Artificial Intelligence and Quasi-Normativity.

Roberto Redaelli

Abstract

This paper aims to investigate the notion of digital normativity, understood as the binding force exerted on the human subject by the predictions and standards established by artificial intelligent systems. To this end, we focus our attention on the merits and limits of the recent treatment of digital normativity proposed by Fourneret and Yvert, in order to bring a decisive correction to their ideas through the introduction of the notion of quasi-normativity. In order to introduce and develop this notion in the context of enquiry into the digital sphere, we make use of Don Ihde's precise postphenomenological considerations on the nature of the artificial agent as a quasi-other. With the aid of Ihde's postphenomenological approach and the introduction of the notion of quasi-normativity, we intend to shed new light on the binding force of AI and offer some guidance for the resolution of the normative question in intelligent systems.

Keywords: Artificial Intelligence; Digital Normativity; Science and Technology Studies (STS); Postphenomenology; Normativity;

1. Normativity: from humans to AI

Today, many disciplines are devoting special attention from a variety of perspectives to the normative nature of our forms of life. From linguistics to jurisprudence, from anthropology to philosophy, from economics to neuroscience, the subject of normativity constitutes a Gordian knot of the present age, towards which the efforts of scientific and philosophical understanding are directed. Emblematic of the growing interest in this topic is the flourishing of numerous theories of normativity, which are based, in the field of philosophy, on thinkers of both the present and the past. In fact, alongside normative theories that appeal to phenomenology (See, e.g., Smith 2012; Crowell 2013, 2019, 4-26; Loidolt 2019; Heinämaa et al. 2022) or philosophical anthropology

(See, e.g., Schloßberger 2019), a number of theoretical proposals stand out on the current philosophical horizon that draw on the thought of Aristotle (See, e.g., LeBar 2008), Hegel and Kant, to name but a few examples. This is why the *vexata questio* still echoes today in various areas of philosophy, this time in reference to the normative sphere: Aristotle or Plato? Kant or Hegel?[1]

With regard to this last question, at least in the field of analytical ethics, Kant's work seems to have a certain pre-eminence. Indeed, though accused by the opposing currents of formalism and individualism, his work has, in recent years, been given a prominent place, making the philosopher from Königsberg a key reference point for all those who approach the problems of ethics and meta-ethics. As we are well aware, this veritable Kant-Renaissance was initiated by Rawls' essay *Kantian Constructivism in Moral Theory* (Rawls 1980), which inaugurated the theoretical position that, naturally, was called constructivism. According to this position, "insofar as there are normative truths, they are not fixed by normative facts that are independent of what rational agents would agree to under some specified conditions of choice" (Bagnoli 2017).

One of Rawls' most promising pupils, Christine Korsgaard, has adhered to this position, which has evolved in numerous Kantian and non-Kantian forms. Korsgaard is credited with initiating, at the Tanner Lectures on Human Values in 1992, a heated debate on what the sources of moral normativity are. The philosopher has, in fact, placed at the centre of her lectures what she has grouped under the title of normative question, expressing it in various forms: What justifies the claims that morality makes on us? Is there anything we must do? (Korsgaard 1992, 9).

A number of leading representatives of today's diverse philosophical panorama have debated over the answer offered by Korsgaard to these questions – an answer that appeals to the Kantian notion of autonomy and the distinctly existentialist concept of practical identity. Alongside the observations of C.A. Cohen, R. Geuss, T. Nagel and B. Williams, which have constituted a commentary on Korsgaard's Tanner Lectures, more recent observations of a phenomenological and hermeneutic nature have been added, through which an attempt has been made to understand the origin of the cogent force of moral reasons, and thus of their ability to guide our actions. But, in addition to the various responses offered by philosophical enquiry, there are others that, although not directly addressed to Korsgaard's text, have transcended the

[1] The question is posed by Krijnen (2019).

boundaries of the purely philosophical debate and invested disparate fields of enquiry.[2] Some of them have revitalised the ambitious Nietzschean project of tracing a genealogy of morality. This genealogy has nowadays sometimes taken on a phylogenetic point of view, based on a natural history of human morality (See, e.g., Tomasello 2016), and sometimes an ontogenetic perspective, either sociological or based on the findings of child psychology (See, e.g., Fittipaldi 2012; Tomasello 2019). In these cases, psychologists, sociologists and anthropologists have sought to identify the various sources of normativity, highlighting the role played, in the process of emergence of the normative sphere, by phenomena such as cooperation (See, e.g., Tomasello 2009), education and aggression (See, e.g., Wrangham 2019).

Enriching the framework traced by such investigations aimed directly (or indirectly) at the normative question is the problem of the moral status of artificial agents and the moral value of their actions. Artificial intelligence systems in particular raise increasingly urgent ethical issues, necessitating a discussion that considers the moral status not only of the living (human and non-human), but also of the non-living, and in this case of the technological object, too, with its (possible) capacity to convey normative demands. It is to this capacity that this paper is devoted, aiming to understand the origin and nature of the binding force conveyed by AI. To this purpose, it is first necessary to present the normative question posed by Korsgaard and to offer an overview of the main theories concerning the sources of moral normativity. With such an overview we wish to trace in advance the boundaries within which our investigation of the moral capacity of AI will take place (par. 1). Within these boundaries, we then intend to focus on the merits and limits of the recent treatment of digital normativity proposed by Fourneret and Yvert (par. 2), and finally, bring a decisive correction to their considerations by introducing the notion of quasi-normativity (par. 3 & 4). In order to introduce and develop this notion in the context of enquiry into the digital sphere, we will make use of Don Ihde's postphenomenological considerations on the nature of the artificial agent as a quasi-other. Thanks to these insights and the introduction of the notion of quasi-normativity, we believe it is possible to shed new light on the binding force of AI.

2. The normative question: some preliminary remarks

[2] For an overview of different scientific perspectives that investigate the normative question see, e.g., Funke and Redaelli (2021).

The problem of moral normativity raised by Korsgaard concerns the source of the authority of our moral standards, of their ability to guide our actions. What is at stake with this question is, more precisely, the origin of the force they exert on us in the intricate network of our everyday lives. For Korsgaard, in fact, moral reasons do not have a descriptive nature, but rather a prescriptive one, which prompts the agent, confronted with their authority, to raise questions about their justification. Hence, in the face of the pressing demands that morality makes on us, the question arises: why should I be moral? (Wrangham 2019)

This question opens up a constellation of problems concerning the right of morality to impose laws on us; a right that, it should be made clear from the outset, has very little to do with the question of whether and why it is *convenient* for us to be moral (Wrangham 2019). In fact, the authority of morality that Korsgaard seeks is unconditional, that is, independent of whether or not its injunctions are corroborated by our interests and inclinations. Thus posed, the normative question distances itself, on the one hand, from any form of utilitarianism, while on the other hand, it embraces Kant's search for the unconditioned, with which the philosopher intends to put an end to moral scepticism (Wrangham 2019), which endlessly repeats the question "why must I do that?". Korsgaard succeeds in identifying this unconditioned not through an appeal to some intrinsically normative entity, but through recourse to what she defines in terms of reflexive approval. The normative question therefore evaluates the authority of morality, avoiding bringing into play both some form of substantive realism[3] and any version of utilitarianism: it is answered by a Kantianism that is, so to speak, existential[4], as we shall now see.

Precisely because of the peculiar approach taken to the normative question and the theoretical force of her constructivism, Korsgaard's philosophical proposal has had and continues to have a loud echo in the multifaceted contemporary philosophical panorama, within which it is possible to recognise three different approaches used to address this question. W. O. Smith has offered a valuable overview of these approaches (Smith 2012, 6-7), which we would like to take up here, if

[3] Unlike substantive realism, whereby there are values out there that can be intuitively perceived, Korsgaard defines her philosophical position in terms of procedural realism: "the form of realism I am endorsing here is a procedural rather than substantive realism: values are constructed by a procedure, the procedure of making laws for ourselves" (Korsgaard 1992, 112).

[4] Responding to Nagel's remarks, Korsgaard confesses in *The Sources of Normativity* that "Nagel also makes this point, characterizing my appeal to identity as 'rather existentialist' (I think correctly) and also as therefore unKantian (I think incorrectly)" (Korsgaard 1992, 237).

only schematically, in order to define in advance the coordinates within which our investigation into the moral normativity of AI will be carried out.

The first approach, which is the one employed by Korsgaard, is defined by the author herself in the *first person*. In it, an internalist perspective is taken, characterised grammatically by the use of the pronoun 'I'. Here, it must be clarified why a certain reason is binding to the first person, i.e. from the subjective perspective of the agent. In order to offer an adequate answer to this question, Korsgaard appeals, as just mentioned, to a specific human characteristic, that of self-reflection.[5] Unlike non-human animals, which act in accordance with their desires and instincts without these instincts being the object of attention and deliberation, man cannot be content with such incentives, but needs reasons to believe and to act.[6] Such reasons, for Korsgaard, are nothing other than a type of reflective success,[7] i.e. our perceptions and desires provide reasons if they pass the test of reflective scrutiny:[8] it is, in fact, "in the internal world created by self-consciousness, that reason is born" (Korsgaard 2009). Reflection thus seems to free the human subject from the authority of instincts that dominates the animal kingdom and at the same time opens up a normative space in which the human subject binds himself to his/her own moral standards. In this sense, Korsgaard in her argument calls into play the Kantian notion of autonomy, while recognising autonomy as only a formal condition for moral action. Alongside autonomy, the author in fact adds the notion of practical identities, understood as "self-images founded in the concrete life forms" (Staiti 2015, 172). Such practical identities (such as father, mother, husband, manager) provide "constitutive standards" with respect to which we can fail or succeed, or more accurately, standards in the light of which our reasons and obligations arise. Clearly this first-person perspective, combining autonomy and practical identity, has the limitation, underlined by several critics (See, e.g., Smith 2012; Crowell 2013), of not accounting for the normative force of the other: the other, the 'you',

[5] The notion of self-reflection called into question by Korsgaard has been the subject of pressing criticism, including that of Mark Okrent, echoed in part by Steven Crowell. According to Okrent, the notion of self-reflection as presented by Korsgaard gives rise to three different interpretations, leaving this notion for the most part undefined (Okrent 1999; Crowell 2013, 246 ff).

[6] On the difference between man and animal, see Korsgaard (2009), where the author distinguishes two types of autonomy, animal autonomy "governed by the principles of your own causality" and human autonomy characterised by the choice of the principles that define our will (Korsgaard 2009, 108).

[7] From this perspective, Korsgaard observes, "scepticism about the good and the right is not scepticism about the existence of intrinsically normative entities. It is the view that the problems which reflection sets for us are insoluble, that the questions to which it gives rise have no answers" (Korsgaard 1992, 94).

[8] More precisely, Korsgaard states that "each impulse as it offers itself to the will must pass a kind of test for normativity before we can adopt it as a reason for action" (Korsgaard 1992, 91).

is reduced, as Smith correctly observed, to "a source of proto-moral reasons – analogous to bodily incentives or inclinations in practical deliberation" (Smith 2012).

In contrast to Korsgaard's approach, the second-person perspective, which Smith also espouses, is characterised by the point of view of the 'you', where it is no longer the 'I' addressing itself, but rather the encounter with the 'other' that is pre-eminent. In this perspective, the cogent force of moral injunction does not arise from self-reflection, but rather originally from the appeal, from the question that the other addresses to the self. It is a matter here, in Levinassian terms, of a face-to-face encounter with the other and a consequent interpersonal foundation of morality, which has taken and continues to take disparate forms, sometimes radically different from one another, such as those proposed by Darwall (2006), Crowell and Smith himself. While Darwall appeals to Kant and emphasises a symmetry between the 'I' and the 'you', Crowell elaborates a second-person phenomenology (See Crowell 2015, 2019), with which he emphasises in Levinassian terms a dissymmetry between the 'I' and the 'other'.

Referring specifically to Smith, it can be observed that following Crowell's line of thought, he attempts to elaborate a theory that harmoniously brings together a normative first-person moment and a second-person perspective. In this way, Smith intends to give an ontological foundation to Levinas' vis-a-vis reflection through recourse to Heidegger's notions of Mitsein and resoluteness. By virtue of this intermingling, this position assigns a central ethical role to the other, without for that reason diminishing the importance of the first-person moment.

The third-person perspective is, on the other hand, the one embodied by the figure of the disinterested spectator, who observes, so to speak, from outside, from nowhere, what happens in the world. This perspective is adopted, paradigmatically, in the field of science, where an objective view of the world is pursued. In the field of ethics, the greatest representative of this approach is, as is well known, Thomas Nagel, whose thinking leads to a form of realism, according to which there are agent-neutral reasons: from the objectivity of these reasons derives their normative force. In this and other forms of realism, the normative question seems, however, to become less interesting since the answer to this question, as Wedgwood states, "corresponds to an appropriate normative truth or fact" (Wedgwood 2007, 7).

Once we have broadly outlined the framework within which we intend to place our reflections on AI, it is appropriate to recall the recent considerations of Fourneret and Yvert (2020), which have

the merit of illuminating the normative issue in relation to AI. These considerations will constitute the starting point for our own reflection, which aims to enrich and supplement the authors' position. To this end, we shall first resort to certain instances present in the debate on the sources of moral normativity recalled here briefly and then introduce, from a postphenomenological perspective, the notion of quasi-normativity.

3. Digital normativity & artificial intelligence

The pervasive presence of AI systems in our lives is giving rise to what has been defined by Fourneret and Yvert as forms of digital normativity. By this expression, the authors mean "the ability of algorithms to establish standards that humans incorporate as what should be considered as normal in their lives and guide their actions" (Fourneret and Yvert 2020, 1). These AI-originated standards thus have a normative power, i.e. the ability to guide, to incline our actions and beliefs. They are not a mere description of what we already do and believe, but rather aim to make us act in certain directions. Algorithms are, in fact, able to retrieve trends that they recognise in the data they possess and thus create "a normalized view" of the problems they face (Fourneret and Yvert 2020). In the terminology of the debate we have just gone over, we could say that these algorithms give us reasons to act and decide in certain situations.[9]

For Fourneret and Yvert, the first form of digital normativity is closely related to a second one, which involves predictive algorithms in particular. These algorithms suggest certain behaviours on the basis of our previous behaviour and that of other users, without any consideration of the reasons for our behaviour. In this way the individual is "objectivized (normalized) by the algorithm" (Fourneret and Yvert 2020, 1). In line with the first form of digital normativity, which creates a normalised view of the problems for which AI was created, the second form leads to a normalised view of ourselves. In this sense, the algorithm reduces the individual and his behaviour to his data (and the data of other individuals), without any consideration of the motivations for his behaviour and the ethical scope of his actions. This second form of digital normativity has a recursive and dynamic character, which in our view is also present in the first, whereby

[9] It should be emphasised here that the action of technological artefacts is not susceptible to reasons explanation and therefore, as Johnson notes, they are not moral agents (Johnson 2006). However, this limitation does not prevent these artificial agents from offering reasons for human action and beliefs.

"algorithmic recommendations emanating from previous human actions in turn influence their next actions" (Fourneret and Yvert 2020, 2).

A third form of digital normativity derives from the predictive *efficiency* of algorithms. This capacity now surpasses in many fields that of humans, imposing norms to which humans adapt. This subordination to the effectiveness of the algorithms occurs even when the process by which the system arrived at the result is unclear. This raises the problem of moral and legal responsibility for the effects produced by AI. Precisely in order to deal with the so-called black box problem, there is a tendency today to point to transparency as indispensable values for the development of a trustworthy AI. In this sense, it has been proposed to introduce checkpoints and/or the repetition of stress-tests in intelligent systems.[10]

Now, the important thing to emphasise about the considerations made by Fourneret and Yvert is the role that, according to the authors, these types of digital normativity can play within the complex process of human subjectivation, whereby this term is understood to mean the "construction process leading someone to become and be aware of being a subject, i.e., being free and responsible for one's actions and at the foundation of one's representations and judgments" (Fourneret and Yvert 2020, 2).

For the authors, in fact, AI systems are capable of exerting a decisive influence on the process of subjectivation, according to a twofold dynamic: on the one hand, they can help us to stop performing certain burdensome tasks, but on the other hand, they can also lead us to stop making decisions in certain areas of our lives due to the predictive effectiveness of algorithms. One example that has become famous is that of the judge who in some legal systems uses algorithms to determine the risk of recidivism. In this case, a conflict could arise between the judge's decision and that of the algorithm, or as in the Compas case,[11] there could be biases that negatively influence the judge's decision supported by intelligent systems.

Beyond these limitations, the extraordinary possibilities offered by predictive algorithms, according to the authors' diagnosis, could activate in future generations an "AI governmentality" (Fourneret and Yvert 2020, 2) with a consequent process of de-empowerment, whereby humans would blindly rely on the results obtained by intelligent systems. In fact, although our ability to

[10] On this point, see Cristianini (2023, here: 155-159).
[11] For the Compas case, see Brennan et al. (2009).

resist the normative force arising from models established by AI always remains open (Fourneret and Yvert 2020, 2), the effectiveness of this technology could ultimately lead to an imposition, albeit non-violent, of the AI's choices. Such an imposition would take place, and already does, through a radical change in our practices of living and knowing. Faced with this risk, the authors suggest accompanying the development of AI with an ethical reflection such as the one promoted by P.P. Verbeek, which, at the level of design, identifies the value system of which the technology is the bearer and reflects on the principles to be followed to protect the process of subjectivation.

In order to highlight the strengths and shortcomings of Fourneret and Yvert's succinct considerations, it is now necessary to read the proposal put forward by the authors in the light of certain elements of the first- and second-person theories we mentioned in the previous section. First of all, artificial intelligence, by virtue of its effectiveness, presents demands with a normative character to which man is called upon to respond. In this sense, AI is recognised as the other that presents itself to the self, so that in the encounter with the face of the other, as Levinas argues and Smith emphasises, there is an ethical interdiction or the "calling into question of my freedom" (Smith 2012, 59). The authors respond to this calling into question of human freedom by appealing to an ethics of design that clearly performs the function of self-reflection proposed by Korsgaard's theory. According to this theory, the demands of the other need to pass reflexive scrutiny in order to become reasons for the 'I' to act and believe.

Following this direction of enquiry, one can thus recognise in Fourneret and Yvert's thought processes a second-person perspective, whereby moral normativity arises in the face-to-face with the other, and a first-person moment, whereby the other's injunction has no normative force until it obtains the 'I''s approval. From the intertwining of these two moments two crucial questions stand out clearly, the resolution of which can help clarify the nature of digital normativity. The first question concerns the perception of AI as the other: can the AI be treated as the other (living human and non-human)? And if the answer is yes, then do AI's injunctions to act have the same normative force as human ones? This last part of the first question is closely linked to the second question concerning the nature of digital normativity: can one base the normative force of the demands made by AI on mere efficiency, thus according to a utilitarian perspective that Korsgaard's Kantian position seems to oust from moral discourse?

In order to answer the first question, i.e. the problem of the identification of AI with the other, we shall make use of Ihde's postphenomenological analyses of the mediating role played by technologies in our lives, while for the resolution of the second question we shall employ the notion of proto-normativity or quasi-normativity, which we have already developed elsewhere (See, e.g., Redaelli 2021) and which we feel it useful to recall here in order to clarify what links (legitimately or illegitimately) the normative force of AI to efficiency.

4. Artificial intelligence from a postphenomenological perspective and the notion of quasi-normativity. Some indications for a solution to the normative question in the field of AI ethics

The postphenomenological approach inaugurated by Ihde has the undisputed merit of combining philosophical analysis and empirical investigation in the search for the relationships between human beings and technological artefacts[12]. In this type of inquiry, Ihde makes a decisive contribution to Science and Technology Studies (STS) by highlighting different forms of technological mediation. For the author, in fact, technologies mediate our relationship to the world in a polyform manner and direct our actions. By virtue of this role, the human-world relationship typically assumes the form 'human-technology-world'.

Ihde's thinking is now renowned, so a synoptical outline of the major notions of his proposal is herein satisfactory. As we know, Ihde identifies four forms of technological mediation: embodiment relations, hermeneutic relations, background relations, and alterity relations. This list, as we shall now see, is not meant to be exhaustive, and one form of mediation does not exclude the other, so the same technology can realise different forms of mediation.

The first type of human-technology relationship is characterised by the fact that, as the philosopher states, "I take the technologies *into* my experiencing in a particular way by way of perceiving *through* such technologies and through the reflexive transformation of my perceptual and body sense" (Ihde 1990, 72). In this relationship, therefore, technology becomes an extension of the human body, it becomes a quasi-me. Examples of this kind of relationship are offered by optical technologies such as the telescope or eyeglasses. Both play a mediating role between the seer and the seen, whereby "one sees *through* the optics" (Ihde 1990, 72). The same can be said

for hearing aids that become an extension of one's body. However, regardless of the different type of technology used, what characterises, according to Ihde, the *embodiment* relationship is the fact that the technology involved in such a relationship must somehow withdraw from our attention in order to be incorporated. Therefore, the closer the technology is to transparency, the more it allows the extension of "one's own bodily sense" (Ihde 1990, 74).

The second type of human-technology relationship is defined by Ihde in terms of a hermeneutic relationship. In this case, in short, technologies offer a representation of the world that requires interpretation. For this reason, Ihde speaks of a *hermeneutic* relation explicitly referring to the study of interpretation. An example is provided by the thermometer that allows us to "read" temperature through numbers (Ihde 1990, 85).

Another type of relation identified by the philosopher is the *background* relation. In this type of relation, technology gives form to our environment, although it is, so to speak, *unthematised*. The thermostat that regulates the temperature in the room and the bright light that allows us to see the objects around us are both part of our environment, although we do not have a direct relationship with these technologies. They remain in the *background*.

The last type of relationship is the one we wish to focus on most because it explicitly involves artificial intelligence: it is the *alterity* relationship. In this type of relationship, technology does not mediate our relationship to the world, rather we establish a relationship *with or to* technology (Ihde 1990, 97). Therefore, in this case, technology is the *terminus* of the relationship: technologies emerge as focal entities.

Taking the term alterity from Levinas to denote the radical difference that separates every man from every other man, Ihde (1990, 102) defines, within the alterity relation, technology as a quasi-other, referring to AI specifically in such terms. While pointing out the risks of anthropomorphizing technology by recognising a quasi-other in it, Ihde (1990, 103) emphasises that the automaton exemplified by a robot is a quasi-other to which we can pay attention, because although the robot has a different experience of the world than we do,[13] it presents itself as the other end of a possible social relation.

In order to support his argument, Ihde also cites the example of the video game, which involves not only a relationship of embodiment ("the joystick that embodies hand") and

[13] Ihde emphasises the difference between humans and robots, highlighting how the robot's experience of the world is significantly different from the human experience already at a sensory level.

hermeneutics (the context of the game, e.g. some sport analogue), but also the sense of "*interacting with* something other than me, the technological *competitor*".[14] In the competition that materialises in the video game there is, in fact, an exchange or dialogue with the quasi-other.[15]

In terms of the dialogue between man and machine, in order to better understand the alterity relation, we can think not only of robots, but also of the chatbots with which most of us now interact on a daily basis. These technologies, exemplified by digital assistants, simulate different types of conversations, respond to our requests and increasingly customise their activities according to our preferences. They apply predictive intelligence and perform an analysis of huge amounts of data through which they are able to offer us increasingly accurate suggestions, anticipating our needs. Precisely because of these characteristics, we could say that we perceive such chatbots as quasi-others in the same way as Ihde has referred to robots. In fact, (ro)bots,[16] as just mentioned, although they are not human, do not appear to us as mere tools: they are something other than us, an 'other than us' with whom we can have a social relationship. Robots and chatbots co-form our practices of living and knowing with a degree of autonomy, adaptability and interaction[17] that other tools do not have. We need only think of the extreme cases of care robots or sex robots with which individuals have relationships involving the affective sphere, or some types of chatbots that are so advanced that it is difficult to understand whether we are interacting with a human or an artificial agent.

Starting from this characterisation of robots and chatbots, we can now elaborate on a crucial point developed by Fourneret and Yvert that is linked to the question of digital normativity. Synthesising the authors' thinking, we could say that the higher the efficiency of robots in relating to us, the more the responses and demands they make will have normative force, i.e. the force to direct our actions.[18] And similarly, the more satisfactory the predictive capacity of AI, the more we will trust the results it returns, up to the point where we blindly accept the results obtained from AI. In this sense, certainly Fourneret and Yvert are right in linking the normative capacity of AI to its efficiency, with a consequent risk to the process of subjectivation. And yet, while, according to

[14] Ibid., 100.
[15] Ibid.
[16] We make use here of the spelling used by Wallach and Allen to refer to both robots and chatbots. See Wallach and Allen (2009).
[17] Although there is no consensus definition of AI, it is customary to recognise autonomy, interactivity and adaptability as characteristics of intelligent systems (see Floridi 2004).
[18] In this sense, one could observe that the moral authority of AI derives from the efficiency of its predictions.

the authors, the efficiency of intelligent systems lends normative force to the results they obtain, the possibility of taking a position in the face of such results remains open (Fourneret and Yvert 2020, 2). This capacity, mentioned by the authors only in passing and implicit in the appeal to ethics, actually shows how the digital efficiency-normativity link does not entirely resolve the normative issue in intelligent systems. In fact, if one understands normativity as the cogent force, that is, the authority that the standards (which may or may not be moral) created by predictive systems exert on us, giving us (or not) reasons to act in certain directions, it can be observed that the instances promoted by AI do not give rise to a pure form of normativity, because this only occurs after passing the test of reflection. Instead, they give rise to what we propose to call, in accordance with Ihde's terminology, quasi-normative instances, with the 'quasi' here indicating that such instances are not normative in the full sense, but neither are they morally neutral. They are, in fact, characterised by the fact that, although they require our approval in order to be normative in the strict sense, they already possess an injunctive force, linked to the predictive efficiency of algorithms, that redefines to a certain degree our space of freedom and directs our action.

An example proving the validity of this thesis, according to which AI is the bearer of a quasi-normative and *not* normative instance in the strict sense, is offered to us in the postphenomenological field by P.P. Verbeek who, in the opening of his *Moralizing Technology*, introduces his proposal – that technologies have moral significance as mediators – from the effects of the use of certain medical devices. A good example is obstetric ultrasonography (See Verbeek 2008b, 2011).[19] For the philosopher, one cannot attribute to this technology a merely functional role (that of making visible an unborn child in the womb), but rather it redefines both the ontological status of the child and our experience of them. Indeed, such technology, capable of predicting whether the child will be affected by certain diseases or not, transforms the child into a possible patient and its parents into makers of decisions about its life. In this way, medical devices redefine the relationship between child and parents in hitherto unexpected terms, whereby pregnancy becomes a process of choice (Verbeek 2011, 25). In fact, the possibility to have "sonograms made, and therefore to detect congenital defects before birth irreversibly changes the character of what used to be called 'expecting a child'" (Verbeek 2011, 25). In our terms, we could

[19] On the use of AI in prenatal diagnosis see He et al. (2021).

say that such technology is the bearer of quasi-normative demands, i.e. having a certain capacity to direct action; these demands, as already mentioned, can only become normative, i.e. binding, with human approval. In fact, while we only need to consider that since the introduction of these technologies, the very choice not to have an ultrasound scan made is also a choice, on the other hand, even in the case of serious illnesses diagnosed in the unborn child, these predictive systems do not exert such a coercive force on the parents that they are induced to abort. In this case, therefore, while it is clear that the redefinition of reality by technologies entails the emergence of a quasi-normativity – in this case, a push to choose that was not as strong before the introduction of such devices –, on the other hand, it is even clearer that any predictions that emerge from the use of such technologies lack the authority possessed by normative issues subjected to the scrutiny of reflection.

5. Brief conclusions: the hybrid origin of normativity and AI

In light of these considerations, it can be observed that the emergence of normativity, as a binding force prescribing a certain behaviour to the subject, is not attributable to technology alone but is the effect of the interrelation between non-humans and humans. In this sense, according to our hermeneutic hypothesis, normativity is inextricably linked to the quasi-normative sphere of technological apparatus, without, however, imputing to the latter a role other than that of mediation ascribed to it by Verbeek (2008a) and before him, by Ihde.[20] In this sense, it can be observed that the normative force of reasons has a relational character, so to speak: it has its terrain of origin in the relationship between self and other or, in the case of AI, quasi-other. And it is precisely this relationship, we may add as an afterword to our discourse, that gives rise to a 'we-society' composed today not only of human and non-human living beings, but also of some intelligent systems. The latter are in fact part of our society, they are quasi-others, because the relationships we have with them clearly share with other types of social relationships the fact that they, in the words of Margalit and Raz, "depend for their existence on the sharing of patterns of expectations, on traditions preserving implicit knowledge of how to do what, of tacit conventions regarding what

[20] It should also be noted here that the models proposed by the algorithms are extrapolated from our data and the algorithms themselves are produced by us. Therefore, from a genetic point of view, it can be observed that the instances raised by AI are already a product of the interrelation between man and technology. In order to understand this interrelation between humans and technology, one refers in the postphenomenological field to the notion of composite intentionality (see Verbeek 2011). On the notion of composite intentionality in the field of AI see Redaelli (2023).

is part of this or that enterprise and what is not, what is appropriate and what is not, what is valuable and what is not" (Margalit and Raz 1990). In this precise sense, intelligent systems are products of our society, of which they are the spokesmen, and hence the need for that appeal to the ethics of design that allows us to decide what demands these robots *should* make; a decision that protects our process of subjectivation, albeit only in part, from the risks of degeneration highlighted by Fourneret and Yvert (2020). In fact, as long as the disruptive force of AI is recognised as a form of quasi-normativity, man will always remain responsible for the standards he intends to adopt, and the efficacy of AI itself cannot replace human intelligence; if anything, it can enhance it by expanding the human field of freedom and responsibility. In this sense, paraphrasing the famous words of G. Anders, the possibilities opened up by technology, and more specifically by AI, *do not have to* be realised, but rather *remain open possibilities on which we have to take a stand* in the ongoing process of human self-formation.

References

Bagnoli, C. (2017). Constructivism in metaethics. In: Zalta, E. N. (Ed), *The stanford encyclopedia of philosophy* (Winter 2017 edn).

Brennan T, Dieterich W, Ehret B (2009). Evaluating the predictive validity of the COMPAS risk and needs assessment system. *Crim Justice Behav*, *36*, 21-40. https://doi.org/10.1177/0093854808326545

Cristianini, N. (2023). *The shortcut: why intelligent machines do not think like us*. CRC Press, Roca Raton/London/New York.

Crowell, S. (2013). *Normativity and phenomenology in Husserl and Heidegger*. Cambridge University Press, Cambridge.

Crowell, S. (2015). Second-person phenomenology. In: Szanto, T., Moran, D. (Eds), *Phenomenology of sociality. Discovering the 'we'*. Routledge, New York, pp 70-91.

Crowell, S. (2019). Second-person reasons: Darwall, Levinas, and the phenomenology of reason. In: Fagenblat, M., Erdur, M. (Eds.), *Levinas and analytic philosophy: second-person normativity and the moral life*. Routledge, New York, pp 4-26.

Darwall, S. (2006). *The second-person standpoint: morality, respect, and accountability*. Cambridge University Press, Cambridge.

Fittipaldi, E. (2012). *Everyday legal ontology*. LED, Milano.

Floridi, L., Sanders J. W. (2004). On the morality of artificial agents. *Minds Mach* 14:349-379. https://doi.org/10.1023/B:MIND.0000035461.63578.9d

Fourneret, E., Yvert, B. (2020). Digital Normativity: a challenge for human subjectivation. *Front Artif Intell* 3:27. https://doi.org/10.3389/frai.2020.00027

Funke, A., Redaelli, R. (2021). Rethinking the sources of normativity in ethics. *Ethics & Politics 2*, 148-151.

He, F., Wang, Y., Xiu, Y., Zhang, Y., Chen, L. (2021). Artificial intelligence in prenatal ultrasound diagnosis. *Front Med* 8:729978. https://doi.org/10.3389/fmed.2021.729978

Heinämaa, S., Hartimo, M., Hirvonen, I. (2022). *Contemporary phenomenologies of normativity norms, goals, and values*. Routledge, New York.

Ihde, D. (1990). *Technology and the lifeworld: from garden to earth*. Indiana University Press, Bloomington and Indianapolis.

Johnson, D. G. (2006). Computer systems: moral entities but not moral agents. *Ethics Inf Technol 8*, 195-204. https://doi.org/10.1007/s10676-006-9111-5

Korsgaard, C. M. (1992). *The sources of normativity*. Cambridge University Press, Cambridge.

Korsgaard, C. M. (2009). *Self-constitution: agency, identity, and integrity*. Cambridge University Press, Cambridge.

Krijnen, C. (2019). *Concepts of normativity: Kant or Hegel?* Brill, Leiden.

LeBar, M. (2008). Aristotelian constructivism. *Soc Philos Policy 25*, 182-213. https://doi.org/10.1017/S0265052508080072

Loidolt, S. (2019). Experience and normativity: the phenomenological approach. In: Cimino, A., Leijenhorst, C. (Eds), *Phenomenology and experience: new perspectives*. Brill, Leiden and Boston.

Margalit, A., Raz, J. (1990). National self-determination. *J Philos 87*, 439-461. https://doi.org/10.2307/2026968

Okrent, M. (1999). Heidegger and Korsgaard on human reflection. *Philos Top*, 27:47-76. https://doi.org/10.5840/philtopics19992724

Rawls, J. (1980). Kantian constructivism in moral theory. *J Philos*, 77:515-572. https://doi.org/10.2307/2025790

Redaelli, R. (2021). A relational account of moral normativity: the Neo-Kantian notion of we-subject. J Transcend Philos, *2*, 303-320. https://doi.org/10.1515/jtph-2021-0014.

Redaelli, R. (2023). From tool to mediator. A postphenomenological approach to artificial intelligence. In: Possati, L. (Ed), *Humanizing AI*. De Gruyter, Berlin (forthcoming 2023).

Schloßberger, M. (2019). *Phänomenologie der Normativität. Entwurf einer materialen Anthropologie im Anschluss an Max Scheler und Helmuth Plessner*. Schwabe, Basel-Berlin.

Smith, W. H. (2012). *The phenomenology of moral normativity*. Routledge, London and New York.

Staiti, A. (2015). Praktische Identität aus phänomenologischer Sicht: Korsgaard und Husserl. *Phänomenol Forsch*, 1, 171-188. https://doi.org/10.28937/1000107764

Tomasello, M. (2009). *Why we cooperate*. Cambridge University Press, Cambridge.

Tomasello, M. (2016). *A natural history of human morality*. Cambridge University Press, Cambridge.

Tomasello, M. (2019). *Becoming human: a theory of ontogeny*. Cambridge University Press, Cambridge.

Verbeek, P.-P. (2008a). Cyborg intentionality: rethinking the phenomenology of human–technology relations. *Phenomenol Cogn Sci*, 7, 387-395. https://doi.org/10.1007/s11097-008-9099-x

Verbeek, P.-P. (2008b). Obstetric ultrasound and the technological mediation of morality: a postphenomenological analysis. *Hum Stud*, *31*, 11-26. https://doi.org/10.1007/s10746-007-9079-0

Verbeek, P.-P. (2011). *Moralizing technology: understanding and designing the morality of things*. University of Chicago Press, Chicago and London

Wallach, W., Allen, C. (2009). *Moral machines: teaching robots right from wrong*. OUP, Oxford

Wedgwood, R. (2007). *The nature of normativity*. OUP, Oxford

Wrangham, R. (2019). *The goodness paradox: the strange relationship between virtue and violence in human evolution*. Pantheon, New York