Füller, J. (2023). Managing innovation in the era of AI. *Technology in Society, 73*, Article 102254. https://doi.org/10.1016/j.techsoc.2023.102254

Umbrello, S., & Gambelin, O. (2023). Agile as a vehicle for values: A value sensitive design toolkit. Eds.. In A. Fritzsche, & A. Santa-María (Eds.), *Rethinking technology and engineering: 45. Rethinking technology and engineering* (pp. 169–181). Springer International Publishing. https://doi.org/10.1007/978-3-031-25233-4_13

Umbrello, S., & van de Poel, I. (2021). Mapping value sensitive design onto AI for social good principles. *AI and Ethics, 1*, 3. https://doi.org/10.1007/s43681-021-00038-3

Ustek-Spilda, F., Powell, A., & Nemorin, S. (2019). Engaging with ethics in Internet of Things: Imaginaries in the social milieu of technology developers. *Big Data & Society, 6*. https://doi.org/10.1177/2053951719879468, 205395171987946.

Vallor, S. (2016). *Technology and the virtues*. Oxford University Press.

Winkler, T. J. (2020). Value sensitive design and agile development: Potential methods for value prioritization. *Societal Challenges in the Smart Society*.

Xu, Y., Bradford, N., & Garg, R. (2023). Transparency enhances positive perceptions of social artificial intelligence. *Human Behavior and Emerging Technologies, 2023*(1), Article 5550418. https://doi.org/10.1155/2023/5550418

Zuber, N., Gogoll, J., Kacianka, S., Nida-Rümelin, J., Pretschner, A., et al. (2024). Value-sensitive software design: Ethical deliberation in agile development processes. Eds.. In H. Werthner, C. Ghezzi, J. Kramer, J. Nida-Rümelin, B. Nuseibeh, E. Prem, et al. (Eds.), *Introduction to digital humanism: A textbook* (pp. 339–358). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-45304-5_22

Zuber, N., Gogoll, J., Kacianka, S., Pretschner, A., & Nida-Rümelin, J. (2022). Empowered and embedded: Ethics and agile processes. *Humanities and Social Sciences Communications, 9*(1), 1–13. https://doi.org/10.1057/s41599-022-01206-4