Roberto Redaelli

Composite Intentionality and Responsibility for an Ethics of Artificial Intelligence

> *Morality is neither to be found in the objects themselves, nor in autonomous subjects. It only comes in relations between subjects and objects, where objects have moral significance and subjects are engaged in mediated relations with the world*
>
> (Verbeek 2014, p. 87).

1. Some preliminary remarks

In recent years there has been an exponential growth in the use of machines equipped with artificial intelligence: from the progressive automation of vehicles to the increasingly widespread use of robotics in the medical-healthcare field, to the voice assistants with which our personal devices are designed, these technologies are bringing radical change to our lifestyles. While there is no doubt that these innovations are beneficial to our lives, the pervasiveness of these technologies must not be underestimated and raises increasingly urgent ethical and legal issues. For these reasons, the spread of intelligent systems into the most disparate areas of our lives renders necessary, alongside a re-modelling of our habits, a radical process of re-semantisation of some notions traditionally ascribed to human beings, such as those of intentionality and responsibility (see Doorn & van de Poel 2012). This paper addresses the process of redefining these notions in order to examine the increasingly close relationship between humans and AI. To this end, we propose to apply in the specific field of artificial intelligence the idea of composite intentionality developed by P.P. Verbeek, with the aim to clarify hybrid forms of a) *intentionality* and b) *responsibility*, and take positive steps towards an *extended agency theory* (see, for example, Hanson 2009; Gunkel 2020).

Scenari, vol. 17, 2/2022 • ISSN 2420-8914 • DOI 10.7413/24208914133

2. Verbeek and the moralisation of technologies

A large part of Verbeek's efforts (Kroes & Verbeek 2014; Verbeek 2011; 2005a) is aimed at accounting for the mediating role played by technologies in our habits and then ascribing to them a clear moral significance. Distancing himself from both the reflections on technology of Heidegger in his later period and Jaspers (see Verbeek 2005a), as well as from any purely instrumental or alienating view of technology, Verbeek elaborates, in an original way, upon many of the fruitful insights found in Latour (see, e.g., Latour 1993; 1994; 2002) and Ihde (see, e.g., 1979; 1990; 1993) in order to highlight the mediation function that technologies perform in our everyday lives.

In fact, Verbeek credits Latour with not reducing technologies to the product of networks of social interactions, as is the case with the social constructivist conception of technology, and rather with highlighting the ways in which "technologies themselves coshape the interactions" (Verbeek 2005a, p. 103), that is, their active role. Despite this merit, Latour's approach, according to Verbeek, fails to adequately bring to the foreground the ways in which technology "coshapes the access human beings have to reality" (Verbeek 2005a, p. 104). In other words, Latour, from an outside perspective[1], reduces the connections between entities to the mere terms of associations (Verbeek 2005a, p. 165), whereas Ihde's postphenomenological approach offers a more nuanced look at these connections, analysing them in terms of experience and behaviour, based on an internal point of view. However, this difference in viewpoints does not lead, according to Verbeek, to a conflict between the two perspectives; rather, they are compatible: both aim, though in different ways, to overcome the subject-object dichotomy. This conviction is the starting point for Verbeek's attempt at "forging from these two approaches a fruitful way to analyze technological mediation" (Verbeek 2005a, p. 168) through a translation of Latour's vocabulary in a postphenomenological perspective.

More precisely, by virtue of this admixture of the two perspectives, ethics has the task of extending its field of inquiry beyond the human sphere, overcoming the modernist subject-object dichotomy in the direction of an "amodern" worldview, inaugurated first by Latour. In fact, Verbeek,

[1] Verbeek defines the perspective taken by Latour as a perspective 'from the outside' in these terms: "Latour argues not from the standpoint of human beings who are concretely situated in the world, but from the standpoint of an analyst who describes configurations equally from the perspective of humans and non-humans. […] What postphenomenology contributes to actor-network theory is the situated perspective, the perspective 'from inside out'" (Verbeek 2005a, p. 168).

who credits Latour more generally with focusing on "nonhuman forms of agency" (Verbeek 2011, p. 17), proposes an ethical reflection that reconsiders the original fusion of humans and technology, as a consequence of which humans themselves become "technological beings" (Verbeek 2011, p. 4) or cyborgs (Verbeek 2008; see also Stiegler 1998). To this end, he uses – as just mentioned – what is commonly called a post-phenomenological approach, which was inaugurated and developed by Don Ihde (see, e.g., 1993)[2]. This approach, which has no foundational claim (see Verbeek 2011, p. 15), is presented in Verbeek's words as a "philosophical analysis of the structure of the relations between human beings and their lifeworld" (Verbeek 2011, p. 7). Within this type of analysis, which takes the form of experimental phenomenology (see Ihde 1977), emerges the idea of a philosophy of mediation. For Verbeek, this philosophy must take into account the by-no-means-ethically-neutral role played by technologies in the relationship between humans and the world. In fact, technological devices help us shape our experiences, influence our moral decisions, and have important repercussions on what we do and think: they co-shape practices of living and knowing. In this sense, we humans have a mediated, so to speak, hybrid experience of reality, as our action is so often mediated by technological devices. Therefore, although artifacts cannot be defined as *human-like* moral agents[3], they help humankind to make decisions, raise ethical dilemmas and offer tools for resolving them: they are bearers of moral demands, and, according to the philosopher, they possess a clear moral significance.

In order to explain this significance, Verbeek's proposal hinges, among other things, on an original notion of intentionality, which allows us to rethink the relationship between humans and reality, and, as we will try to show, to better understand the notion of responsibility in our digital era.

3. The concept of technological intentionality

As a starting point for the development of a non-humanist approach to ethics, Verbeek addresses the problem of the intentionality of artifacts. From both deontological and consequentialist perspectives, technological objects are excluded from ethics because they lack intentionality and freedom. Nevertheless, the moral role of these technologies in our society has become increasingly evident in recent years. As

[2] An accurate introduction to post-phenomenology can be found in Rosenberger & Verbeek 2015.

[3] On the moral agency of technological artifacts see, for instance, the different positions of Floridi & Sanders 2004; Johnson 2006; Sullins 2006; Wallach & Allen 2009.

proof of this role, numerous studies have demonstrated how intelligent systems, in addition to being able to perform the function of artificial moral advisor (Giubilini & Savulescu 2018), possess the ability to embody values and biases (Flanagan, Howe, & Nissenbaum 2008; van de Poel 2020). In fact, these devices demonstrate a tendency to convey values (and disvalues) reflecting opinions, idiosyncrasies, and evaluations of the various social actors involved in the development and use of AI systems. In order to avoid such transmission to the algorithm of discrimination and prejudices, as well as to discourage bad practices related to AI (see Floridi 2022), it was decided, on the one hand, to adopt ethical codes with which these technologies must be aligned during the programming processes (see Gabriel 2020), while on the other hand there was an attempt to disseminate among users some rules of conduct aimed at respecting the other parties in digital relationships.

However, beyond these solutions, which do not constitute a definitive answer to the axiological question, the paradox remains that artificial intelligence is a bearer of moral values, even though it is not recognized as having any intentionality. For the deontological approach, in fact, the morality of an action depends on the fact that an agent *intends to act* according to certain rational criteria and is *free* to do so. The artifacts seem to be devoid of this intention to act and this freedom. From the consequentialist perspective, however, what matters is the result of the action, the value of the outcomes. But although technology can achieve morally significant results, the moral responsibility for these results lies, according to this perspective, only with the human being who uses technology as an instrument for the realization of goals. In this way, both approaches exclude artifacts from the realm of ethics, assigning them a causal responsibility and a merely instrumental character[4].

In contrast to these two ethical approaches, Verbeek presents a third that takes into account the active role of technological mediation, whereby "things can be seen as part of the moral community in the sense that they help to shape morality" (Verbeek 2011, p. 42). According to this view, it is not legitimate to exclude technological objects from the ethical world, but rather it is appropriate to extend the notion of agency to include technology. To this end, Verbeek, first of all, observes – as Winner's (1986) pioneering work indicated – that artifacts embody human intentions *in a material way* and at the same time present emergent properties, which seem to assign them a certain degree of freedom. In this sense, one could observe that moral agency is distributed between humans and

[4] In Verbeek's eyes, even Ihde's reflection that recognizes several relationships human beings can have with technological artifacts (embodiment, hermeneutic, alterity, background relation) seems to lack a careful analysis of some forms of intentionality (see Verbeek 2008).

nonhumans; therefore, moral actions and decisions are the product of human-technological associations.

Given this premise, Verbeek reworks in human-nonhuman relational terms both the concept of intentionality and that of freedom, traditionally ascribed to the human agent. In the field of intentionality, two concepts must first be distinguished: the ability to form intentions typical of ethical theory and the phenomenological concept of the "directedness of human beings toward reality" (Verbeek 2011, p. 55). For Verbeek, the first type of intentionality is based on the second, because the ability to form intentions to act cannot exist without "being directed at reality and interpreting it in order to act in it" (Verbeek 2011, p. 55). Starting from Ihde's reflections, but going beyond them, the author then emphasizes how intentionality in the post-phenomenological sense is mediated by technological devices, or in his terminology, is *composite*, whereby "when this 'directedness' of technological devices is added to human intentionality, a *composite intentionality* comes about: a form of intentionality that results from adding technological intentionality to human intentionality" (Verbeek 2011, p. 145, see also Verbeek 2008).

From eyeglasses to thermometers, from air conditioning to artificial intelligence systems, technologies shape in ever new ways our experience of the world, and grant access to it. And it is on the basis of the idea of composite intentionality, to which Verbeek assigns both a representative and a constructive function (Verbeek 2011, p. 145), that it is possible to attribute to technologies a certain intentionality, or rather, recognize that the same intentionality is distributed between human and nonhuman subjects. In fact, technologies can change our behaviour, our perception of the world, and therefore have intentionality in the sense of the *directing role of human action* (Verbeek 2011, p. 57). In this sense, their intentionality is reducible neither to that of the designers nor to that of the users, but rather takes on a character that is, so to speak, emergent with respect to the first two; a character to which, according to Verbeek, a certain notion of freedom is linked. The effects of technologies are not completely predictable, or controllable by humans. They are characterized by "emergent forms of mediation" (Verbeek 2011, p. 127) and "feedback effects" that affect both us and the world we live in (see Di Martino 2017; Tenner 1996).

These traits of composite intentionality now allow us to understand how moral decisions are not merely a human product, but that the technological apparatuses with which humans access the world are already involved therein. Therefore, it does not appear unjustified to state that "moral decision making is a joint effort of human beings and technological artifacts" (Verbeek 2011, p. 58). Humans never decide or act in a vacuum, but rather do so in a technologically shaped world, within

which their choices are *directed* by technology, which opens up new possibilities for action, redefining the scope of reality and forming our habits.

In regard to the directive nature of technologies (Verbeek 2011, p. 57), linked to the notion of intentionality, the case study offered by the use of certain medical devices such as those used in prenatal diagnosis is particularly interesting. A useful example is made of obstetric ultrasound technology, to which Verbeek (2008a; 2011) cannot attribute a merely instrumental role in making the fetus visible inside the uterine cavity. The use of ultrasound tech, which makes it possible to predict many congenital defects before birth, involves a significant redefinition both of the ontological status of the unborn child, who becomes a possible patient, and of the parents, who can make extreme decisions on the basis of this diagnosis. In this case, the human-machine association conveys composite intentionality and the role played by technology in offering humans new directions of action is evident, which raises far-reaching ethical dilemmas (see, e.g., Mitchell 2001 for the ethical consequences of the use of these technologies).

Starting from this notion of composite intentionality, exemplified by the case study mentioned above, the application of Verbeek's post-phenomenological perspective, according to our working hypothesis, allows us to better understand the moral status of artificial intelligence. Indeed, compared to, so to speak, traditional technologies, artificial intelligence makes autonomous decisions, adapting its behaviour to the conditions in which it operates. In this capacity composite intentionality is revealed as a result, not merely a summation, of the human-machine association. Clarifying this type of intentionality and applying it to ethical problems[5] has become an increasingly urgent task, given the pervasiveness with which intelligent machines, capable of autonomy, adaptability and interaction (Floridi & Sanders 2004), are influencing our lives.

With respect to the position of Floridi and Sanders (2004) it can be observed that the notion of composite intentionality proposed by Verbeek has the merit of avoiding those positions defined by the authors as anthropocentric, whereby moral philosophy remains "unduly constrained by its anthropocentric conception of agenthood" (Floridi and Sanders 2004, p. 350), without, however, making use of the method of abstraction about whose tenability several objections have arisen (see Gunkel

[5] For instance, the problem of the 'contamination of the algorithm', that is, the transmission to the AI of the prejudices of programmers and users, can be clarified, if understood as a result of the addition of human intentionality to that of artificial intelligence. Therefore, a resolution of this problem can be elaborated from the recognition of this composite agency to which the notion of composite intentionality corresponds.

2017 p. 73; Johnson and Miller 2008). In fact, Verbeek's posthumanist position brings the human-machine relationship back within the notion of extended agenthood, attributing to it a composite intentionality, which can account for human-AI entanglement without reducing the latter's technological intentionality to that of humans. In this sense, the idea of composite intentionality developed by Verbeek seems to be particularly suitable for explaining the intentionality present in systems with artificial intelligence. Such systems, in fact, present a technological intentionality understood as a directedness (Verbeek 2008, p. 392) that guides our action and thinking, a form of intentionality that, as we have seen, is indeed connected to the man who designs the machine and uses it, but which, at the same time, presents an emergent property with respect to human intentionality.

Therefore, in such intelligent systems, the human-machine compositionality highlighted by Verbeek is manifested to the highest degree, whereby human intentionality is directed at the technological intentionality of the machine, which represents and constitutes the real not as a mere extension of the human, but by virtue of its own relevance (Verbeek 2011, p. 146). Technology, in this case, is developed in order "to reveal a reality that can be experienced only by technologies" (Verbeek 2011, p. 146).

4. Composite responsibility for rethinking the human-technology relationship

The question of the intentionality of artificial intelligence is linked to the second topic of our investigation, namely that of the moral and legal responsibility of these technologies. The use of artificial intelligence in the legal field as an aid to a judge's decision, as well as the use of technologies in the medical field for diagnostic purposes, raise the problem of the attribution of moral and legal responsibility for the indirect effects deriving from their use (for the legal field, see Funke 2022; for the moral field, Coleman 2004). In fact, if machines equipped with artificial intelligence make it possible in some areas to achieve results that are far more reliable than human ones, this does not free these technologies from error. These systems, used for example in the legal field to predict the risk of recidivism (for the Compas case, see Brennan et al. 2009), may favor discrimination, since their decisions are based on previous resolutions, according to a bottom-up statistical approach (for the limits of this approach, see Di Giulio 2020). This approach is based, in fact, on the ability of intelligent systems to learn, from the data available, how to acquire new knowledge and make decisions. Therefore, where the data set used

for learning is somewhat misleading (incomplete or unrepresentative), there may be unforeseen consequences produced by the algorithms. Because of this limitation, it is becoming increasingly urgent to identify an exact criterion for attributing responsibility for these consequences. To reach this goal, different proposals have been put forward, ranging from increasing human control over artificial intelligence (with the consequent attribution of responsibility to designers, producers and users) to the recognition of some moral or legal responsibility of the artificial agents themselves. In the latter case, it has even been proposed to assign to technologies equipped with artificial intelligence an "electronic personality", with its own legal subjectivity (in this regard, see Pacileo 2020).

Faced with these various proposed solutions, we aim to elaborate a post-phenomenological notion of responsibility, in line with that of intentionality, which reflects the composite character of human-artificial intelligence associations. This does not mean assigning a moral responsibility to artificial intelligence, but rather *recognizing a precise role in the process of forming the moral responsibility of the human agent* (Verbeek 2008, 2009; Hanson 2009; Gunkel 2020) both in the case of *backward-looking responsibility* and in the case of *forward-looking responsibility* (see de Poel 2011). In this way, this notion can contribute to the resolution of the so-called "responsibility gap" (Matthias 2004), whereby more complex and autonomous technologies involve less human intervention, so that it is less easy to assign responsibility to humans for the behavior of such machines.

In order to offer such a contribution, it is useful to demonstrate how the advent of intelligent systems makes the human-nonhuman association even more complex and inextricable. If Selinger and Engström (2007), following Ihde (2002), observed that the human subject modifies itself when using technological tools, the advent of technologies with operational autonomy entails a radical extension of agency with significant repercussions on the attribution of responsibility. Depending on how this redistribution is understood, two different approaches to the question can be recognized: some scholars identify a certain degree of moral responsibility in technological artifacts, assigning to them, in some cases, attributes similar to those of the human moral agent; others recognize intelligent systems as a "quasi-responsibility," that is, a conductive character of moral action (Verbeek 2011), for which the responsibility falls on the extended agent, understood as a human-machine association as a whole (Gunkel 2020; Hanson 2009).

This second hypothesis, which appears to be the most suitable to address the problem of responsibility at today's level of AI-enhanced technologies (see Gunkel 2020), can perhaps be improved by defining the notion of responsibility to better account for the complexity of the

human-*artificial intelligence* connection. Indeed, while it is sometimes possible to make an immediate distinction in the human-technology relationship between the causal responsibility of technology and the moral responsibility of the human agent, it is not so easy to make such a distinction in situations in which artificial intelligence is involved. Repurposing an example from Hanson (2009), we can observe that if no problem is created when we assign a mere causal responsibility to the rope used by rescuers to save a child who has fallen into a well, it is more difficult to establish responsibility when there are machines in action with operational autonomy, that is, capable of carrying out tasks, deciding and acting by changing their behavior in the face of unforeseen situations. In these cases, in order to properly allocate responsibility, it is necessary to take into account the various social actors involved (e.g. manufacturers, programmers, users and machines) and redistribute responsibility between them. However, achieving a fair redistribution is far from easy in cases where the artificial agent makes decisions in ethically relevant contexts (Wallach & Allen 2009), as in the case of unavoidable collisions of self-driving vehicles. In fact, the plurality of agents involved in these cases, from programmers to car manufacturers, as well as the regulators who allowed the circulation of such vehicles, raises the so-called "many hands" problem (see Nissenbaum 1994; van de Poel et al. 2015), so evidently the contribution of a multiplicity of agents to the action makes it difficult to clearly identify those responsible. To address this difficulty, it has been proposed in recent years to introduce *meaningful human control* (Santoni De Sio & Van de Hoven 2018) over the activities carried out by artificial intelligence, to ensure a human "controller" has sufficient information and time to intervene on the nonhuman agent. The implementation of such a control raises, however, technical and theoretical issues that are difficult to resolve, which once again involve human-machine collaboration, and interfere, among other things, with progressive automation of intelligent machines.

This complex situation must be taken into account in order to advance a notion of composite responsibility, which can be developed by implementing some valuable indications present in Hanson (2009), Gunkel (2020) and Verbeek (2011). This composite or hybrid responsibility must first and foremost account for the dense web of relationships woven by social actors, including individuals, organizations, natural entities and technology, so as to recognize the role played by the parties involved (Hanson 2009). In this analysis, it is necessary to pay particular attention to the function exercised by technological intentionality, understood as the ability to incline the user towards a certain purpose (see Benanti 2021).

Secondly, as observed by Gunkel (2020), an *extended agency theory* to

which a hybrid responsibility corresponds must have a communitarian perspective, which overcomes ethical individualism. In this direction, it is a question of grasping not only the composite character of responsibility, but also and above all the *shared* responsibility. According to our proposal, this quality of being shared requires an ethical commitment on the part of all those involved in the development and use of artificial intelligence, and thus there is an increasing need today to establish unambiguous ethical principles that guide the planning and use of these technologies.

Finally, in the development of this notion of composite responsibility it is necessary to consider not only the relationship between humans and artificial intelligence, but also the machine-to-machine relationship. Indeed, as Wiener (1988) predicted with great foresight, our technological world is no longer formed by the human-machine relationship alone, but also and above all by the interactions between machines. The notion of responsibility must therefore make sense of this last segment, too, in order to ensure the correct distribution of responsibilities. In fact, machine-to-machine interactions can on their own cause unwanted indirect effects with dramatic repercussions on users. For this reason, a notion of composite responsibility must take into account such a relationship in order to trace the human agents who have programmed, produced, and used the machines involved.

In order to avoid misunderstandings and incurring the criticism of being a-moral or even anti-moral, it is necessary here to clarify that with such a notion of composite responsibility, characterized by the elements listed above (the considerations of all actors, a communitarian perspective, a consideration of machine-machine relations), there is no intention to assign some moral or legal responsibility to the machine, but rather to highlight the role that technology plays in the process of shaping our moral responsibility. This role is particularly evident in intelligent systems, which can promote discrimination and prejudice by making use of data imbued with these disvalues, but this does not make such machines morally responsible for the results they produce. In this sense, Verbeek correctly observes that "technologies also contribute to the moral responsibility of human beings for the actions that come about in human-technological interaction. But […] this does not imply that technologies should be held morally accountable for their mediating roles in human behaviour – just as it does not make sense to consider technologies full-fledged moral agents in the way human beings are moral agents" (Verbeek 2011, p. 108). So, in more precise terms, with the notion of composite responsibility we intend here to account for the active role that technologies play in the sphere of moral responsibility, without thereby assigning some form

of responsibility to them, which could lead to the de-responsibilization of designers and users.

In this sense, the idea of composite responsibility does not relieve designers of responsibility for the effects produced by the technologies they develop, just as it does not relieve users of their responsibility for the effects resulting from the (inappropriate) use of such technologies. In fact, the former can anticipate, albeit within certain limits, the moral impact of the technologies they are developing, while the latter have the duty of appropriate use of those technologies according to the use plan[6]. It is precisely at the design and use stage that one can finally identify the space of human moral responsibility to both appropriately design and use the technologies. In this space of freedom, the moral mediation of technologies certainly plays a central role, sometimes revealing a discrepancy between the values intended and the values actually realized through the use of technologies[7]. This discrepancy does not free the human from his responsibilities, but rather assigns him new responsibilities[8]. To return to the example of intelligent systems that (unintentionally) promote discrimination and bias (see the Compas case), the choice of the data on which the algorithms are to be trained turns out to be crucial, and the human is solely responsible for the decision regarding the dataset to be used for training. If anything, the algorithm expands or shrinks the human space of freedoms by virtue of its ability or inability to incorporate certain values, and in this way contributes to the formation of human responsibility that is already always mediated by technology.

[6] "A use plan is a plan that describes how an artifact should be used to achieve certain goals or to fulfill its function. In other words, a use plan describes the proper use of a technical artifact, and that proper use will result (in the right context and with users with the right competences) in the artifact fulfilling its proper function" (Van de Poel 2020, p. 391).

[7] We use here the distinction proposed by Van de Poel between intended, embodied and realized values: "The intended values are the values intended by the system's designers. However, these intended values may be different from the embodied values when an artifact (or institution or system) has not been properly designed. The embodied value is the value that is both intended (by the designers) and realized if the artifact or system is properly used. The realized value, in turn, may be different from the embodied value: for example, because a technology is used differently than intended or foreseen" (Van de Poel 2020, p. 389)

[8] In regard to the antenatal diagnosis achieved thanks to the sonogram, for instance, Verbeek observes significantly that "the mere availability of testing possibilities had made us feel responsible for *not* testing and for accepting the 'risks' connected with that. The decision not to be put in the position of having to make a decision appeared to be a decision as well" (Verbeek 2011, p. VII). And again: "by making it possible to detect specific diseases, medical diagnostic devices do not simply produce images of the body but also generate complicated responsibilities, especially in the case of antenatal diagnostics and in situations of unbearable and endless suffering" (Verbeek 2011, p. 1).

5. Final remarks

In order to understand the scope of Verbeek's post-phenomenological perspective, it is appropriate, in closing our examination, to provide some clarification of terminology. The use of expressions that we have employed in the text, such as those regarding the morality of things and technological intentionality, has often raised a variety of suspicions regarding Verbeek's proposal. In particular, the conceptual framework elaborated by the philosopher has given rise to a number of criticisms regarding the idea that there is some kind of distribution of intentionality between humans and machines, to which the author himself has already responded in part (see Verbeek 2014). Here we will simply emphasize a point related to the notion of intentionality, which, for example, the articulate criticisms of Peterson and Spahn (2011, pp. 416ff.) do not seem to grasp. Clarification of this point is of vital importance, since Verbeek's entire philosophical project and, consequently, the reflections we have made so far hinge upon it.

With the expression 'technological intentionality' Verbeek aims to highlight the directive nature of technologies, but without attributing any intention to act to technologies, as Peterson and Spahn erroneously believe. Similarly, Verbeek's idea that technologies 'actively co-shape' our being cannot be understood as an action independent of humans, and thus – Peterson and Spahn erroneously observe – "technological objects certainly have an impact on us and our actions, like many other natural and nonnatural objects, but this impact is not active in the sense that it is independent of the designer or inventor who decides to produce or sell the new artifacts" (Peterson and Spahn 2011, p. 414).

Both of these observations of Peterson and Spahn miss the mark because they lose sight of the *co-constitutive* and therefore interdependent character of technologies, which invokes the notion of composite intentionality. Indeed, although Verbeek emphasizes both the emergent properties of technologies and the idea of technological intentionality, this does not mean that these technologies act *completely* independently in the sense intended by Peterson and Spahn. Verbeek in fact states that humans and technologies "do not have a separate existence anymore" (Verbeek 2008a, p. 14; see Peterson and Spahn 2011, p. 414). In this sense, technologies co-shape (and the emphasis should be on 'co-') our world and co-exist with us.

While such criticisms do not seem to impact Verbeek's reflection, certainly, as Coeckelbergh rightly observes, "some of these objections could be avoided if Verbeek would not use terms and phrases such the 'morality of things' and the 'moral agency of things' but stay with the claim that technologies mediate morality" (Coeckelbergh 2020, p. 67). Indeed, although one can understand Verbeek's use of such language in order to

challenge the "common understandings of technology" (Coeckelbergh 2020, p. 67), the post-phenomenological approach cannot refrain from constant vigilance over the type of language we should use when technologies raise increasingly urgent ethical issues. This is the direction in which this paper moves, and it aims to be an invitation to make use of the notions of intentionality and composite responsibility to address the challenges of the present, but without forgetting the ambiguity that at times still affects/constrains these notions, obscuring their ability to account for the intricate web that inextricably binds man and technology.

References

Benanti, P.
2021 *Le macchine sapienti. Intelligenze artificiali e decisioni umane*, Marietti, Bologna.

Brennan, T., Dieterich W., Ehret, B.
2009 "Evaluating the Predictive Validity of the COMPAS Risk and Needs Assessment System", in *Criminal Justice and Behaviour*, vol. 36, pp. 21-40.

Coeckelbergh, M.
2020 *Introduction to Philosophy of Technology*, Oxford University Press, Oxford.

Coleman, K. G.
2004 "Computing and moral responsibility", in E. N. Zalta (edited by), Stanford Encyclopedia of Philosophy.

Di Giulio, M.
2020 "Decisioni artificiali. La capacità di giudizio delle macchine intelligenti", in M. Bertolaso, G. Lo Storto (a cura di), *Etica Digitale. Verità, responsabilità e fiducia nell'era delle macchine intelligenti*, Luiss University Press, Roma, pp. 69-81.

Di Martino, C.
2017 *Viventi umani e non umani. Tecnica, linguaggio, memoria*, Cortina, Milano.

Doorn, N., & van de Poel, I.
2012 "Editors' Overview: Moral Responsibility in Technology and Engineering", in *Science and Engineering Ethics*, vol. 18, pp. 1-11.

Flanagan, M., Howe, D., & Nissenbaum, H.
2008 "Embodying Values in Technology: Theory and Practice", in J. van den Hoven, J. Weckert (edited by), *Information Technology and Moral Philosophy*, Cambridge University Press, Cambridge, pp. 322-353.

Floridi, L.
2022 *Etica dell'intelligenza artificiale. Sviluppi, opportunità, sfide*, Raffaello Cortina Editore, Milano.

Floridi, L., & Sanders, J. W.
2004 "On the Morality of Artificial Agents", in *Minds and Machines*, vol. 14, pp. 349-379.

Funke, A.
2022 "Ich bin dein Richter. Sind KI-basierte Gerichtsentscheidungen rechtlich denkbar?", in A. Adrian, M. Kohlhase, S. Evert & M. Zwickel, (edited by), *Digitalisierung von Zivilprozess und Rechtsdurchsetzung*, Duncker & Humblot, Berlin.

Gabriel, I.
2020 "Artificial Intelligence, Values, and Alignment", in *Minds & Machines*, vol. 30, pp. 411-437.

Giubilini, A., & Savulescu, J.
2018 "The Artificial Moral Advisor. The 'Ideal Observer' Meets Artificial Intelligence", in *Philos Technol*, vol. 31, pp. 169-188.

Gunkel, D. J.
2017 *The Machine Question: Critical Perspectives on AI, Robots, and Ethics*, The MIT press, Cambridge, Massachusetts.
2020 "Mind the gap: responsible robotics and the problem of responsibility", in *Ethics and Information Technology*, vol. 22, pp. 307-320.

Hanson, F. A.
2009 "Beyond the skin bag: on the moral responsibility of extended agencies", in *Ethics and Information Technology*, vol. 11, pp. 91-99.

Ihde, D.
1977 *Experimental Phenomenology: An Introduction*, State University of New York Press, New York.
1979 *Technics and Praxis*, Reidel, Dordrecht.
1990 *Technology and the Lifeworld: From Garden to Earth*, Indiana University Press, Bloomington.
1993 *Postphenomenology*, Northwestern University Press, Evanston.
2002 *Bodies in Technology*, University of Minnesota Press, Minneapolis.

Johnson, D. G.
2006 "Computer systems: Moral entities but not moral agents", in *Ethics and Information Technology*, vol. 8, pp. 195-204.

Johnson, D. G., & Miller, K. W.
2008 "Un-making artificial moral agents", in *Ethics and Information Technology*, vol. 10, pp. 123-133.

Kroes, P., & Verbeek, P.P. (edited by)
2014 *The Moral Status of Technical Artefacts.* Philosophy of Engineering and Technology, vol 17, Springer, Dordrecht.

Latour, B.
1993 *We Have Never Been Modern*, eng. trans. by Catherine Porter, Harvard University Press, Cambridge, Massachusetts.
1994 "On Technical Mediation: Philosophy, Sociology, Genealogy", in *Common Knowledge*, vol. 3, pp. 29-64.
2002 "Morality and Technology: The Ends of the Means", in *Theory, Culture and Society*, vol. 19, pp. 247-260.

Matthias, A.
2004 "The responsibility gap: Ascribing responsibility for the actions of learning automata", in *Ethics and Information Technology*, vol. 6, pp. 175-183.

Mitchell, L.
2001 *Baby's First Picture: Ultrasound and the politics of Fetal Subjects*, University of Toronto Press, Toronto.

Nissenbaum, H.
1994 "Computing and Accountability", in *Communications of the Association for Computing Machinery*, vol. 37, pp. 72-80.

Pacileo, F.
2020 "L'uomo al centro. IA tra etica e diritto nella responsabilità d'impresa", in M. Bertolaso, G. Lo Storto (a cura di), *Etica Digitale. Verità, responsabilità e fiducia nell'era delle macchine intelligenti*, Luiss University Press, Roma, pp. 83-99.

Peterson, M. & Spahn., A.
2011 "Can Technological Artefacts Be Moral Agents?", in *Sci Eng Ethics*, vol. 17, pp. 411-424.

Rosenberger, R. & Verbeek P.P.
2015 "A Field Guide to Postphenomenology". In R. Rosenberger and P.P. Verbeek (eds.), *Postphenomenological Investigations: Essays on Human-Technology Relations*, Lexington Books, London, pp. 9-41.

Santoni De Sio, F., & Van De Hoven, J.
2018 "Meaningful Human Control over Autonomous Systems: A Philosophical Account", in *Frontiers in Robotics and AI*, vol. 5, https://doi.org/10.3389/frobt.2018.00015.

Selinger, E., & Engström, T.
2007 "On Naturally Embodied Cyborgs: Identities, Metaphors, and Models", in *Janus Head*, vol. 9, pp. 553-584

Stiegler, B.
1998 *Technics and time 1: The fault of Epimetheus*, Stanford University Press, Stanford.

Sullins, J. P.
2006 "When is a Robot a Moral Agent?", in *International Review of Information Ethics*, vol. 6, pp. 23-30.

Tenner, E.
1996 *Why Things Bite Back: Technology and the Revenge of Unintended Consequences*, Vintage Books, New York.

van de Poel, I.
2011 "The Relation Between Forward-Looking and Backward-Looking Responsibility", in N. Vincent, I. van de Poel, J. van den Hoven (edited by), *Moral Responsibility*, Library of Ethics and Applied Philosophy, vol 27, Springer, Dordrecht, pp. 37-52.
2020 "Embedding Values in Artificial Intelligence (AI) Systems", in *Minds & Machines*, vol. 30, pp. 385-409.

van de Poel, I., Royakkers, L., Zwart, S., D. (a cura di)
2015 *Moral Responsibility and the Problem of Many Hands*, Routledge, London.

Verbeek, P.P.
2005 *Beyond the Human Eye: Mediated Vision and Posthumanity*, in P.J.H. Kockelkoren (edited by), Proceedings of AIAS Conference 'Mediated Vision', published online at: http://www.aias-artdesign.org/mediatedvision.
2005a *What Things Do – Philosophical Reflections on Technology, Agency, and Design*, Penn State University Press, Penn State.
2008 "Cyborg Intentionality: Rethinking the Phenomenology of Human–Technology Relations", in *Phenomenology and the Cognitive Sciences*, vol. 7, pp. 387-395.
2008a "Obstetric Ultrasound and the Technological Mediation of Morality: A Postphenomenological Analysis", in *Human Studies*, vol. 31, pp. 11-26.
2011 *Moralizing Technology: Understanding and Designing the Morality of Things*, University of Chicago Press, Chicago and London.
2014 "Some Misunderstandings About the Moral Significance of Technology", in P. Kroes, P.P. Verbeek (edited by), *The Moral Status of Technical Artefacts.* Philosophy of Engineering and Technology, vol 17, Springer, Dordrecht, pp. 75-88.

Wallach, W., & Allen, C.
2009 *Moral Machines. Teaching Robots Right from Wrong*, Oxford University Press, Oxford.

Wiener, N.
1988 *The Human Use of Human Beings: Cybernetics and Society*, Da Capo Press, Boston.

Winner, L.
1986 "Do Artifacts Have Politics?" in *The Whale and the Reactor*, University of Chicago Press, Chicago.

Composite Intentionality and Responsibility for an Ethics of Artificial Intelligence

The decisions, forecasts and operations produced by intelligent systems reveal new possibilities for action in various areas of society and radically transform our lifestyles. This epochal change has triggered, in parallel with a remodeling of our habits, an important process of re-semantisation of some notions traditionally ascribed to human beings, such as those regarding intentionality and responsibility. This paper addresses the process of redefining these notions with the aim to shed light on composite forms of intentionality and responsibility in the increasingly close relationship between humans and AI. To this purpose, we propose to apply in the specific field of AI the idea of *composite intentionality* developed by P.P. Verbeek and extend it to the notion of responsibility. Despite some terminological problems related to Verbeek's proposal, his post-phenomenological approach has the merit of taking the human-technology association as its main focus, helping us to better understand the moral status of artificial intelligence.

KEYWORDS: Philosophy of Technology; Post-phenomenology; Artificial Intelligence; Responsibility; Intentionality;