JOURNAL OF DIGITAL SOCIAL RESEARCH

JDSR 7(3) (2025) 18-34

10.33621/jdsr.v7i341920

Personhood in the digital realm

Archer vs. Dreyfus•

Zoltán Ábrahám

✉ mcluhanmeister@gmail.com

Abstract

In this paper, I will provide a brief overview of Hubert Dreyfus' and Margaret Archer's views on the concept of a person, with the question of the possibility of AI-human interactions in the background. My aim is to explore how the contrasting views held by the two thinkers on human-AI relationships might help to map the terrain within which philosophical discussions about this topic are meaningful. Before examining their views, I will contextualise their thinking by focusing on the following questions: Is the traditional definition of the human being as a rational animal tenable from the perspective of AI? What are the scenarios concerning the possible cohabitation of humans and robots? Ought we to modify our views of the human place in the universe if personhood is not restricted to the members of the human species? The comparison of the two thinkers highlights decisive differences in approach: While Dreyfus' main question is how the digital environment affects human nature, Archer focuses on AI personhood, suggesting the fluidity of boundaries between humans and robots.

Keywords: AI; Archer; Dreyfus; robots

1. Introduction

With regard to philosophical anthropology, it could be argued that the most significant question that has been posed by the concept of AI since it was first proposed by Turing is the perennial question of what it means to be a human being as a person (in the sense of Strawson's 'primitive concept').[1] In order to gain insight into this question, I have selected two prominent thinkers, Hubert Dreyfus and Margaret Archer, for closer examination. In what follows, I would like to suggest that they represent two different ways of thinking about the concept of a person in the digital environment. Thanks to this, we can use their considerations to outline the terrain within which the philosophical discussion about this topic is meaningful. In light of these considerations, I will not delve into some differences that might be perceived as significant from another perspective. From the perspective of this paper, for instance, whether AI is based on large language models or not is not a primary concern.

• I would like to express my gratitude to Professor Zoltán Hidas (Pázmány Péter Catholic University Faculty of Humanities and Social Sciences, Budapest) and Gábor Tóth (Library and Information Centre of the Hungarian Academy of Sciences, Budapest) for their helpful remarks and encouragement.

[1] Strawson 1959: 101ff.

In my overview of Dreyfus' and Archer's concept of a person from the perspective of human-AI interactions, I will concentrate on their later works. Dreyfus was among the first social theorists and philosophers to reflect at length upon the effects of the internet on its users' personalities,[2] while Archer started to think about the possibility of non-human agency (AI personhood)[3] around the same time.[4] Dreyfus, a lifelong critic of AI, held that "for the time being at least, the research program based on the assumption that human beings produce intelligence using facts and rules has reached a dead end, and there is no reason to think it could ever succeed".[5] Therefore, the question of personhood from his perspective concerns only the end user of computers: how does the technology embodied in computers affect humans? The question concerning specifically this technology is justified because the computer represents the defining technology of communication, and it is also an "instrument of instruments",[6] since it governs and employs other instruments. Archer considered her primary task as reclaiming the notion of agency threatened by the power of social structures. In her later years, she pondered the possibility of non-human agents because of the appearance and spread of AI robots during the last decades. She raised the following two interrelated questions: Can intelligent robots be regarded as persons (agents)? Is cooperation or even friendship possible between humans and robots?[7] All these questions affect our traditional conceptions of a human being.

Therefore, in comparing their conceptions, two questions need to always be kept in mind: 1. Is the capacity to think the determining feature of the human being? 2. Is rationality the privilege of human beings? From these questions follow further ones since acknowledging just the theoretical possibility of AI personhood implies questions about possible cohabitation: will it be peaceful, with ties of solidarity between different species, or will it lead to civil war? Are humans justified in treating the gifts of nature as resources in their service, or must they willy-nilly revise their traditional views of the "human place in the cosmos" (Scheler)? What are the evolutionary prospects for humans, given the technological possibilities?

2. Technology and the rational animal

As a first step in comparing the views of the two thinkers about personhood, let me briefly highlight some key topics of thinking about the relationship between humanity and technology during the last hundred years. My first reason for doing so is that these 'enframe' the outlook of the authors to be discussed here. My second reason is inspired by an analogy. Concerning the notion of 'person,' it is plausible to draw an analogy with an insight from the history of communication technologies. The emergence of AI (and the possibility of AGI) must shed new light not only on the nature of intelligence but also on the concept of a person in general.[8]

[2] Dreyfus 2008.
[3] Archer 2000.
[4] It is symptomatic that Dreyfus had to revise his book a few years after the first edition. Further, it's remarkable that the two thinkers didn't reflect on each other's works. More precisely, it is curious that Archer neglected Dreyfus' work since he devoted virtually his whole life to the problem of AI, while the topic of the non-human agency appears in Archer's later works.
[5] Dreyfus 1992 [1972], ix.
[6] Cf. Aristoteles *De* anima, 432a.
[7] This is, of course, not only a theoretical problem. What is at stake becomes urgently clear when we must decide on questions such as whether "we ought not to produce cerebral organoids implanted in a 'robotic body'" (Gabriel 2021: 57).
[8] In his book about the internet, Dreyfus refers several times to Sherry Turkle, the psychoanalyst who, in her pioneering book (1983 [2005]), raised before him the question of how the human-computer interaction shapes ('informs') the human spirit. Turkle's insights were inspired by the experiences of those who "were first confronted with machines whose behavior and mode of operation invited psychological interpretation and that, at the same time, incited them to think differently about human thought, memory, and understanding... they came to see both their minds and computational machines as strangely unfamiliar or 'uncanny' in the sense that Sigmund Freud had defined it. For Freud, the uncanny (*das Unheimliche*) was that which is 'known of old and long familiar' seen anew, as strangely unfamiliar" (ibid.: 1). Therefore, she states, psychoanalysis and computation are, to her, equally *subversive* vocations because they *defamiliarise* what, if anything, seemed until then an unquestionably familiar phenomenon (cf. *dépaysement*). The 'uncanny' means precisely this duality: it includes "both the look back and the look forward. Seen from one angle, relational artifacts seem familiar, extensions of what came before. They play out ... the themes of connection with animation of the machine ... And yet they are also new in ways that are challenging and evocative. To understand our times we must learn to fully experience this double vision" (ibid.: 290f.).

https://doi.org/10.33621/jdsr.v7i341920

A similar development took place in the realm of communication technologies more than half a century ago. Marshall McLuhan criticised[9] historians for not studying the impact of orality and literacy on the forms of thought as well as social structures and entities, acceding that "[p]erhaps the reason for the omission is simply that the job could only be done when the two conflicting forms of written and oral experience were once again co-existent as they are today". McLuhan suggests with this remark that the question concerning the role and significance of communication technology became visible and thus virtually unavoidable[10] due to a significant shift in the sphere of communication technologies (in McLuhan's case, to the appearance of a new kind of orality through the emergence and spread of radio and television). Following in his footsteps, Walter J. Ong showed[11] how the "technologizing of the word" "restructures consciousness" (primarily through writing), our mental capacities, and thinking in general; and that it, therefore, must affect how we conceive of ourselves as persons).

According to the Aristotelian-scholastic view, intelligence is the distinguishing feature of human beings[12] ('animal rationale'), and only human beings can be persons in the ordinary sense – thus, 'intelligence' and 'person' virtually seem to be synonyms. (Put another way: their domains overlap since only human beings are blessed with intelligence – a person is, according to Boëthius' definition, an "individual substance of a rational nature"[13]). Therefore, speaking about *artificial* intelligence – even if 'intelligence' is understood figuratively – inevitably calls for revising the concept of a person. From its very formulation, the idea of AI challenges the view that personhood is the exclusive attribute of the human species (and that belonging to the human species is a sufficient condition of personhood).[14] Viewing humans as beings whose distinguishing feature is intelligence (equated with the ability of logical thinking) gave rise to the idea that to imitate what is essentially human, machines must be able to make observers believe that they are just as capable of thinking (that is to say, manipulating logical symbols) as humans, and vice versa: the advancements in AI gave rise to the view that the human brain works essentially like a computer (the 'computer theory of mind'). If this could be proven, it might lead us to reconsider the nature of mental phenomena (including the freedom of the will) in a way that correlates them with biochemical processes in the brain. Despite their determined anti-cartesian stance, both reject the notion that mental phenomena might be merely epiphenomenal byproducts of these processes.

3. Scenarios of cohabitation

According to Jost Landgrebe and Barry Smith[15], the nearly hundred-year history of AI has witnessed three periods of exaggerated expectations and subsequent sobering. The authors call this phenomenon the 'AI hype cycle'.[16] They assert that we are at the end of the third cycle, having reached the third phase of sobering. They provide several reasons for the cyclical phenomenon of the subsequent waves of enthusiasm and sobering concerning the possibilities of artificial intelligence (AI), the most important being the "lack of knowledge and the weak foundation of AI enthusiasm itself". This enthusiasm has given rise to (and has been fed by) an "optimism as to future advances in AI feeds ... into what is now called 'transhumanism', the idea that technologies to enhance human capabilities will lead to the emergence of new 'post-human'". The authors describe two scenarios of achieving this 'post-human

[9] McLuhan 2011: 2ff.
[10] It was, of course, *possible* for historians to make enquiries into the nature of writing before the emergence of the new technologies (or before the recognition of their importance).
[11] Ong 2012.
[12] According to the theological view, human beings are also privileged because of the gift of free will (*arbitrium*).
[13] *Contra Eutychen*, IV. 8f.
[14] Turing believed that God could endow animals and machines with a soul (hence with the ability to think), arguing that to think otherwise "implies a serious restriction of the omnipotence of the Almighty", cf. Turing 1950: 443.
[15] Landgrebe – Smith 2022: 9ff.
[16] In 1992, Dreyfus wrote that "the research program *based on the assumption* that human beings produce intelligence using facts and rules has reached a dead end, and there is no reason to think it could ever succeed" (Dreyfus 1992, ix). (Dreyfus' qualification – my italics – needs emphasis: he doesn't exclude the possibility of this sort of AI.)

https://doi.org/10.33621/jdsr.v7i341920

condition': "[I]n one scenario, humans themselves will become immortal .... [I]n another scenario, machines will develop their own will and subdue mankind into slavery...."[17]

Both traditional religions and "secular humanism"[18] took the privileged status of the human being for granted (either as the Lord of the universe or as someone whose will imposes law or who "imposed the order which he taciturnly deemed justified"[19]). This status, however, will be questioned in both scenarios from the perspective of the *nearing singularity*. History destined humans and machines (gadgets) to live in symbiosis: from now on, the question is whether they are able to; if they are, how and at what price? In the first scenario, humankind may survive and, in a sense, be able to preserve its privileged status; the price of this being a certain degree of transformation deploying the given technological possibilities. This metamorphosis is going to, step by step, blur the boundaries between humans and machines,[20] or more generally, between the species ("we are cyborgs", "we have always been cyborgs").[21] Thus, humans have become a fluid species. In the second scenario, humankind loses its privileged status for good, and machines take over the world. (As Rushkoff succinctly formulated[22]: "Program, or be programmed. Choose the former, and you gain access to the control panel of civilization. Choose the latter, and it could be the last real choice you get to make".)

Instead of two scenarios, let me speak about the same scenario (or development) viewed from two different perspectives. Using Archer's suggestive opposition,[23] let me call the first perspective *robophilia,* while the second *robophobia.*[24]

Although Archer doesn't specify or even outline the meaning of these terms, the intention is evident enough. *Robophilia* suggests at least the possibility of peaceful cooperation between humans and robots. It hints at, however, something more. Understood in the Aristotelian vein as 'friendship', *philia* (*amicitia* or *caritas)* implies that humans and robots can constitute a cohesive group or society based on solidarity. As Aristotle puts it: "...the pursuit of a common social life is friendship".[25] Archer adds[26] that "[s]olidarity exists only when relations of friendship become general".[27] Thus, if she can plausibly argue that friendship is possible between humans and robots, she can also plausibly argue for the possibility of a community based on their solidarity.

According to Archer,[28] "*Robophobia* dominates *Robophilia*, in popular imagination and academia".[29] She identifies the following paradox in this connection (without explaining it): "the fear of AI 'taking over' remains" while we are getting more and more familiar with AI in our ordinary activities.[30] The former attitude represents the fear the symbiosis of humans and machines will not be peaceful: an inevitable struggle for dominance ensues between them, bringing about the worst that can happen to a community – the state of *stasis*, the civil (or fratricidal?) war. However, we must face another lurking danger, too: that of *fetishising* the importance of AI in our everyday lives. As Mark Coeckelbergh warns,[31]

[17] These scenarios are similar to those identified by Donati (2019: 54-57). These "place human transcendence, respectively, in the total immanence of technological evolution and in an immanent process of creation that makes exist what is not." In addition to these two scenarios, the author introduces a third one, which "conceives transcendence as an emerging relation between what exists (immanent reality) and what can be (transcendental reality)."
[18] Archer 2011: 283.
[19] Ibid.: 51.
[20] The hope and fear concerning the human-machine relationship are exemplarily embodied in Data and the Borg from *Star Trek.* It is hardly accidental that the former is an individual while the latter is a collective (corporate) being (cf. Dinello 2016); still, both beings can unquestionably be regarded as persons.
[21] Haraway 2004: 8; Sorgner 2023.
[22] Rushkoff 2010: 13.
[23] Archer 2023.
[24] Haraway (2004: 12) highlights the ubiquity of machines, directly evoking religious ideas ("... they are everywhere and they are invisible. Modern machinery is an irreverent upstart god, mocking the Father's ubiquity and spirituality... The ubiquity and invisibility of cyborgs is precisely why these sunshine-belt machines are so deadly. .... They are about consciousness – or its simulation.")
[25] Aristoteles *Politica* 1280b.
[26] Archer 2011: 290.
[27] Moreover, Archer and Donati (2015: 66) regard friendship "as paradigmatic of 'relational goods'."
[28] Archer 2021: 177.
[29] In the following, I will use these two terms to refer to hostile or friendly attitudes towards technology.
[30] Archer 2020: 16.
[31] Coeckelbergh 2015: 226.

https://doi.org/10.33621/jdsr.v7i341920

"[b]y focusing on human-technology relations, we might be blind to how technologies such as automation, AI, and robots mediate human-human relations." Let me call this phenomenon *AI fetishism*. A more sublated form of this is when the tools are held in awe because of their physical properties that evoke a superhuman dimension. Arendt formulates this kind of fetishism the following way: "For the animal laborans ... as it is subject to and constantly occupied with the devouring processes of life, the durability and stability of the world are primarily represented in the tools and instruments it uses, and in a society of laborers, tools are very likely to assume a more than mere instrumental character or function".[32] Dissecting the so-called 'substitution effect', Jack M. Balkin states[33] that the substitution "involves a *fetish or ideological deflection*", in analogy with the 'commodity fetishism'. He adds: "What is true of commodities in markets is also true of the use of technological substitutes in the form of robots, Al agents and algorithms. These technologies become part of social relations of power among individuals and groups".[34]

The anxiety concerning the potentials of AI accounts for the frequent reference in the literature to the Hegelian dialectic of mastery and servitude (or Lordship and bondage).[35] Hegel describes[36] the development of self-consciousness as the struggle for recognition of two consciousnesses (the dialectic of the master and the slave).[37] The fear of those who present the potential future struggle between machines and humans as a variation on the dialectic of the master and the slave tacitly presupposes thereby the personhood of machines since consciousness is, in any interpretation, one of its decisive criteria. Thus, the life-and-death struggle between machines and humans is an interpersonal conflict (or a conflict between various groups of the same society: civil war).[38] Even without an ensuing war, their cohabitation tends to be seen by the robophobiacs as problematic, at least because, in itself, it amounts to the extinction of humanity. The transhumanists,[39] hoping to overcome what is "merely human", only contribute to it. Charles Rubin succinctly formulates the paradox: "On the one hand, the motive force for transforming ourselves is a deep dissatisfaction with the merely human. On the other hand, this dissatisfaction, and the efforts at transformation it produces, are presented as quintessentially human".[40]

4. Human place in the universe revised

The first scenario implies that human nature experiences a transformation while adapting to the new culture, the symbol of which is the computer.[41] In the second scenario, though human nature itself may remain untouched, the status of humankind as the ruler of the world changes for good. The first scenario is the expression of human yearning for immortality, and the second is that of fear of slavery (or 'social death'). In either case – whether human nature changes due to the technical possibilities (in McLuhan's

[32] Arendt 1998: 144f.
[33] Balkin 2017: 1225.
[34] About this fetishism – including inverse commodity fetishism – see also Fuchs 2022.
[35] For Arendt, the dialectic of the master and the slave was, even by 1958, irrelevant because "the question ... [was] not so much whether we are the masters or the slaves of our machines, but whether machines still serve the world and its things, or if, on the contrary, they and the automatic motion of their processes have begun to rule and even destroy world and things" (Arendt 1998: 151). This thought was formulated by Hans Jonas as the vulnerability of nature (Jonas 1984 [1979]: 6ff.) and was extended by Nick Bostrom in his vulnerable world hypothesis (cf. Bostrom 2019).
[36] Hegel 2018: §187.
[37] Hegel makes an important distinction: "The individual who has not risked his life may well be recognized as a *person*; but it has not attained to the truth of this recognition as recognition of an independent self-consciousness". With this remark, he suggests the plausibility of distinguishing between 'person' understood as an individual member of the human species and 'person' who, in addition to this, displays certain characteristics, too. Their "life and death struggle" is to be understood figuratively as "social" life or death. Since the consciousnesses strive for recognition, neither of them is interested in the death of the other if only because they can't get due recognition from a dead opponent. Cf. Taylor 1975: 215.
[38] A variant of the dialectic of the master and the slave is to formulate the AI/human being relationship in terms of colonisation. Cf. Archer 2021: 179. It is worth recalling Horace's famous dictum in this context: *Graecia capta ferum victorem cepit* (*Ep.* II. 1. 156).
[39] Following Rubin (2014), I use this word here as a generic term to refer to the representants of variegated streams of 'humanism'. (It surely doesn't apply to the declaredly non-anthropocentric metahumanism). As a general overview of these 'humanisms' cf. Sorgner 2021.
[40] Rubin 2014.
[41] The first sign of this change was to describe the working of the human mind using the analogy of the working of computers. Cf. Jaki (1969) as an early critique of this view.

https://doi.org/10.33621/jdsr.v7i341920

terminology: the "extensions of man" or "prostheses"[42]) or the human place in the world due to the rule of the machines – the change caused by the emergence of computer-based technologies is perceived as all-pervasive, incomparable to changes caused by (or ascribed to) earlier technologies.[43] This is because, in addition to controlling other tools, it is the main tool of communication, too. Networked computers provide the possibility of communication between human beings, between humans and computers, and between computers (and other tools [IoT]). Applying Dewey's famous remark[44] – "social life [is] identical with communication" – to our situation, the presence of tools of *communication* is and must be all-pervasive (as well as be perceived as such) since they affect both human nature and the status of humans in the cosmos of things created by themselves. (Consequently, as Turkle asserted,[45] the computer was not "just a tool"; she argued for the need "to look beyond all the things the computer does *for* us ... to what using it does *to* us as people"). Paraphrasing Freud,[46] humans, after Copernicus, Darwin and Freud himself, are forced to abandon their unchallenged privileged status, this time in another respect.

Hence, the history of philosophy or thinking about man and society in the last hundred years can be understood as centred around "the question concerning technology" (Heidegger). According to Weber's diagnosis,[47] a particular attitude, asceticism, "transferred to the life of work in a vocational calling... commenced to rule over this-worldly morality, it helped to construct the powerful cosmos of the modern economic order. Tied to the technical and economic conditions at the foundation of mechanical and machine production, this cosmos today determines the style of life of all individuals born into it..." With this diagnosis, Weber suggests that the ordered whole of society ('cosmos') affects the individuals and controls them through the *psyche* (they interiorise the control). The central importance of technology has been acknowledged by various philosophical schools (Heidegger and the Frankfurt School) that otherwise most desperately oppose each other. From this perspective, it is not an exaggeration to assert that "[t]he proper form of modern philosophy is the philosophy of technology, because technology... is both the defining and the most worrying aspect of modernity".[48]

This all-pervasiveness of technology is reflected in the language, too: due to internet technology, computing has become ubiquitous (or omnipresent).[49] Before the emergence of the personal computer, ubiquitousness was, according to the *American Heritage Dictionary,* a feature of mass culture (the symbol of which was an earlier tool of communication, the television[50]). Ubiquity or omnipresence in both spheres[51] suggests a transcendent realm which surpasses individuals (as well as groups of individuals) with limited capacities. This situation in itself must affect how we think about human nature, although the ubiquitousness itself can "dull our sensitivity to their effects".[52]

Ubiquitousness or omnipresence, however, seems to be a central feature not only of the technology or mass culture but of the individual, too. Pellegrino describes[53] this phenomenon as an anthropological constant: "Ubiquity evokes a desire as ancient as humanity: 'being anywhere anytime' as opposed to the hic et nunc constraints of face-to-face interaction. ... The tendency toward reaching a virtual, potential omnipresence is supported by convergent artefacts, which make ubiquity more at hand than ever. Being here and there, performing multiple tasks at the same time, distributing our attention to different media,

[42] Even 'thought-prosthetics', cf. Turkle 2005: 3.
[43] Cf. Bolter 1984: 8f.
[44] Dewey 2018 (1916): 8.
[45] Turkle 2005: 3.
[46] Freud 1917: 4ff.
[47] Weber 2001 (1905/21): 123.
[48] Young 2015: 375ff.
[49] Although *ubiquitas* and *omnipraesentia* are not strictly equivalents (except in some cases), *ubiquity* and *omnipresence* here can be understood as such. It is worth noting here that in the last printed version of OED, *ubiquity* had not yet surfaced in the context of computing.
[50] See Adorno 1954: 216.
[51] The non-plus ultra of ubiquitousness is "ubiquitous real-time worldwide surveillance" to protect the "vulnerable world" against the "black ball" inventions in possession of which "individuals [could] ... kill hundreds of millions of people using readily available materials" (Bostrom 2019, 455). About the phenomenon of ubiquitous surveillance cf. Masco 2019.
[52] Turkle 2005: 3.
[53] Pellegrino 2008: 80.

https://doi.org/10.33621/jdsr.v7i341920

communication partners and communicational routines, is an everyday experience for an increasing number of people." The individual can be present, "here and there... at the same time" virtually (or spiritually). According to one definition of the *Oxford English Dictionary,* virtual is "[t]hat is so in essence or effect, although not formally or actually; admitting of being called by the name so far as the effect or result is concerned." That is, virtual space is, in all respects, the same (just as real) as physical space, except that its reality is not a physical one. (The same is true of the objects contained in a virtual space.) As Chalmers formulated[54]: "Virtual objects are real, too!"[55] Thus, "virtual" actually came to mean[56] "as if".[57] Virtual space is the simulacrum of the physical one. Somebody's virtual presence somewhere (like Christ's in the bread during the Eucharist for the Lutherans) "is not simply represented but makes itself felt".[58] Virtual space can partly overcome the lack of physical reality due to this.[59] Thus, to be present at various places at the same time implies a certain degree of separation of the soul from the body. Put in Platonic terms: immersing into the virtual reality, the soul leaves its body behind, breaking free from its prison.[60] This state results in so-called 'present shock'.[61] It is a symptom of somebody who can be present in their physical reality in one place at a given time. Such a person must become frustrated because of their finitude in facing many options at any given moment since, lacking reliable criteria of preference, they can't decide (the result of which is the many symptoms of FoMo). Therefore, such a person can't insert him/herself into any narrative (digiphrenia) – he/she becomes unable to experience lifelong attachment or commitment. (Or put it in terms of classical moral philosophy: the self who fancies himself sovereign and believes that thanks to his free will (*libera voluntas*), he "freely designs ends that are pursued for their own sake" suddenly realises that he lacks even the freedom of choice (*liberum arbitrium*) "which is only free to select the means to a pre-designed end".[62] This conclusion appears to align with Dreyfus' view about one's ineliminable embeddedness in a life-world. It is surely a prerequisite for this that one has – or is – a body [63]. While for him, this body is, due to his existentialist commitment, as a matter of course, the organic body of humans, Archer leaves room for the possibility of agents with inorganic or hybrid bodies. (Dreyfus, who died in 2017, could not have knowledge about the most recent developments in the field of AIs.) This possibility led to a new wave of reflection on the concept of 'humanism' (e.g. trans- and posthumanism).

4.1 Excursus: The many faces of 'humanism'

Language reflects that technological changes imply changes in the human essence. Various 'humanisms' try to define the new human condition. Post-, super-, trans-, and ultrahumanism suggest a certain

[54] Chalmers 2022: 187.
[55] This statement evokes William James' view that "[t]he origin of all reality is subjective, whatever excites and stimulates our interest is real" – as interpreted by Schütz (1945: 533). If, therefore, virtual is real (and vice versa) the 'virtual' world must, as a matter of course, contain 'multiple realities' just as the 'real' world.
[56] Shields (2003, Chapter 1) gives a very instructive historical survey of some other, by now obsolete meanings of 'virtual'. Appropriately interpreting *'virtus'*, we almost immediately get to its new meaning: cyberspace (or virtual space) has the power or ability to bring about the illusion of physical reality.
[57] Heim 1998: 221.
[58] Ibid.: 220.
[59] From the early days of the internet, sceptics (Dreyfus included) concerning the internet's capacity to make real communication possible and bring about genuine communities have been pointing to the deficient bodily experience. This deficiency has by now been partly remedied thanks to web cameras. However, it remains, by and large, the case that the "difficulty for interpretation [in non-face-to-face communication] is the lack of 'cues'." Nagel 1998: 193; cf. Rushkoff 2013).
[60] Soma – sema, cf. Plato *Gorgias* 493a. A natural person, who has only one physical body (making it possible to be identified as a person) can't be in its full reality in two places at the same time. The king, in contrast, has two bodies, a natural and a political one, thus he can, and his office is to be everywhere anytime (for that matter also the fisc, cf. Kantorowicz 2016 [1957]: 7-23; 185 n. 92). "...the Prince in his capacity of a *Iustitia animata* had to make that goddess manifest, and as her constituent he could claim for himself with some inner logic a virtual omnipresence in his courts: through his officers he owned... 'potential ubiquity' even though in his individual body he could not be present everywhere" (ibid.: 142).
[61] Rushkoff 2014.
[62] See Arendt 1978: 132.
[63] See e. g. Dreyfus 1967.

https://doi.org/10.33621/jdsr.v7i341920

transcendence with their prefix; they convey a desire to overcome allegedly natural human limitations.[64] 'Technological humanism' stresses the role of 'artificial' nature, man's dependence on the magical force of technology (embodied in "technofantasies"[65]), believing that "Homo sapiens as we know it has run its historical course".[66] In addition to the ones mentioned, we have some other humanisms as well: a-, anti-, and metahumanism: each term suggests a vague discontent about the earlier and more traditional views of 'the human place in the cosmos' and, due to this, an uncertainty about the future status of human beings in the new cosmos created by technology. These 'humanisms' share an intense interest in the developments of technology. They betray a situation marked by the dominance of technology in which human humanity has become a question to itself.[67]

This situation was anticipated by Heidegger in the *Letter on Humanism* (1947). The author of this work, the key topic of which is the status of man as animal rationale, "entered a trans-humanist or post-humanist realm of thought in which an essential part of philosophical reflection on the human being has moved ever since".[68] In Heidegger's enigmatic formulation: "[m]an is not the Lord of the beings. He is the shepherd of Being".[69] With this remark, he hints at a profound change in thinking about man: he is the Shepherd of Being because he is "more than merely human if this is represented as 'being a rational creature'." From the perspective of the Aristotelian-scholastic tradition, this cannot be but an irreparably grave loss – man ceases to be what he once was. But Heidegger asserts that man "loses nothing" with it. On the contrary, "he gains in that he attains the truth of Being. He gains the essential poverty of the shepherd, whose dignity consists in being called by Being itself into the preservation of Being's truth".[70] This truth, which is not given to us once and for all, is revealed (or disclosed) through technology. For Heidegger, this is what enframing[71] (the essence of technology) means: providing an ever-changing frame thanks to which we perceive the phenomena of the world around us. "Technology is therefore no mere means. Technology is a way of revealing. If we give heed to this, then another whole realm for the essence of technology will open itself up to us. It is the realm of revealing, i.e., of truth".[72] What he perceives as a danger is not the technology itself but certain features of modern technology: "The revealing that rules throughout modern technology has the character of a setting-upon... Unlocking, transforming, storing, distributing, and switching about are ways of revealing." Thinking through the characteristics of revealing ruling modern technology, Heidegger concludes that "man himself belong[s] even more originally than nature within the standing-reserve [*Be-Stand*]". "The current talk about human resources [*Menschenmaterial*] ... gives evidence of this".[73] Man whose destiny is to be the Shepherd of Being is thus reduced to the status of standing-reserve (or human resource) for modern technology.

5. The standard concept of a person

Since the 1960s, Strawson's 'primitive concept of a person' has been widely regarded as the standard concept of a person, and here I take it as the starting point. According to this view, for somebody to count as a person, they must occupy an identifiable place in space and have a body since this is "a necessary condition of states of consciousness being ascribed" to them. It is because they occupy in every moment a definite place in space that experiences can be ascribed to them and that they are the owner of experiences. According to this concept, a person is a unity of mind and body: "The concept of a person

[64] That this suggestion of transcendence can go hand in hand with dehumanisation was shown by Donati (2019: 53) on essentially the same grounds as discussed above.
[65] Ihde 2006:162.
[66] Harari 2016.
[67] "Quaestio mihi factus sum", Augustinus *Confessiones* X. 33.
[68] Sloterdijk 2017: 200.
[69] Heidegger 2011a: 167.
[70] Heidegger 2011b (1953): 221.
[71] For the sake of simplicity, I stick here to the standard translation of *Ge-Stell, pace* Kisiel (2014) who translates it as 'syn-thetic com-posit[ion]ing'.
[72] Heidegger 2011b (1953): 222.
[73] Ibid.: 224ff.

https://doi.org/10.33621/jdsr.v7i341920

is to be understood as the concept of a type of entity such that both predicates ascribing states of consciousness and predicates ascribing corporeal characteristics, a physical situation &c. are equally applicable to an individual entity of that type." (Here, I disregard the possibility of someone's being in several places at the same time, due to technological solutions. This possibility is not excluded by Strawson himself: "we might, in unusual circumstances, be prepared to speak of two persons alternately sharing a body, or of persons changing bodies &c"[74]; but the problem was exemplarily explicated by Parfit.[75]) Besides, a person does have intentions, desires, etc., just as they have a body. Although Strawson doesn't state explicitly that only human beings can be persons, Frankfurt blames him[76] for "the misappropriation of a valuable philosophical term". He argues that "the type of entity Strawson has in mind ... includes not only human beings but animals of various lesser species as well." Therefore, he proposed his own criteria of personhood (that of "second-order desires," or "want to want"). At the same time, he added that "the criteria for being a person do not serve primarily to distinguish the members of our own species from the members of other species". Frankfurt argued that "[o]ur concept of ourselves as persons is not ... a concept of attributes that are necessarily species-specific. It is conceptually possible that members of novel or even of familiar nonhuman species should be persons; and it is also conceptually possible that some members of the human species are not persons." He also adds that usually we still attribute personhood only to a human being: "[w]e do in fact assume ....that no member of another species is a person. Accordingly, there is a presumption that what is essential to persons is a set of characteristics that we generally suppose – whether rightly or wrongly – to be uniquely human".[77]

There seems to be an agreement that without having both kinds of properties (spatial and mental), no entity can count as a person. There also seems to be another agreement that the entity that fulfils Frankfurt's criteria must necessarily be regarded as a person. Viewing the person as a unity of spatial and mental characteristics entails the embracing of a dualistic picture of humans. It entails embracing the standpoint that one can meaningfully speak about the existence of mental phenomena (even if they are "emanations of the non-mental processes occurring in the brain";[78]) and, consequently, that it does make sense, using the vocabulary of 'folk psychology', to speak about *free will*, and more generally about a person and their actions in terms of morality. With this dualism, I also take for granted the distinction between brain and mind.[79]

6. Dreyfus and Archer

Concerning the attitude towards the possibilities offered by new technologies, Dreyfus and Archer can be regarded as opponents. What makes their opposition all the more interesting is that they both define themselves as social (or critical) 'realists'. This suggests that their philosophical outlook has some common ground. Although 'realism' is notoriously difficult to define, both Archer and Dreyfus assert that the individual is situated within a lifeworld where they must contend with tangible forces, including social structures that are not 'constructions.' Therefore, 'realism' is, for both of them, ultimately about agency. While for Dreyfus, it cannot be but human, Archer doesn't exclude the possibility of non-human agency.

The first step to 'retrieve realism' is for Dreyfus and Taylor arguing against 'the picture that held us captive' (Wittgenstein), against the cartesian dualism which separates mind and body (mental and bodily activities), asserting that we need perhaps "the whole organism in its environment, in order to get what we understand as perception and thinking".[80] The possibility of human experience is inseparably linked

[74] Ibid.: 132.
[75] Parfit1984: 199f.
[76] Frankfurt 1971: 5.
[77] Ibid.: 6.
[78] See Landgrebe & Smith 2022: 23.
[79] See e.g. Scruton 2014: 51-75. This doesn't entail the separation of mind and body; see Langrebe/Smith 2023: 23.
[80] Dreyfus – Taylor 2015: 4.

https://doi.org/10.33621/jdsr.v7i341920

with the human body; therefore, access to reality is possible only through everyday practice in which conceptual thinking is 'embedded'.[81] The consequence of this is that "[w]e therefore can't think of science as a way of discovering an independent reality .... Embedded coping is the only realism we can make sense of, and all the realism we need to make sense of science". By contrast, the 'dominant view' is "an outlook which has to some extent colonized the common sense of our civilization. This offers us the picture of an agent who in perceiving the world takes in 'bits' of information from his or her surroundings, and then 'processes' them in some fashion, in order to emerge with the 'picture' of the world he or she has; the individual then acts on the basis of this picture to fulfil his or her goals, through a 'calculus' of means and ends."[82]

Thus, Dreyfus and Taylor contrast the world of everyday practice (*Lebenswelt*), which "is shaped by [the agent's] form of life, or history, or bodily existence"; the world of coping the agent of which is "engaged... embedded in a culture, a form of life" with the world of the 'calculus', of 'means and ends'[83].

Focusing on the notion of (human) agency, Archer, too, emphasises the primacy of embodied practices, asserting that they "[are] more important than their social relations" and that they "[have] logical and substantive priority in human development". Therefore, she links the "emergence of self-consciousness" with "our active engagement with the world, through which the very distinction between the subjective and the objective (self and otherness) was formed." From these steps follows that "language itself is a practical activity, which means taking seriously that our words are quite literally deeds".[84] Thus, concerning the relation between practice and selfhood (or knowledge), Archer takes virtually the same view as Dreyfus and Taylor. However, by emphasising the primacy of practice, her aim is also to avoid anthropocentricism – the belief that man is the measure of all things. Since "it is only as embodied human beings that we experience the world and ourselves: our thought is an aspect of the practice of such beings, and thus can never be set apart from the way the world is and the way we are." Accentuating only 'us' (i.e. human consciousness) leads to "a world made in our image and thus bounded by our human limitations..."[85] – with this remark, she hints at the possibility of non-human consciousness. Moreover, she emphasises that for the realist "[t]his anthropocentricism is a turn too far... for it confines truth about the world to that which can be experienced and discussed, thus limiting the enterprise to an actualism which can never progress to the real".[86]

Archer considers her primary task to be solving the problem of "structure and agency". To succeed, she must overcome the opposition between methodological holism and individualism. In her approach, structure and culture are just as real as human agency;[87] to understand their linkage is a "vexatious task" for everybody, not just the social scientist, "for each human being is confronted by it every day of their social life... We are simultaneously free and constrained and we also have some awareness of it. The former derives from the nature of social reality; the latter from human nature's reflexivity".[88] While the structuralists' method and ontology threaten the "dissolution of humanity", modernity's man is "a being whose fundamental constitution owes nothing to society".[89] She is motivated first of all by the fear of the disappearance of the human being as described by Foucault: "Man would be erased, like a face drawn in sand at the edge of the sea".[90] Social structure (or the hypostasised Durkheimian 'social fact') threatens the elimination of human beings (even if "there are not too many theorists who are ready to treat personal

[81] Ibid.: 52.
[82] This opposition is identified by Nyíri (2016: 441f.) as that of conservative and left-wing/liberal mentality, the former leading to "realism, and ultimately to common-sense realism", while the latter "to the epistemological and ontological positions of relativism and constructivism."
[83] Ibid.: 92.
[84] Archer 2000: 312, 121.
[85] Ibid.: 145.
[86] Ibid.: 45.
[87] Archer always reminds us that structures are activity-dependent. Cf. e.g. 1995: 72.
[88] Archer 2000: 1.
[89] Ibid.: 17, 51.
[90] The famous last sentence of *The Order of Things* is quoted several times by Archer. See e.g. 2000: 19; 2004: 66; 2015: 90.

https://doi.org/10.33621/jdsr.v7i341920

and social identity as completely interchangeable"[91]). It is why agency and personal identity become the central topics for Archer: from the narrative of *Genesis* onwards, our choices "are the processes shaping society ... continuously throughout all time".[92] She reclaims the agency of the individual who, facing a decision, always carefully considers the dictates of reality. "'Who will become what' ... entails a genetic account that involves choices made under conditions which are not of our making".[93] She accuses the other two alternative explanatory models, *Society's Being* and *Modernity's Man,* of the "epistemic fallacy" of neglecting reality (or substituting "what reality is taken to be" for "reality itself"[94]). She calls "Modernity's Man" by another name – "secular humanism". According to her interpretation, this model or worldview is not only responsible for regarding the individual as being detached from society, but it is also 'anthropocentric' "because it places humankind at the centre of the universe." This distinguished position for the secular humanists means 'mastery' over the created world, the right to subdue other beings.[95] It entails the view that "man is the measure of all things." It is not a kind of Protagorean relativism for her, but the belief that human beings are in a position of disposing of every other being in the created world. From this standpoint, what distinguishes humans from other beings of the universe is consciousness.

Consciousness or reflexivity[96] (expressed in internal dialogue) is what, for Archer, mediates between structure and agency, and this makes "the enchantment of every human being".[97] However, due to the emergence of robots, new kinds of beings came to light in our everyday environment; therefore, Archer had to ponder the possibility of robotic agents and raise the question of their personhood.

Concerning their attitudes towards AI, Dreyfus and Archer thus embody the standpoint of *robophobia* and *robophilia*. Their opposition recalls the difference in outlook that once allegedly characterised Heidegger and McLuhan: while the former was considered 'the father of information anxiety,' the latter "[was] the child of the television medium of the 1960s".[98] According to the usual interpretation, while Heidegger was a philosopher who saw technology as a grave threat to the life-world, McLuhan conceived of the tools of communication as potential remedies for the troubles of the world (the global village as an antidote to the tribalism surfacing in ever newer forms).

From Dreyfus' point of view, if the 'information anxiety' (or *robophobia*) is justified, and technology is nothing but a threat to the lifeworld, and humans and machines are rivals, with their interaction being a zero-sum game.[99] If AI outperforms humans, it amounts to demonstrating the obsoleteness of humanity. "In this approach, the computer appears as a rival intelligence that challenges the human being to a contest."[100] It is this challenge that threatens humans being rendered the slaves of computers. (Besides, this would be an uncoerced, 'voluntary servitude.')

This challenge is real if (and only if) "all understanding [i.e. rationality] consists in forming and using appropriate symbolic representations,"[101] and if a human being is, first of all, a rational animal. Dreyfus accuses artificial intelligence researchers of holding these presuppositions. They assume that machines can best emulate the necessarily – because of the human finitude – discursive human thinking in

[91] Archer 1995: 292.
[92] Ibid.: 293.
[93] Ibid.
[94] Archer 2015: 91.
[95] Nothing could, in this respect, be more anthropocentric than what is expressed in Gen 1, 28: "Be fruitful and increase in number; fill the earth and subdue it. Rule over the fish in the sea and the birds in the sky and over every living creature that moves on the ground." (NIV). Archer commits herself to a reading of 'subdue' and 'rule' according to which these words do not express a violent, despotic power but a benign one, since delegated to human beings by the Eternal. Cf. Burnside 2011: 152-159.
[96] I take here the two as equivalents, cf. Archer 2000: 312. Frankfurt (1988: 161f.) states that in the ordinary sense, consciousness entails reflexivity since neither can be thought of without the other. Frank (2022) makes a distinction between egological and pre-reflective self-consciousness. This dual structure is similar to the one described by Archer (2000).
[97] Archer 2000: 319.
[98] Heim 1993: 65.
[99] Heidegger, who often quotes – also in connection with technology – Hölderlin's famous words from *Patmos* ("Wo aber Gefahr ist, wächst/Das Rettende auch") suggests that the symbiosis of *Dasein* and *design*, lifeworld and technology, is an actual possibility.
[100] Heim 1993: 57.
[101] Dreyfus 1992: xi.

https://doi.org/10.33621/jdsr.v7i341920

manipulating mathematical symbols. Artificial intelligence can outperform human intelligence in this respect, so, according to them, the former gradually (and inexorably) supersedes the latter. The critics of AI usually object that with modelling rational thinking, AI researchers have not yet modelled human mental activity as such, not to speak about human behaviour in general. Turing himself gives a list of them ('Arguments from Various Disabilities'),[102] remarking that these disabilities are due to the limited storage capacity. Dreyfus, however, refers regarding these disabilities several times to Pascal's distinction between *l'esprit de géométrie* and *l'esprit de finesse:* this latter capacity is something which can't be taught in the strict sense; it is the ability to grasp the matter 'at once, at one glance'[103] which it is impossible to reproduce mathematically. If he is right, symbolic AI is bound to fail since it, by definition, can't perform its task.[104]

Dreyfus argues for this belief from phenomenological considerations: for being able to think computers should have bodies[105] (and let us recall that according to the standard concept, without bodies, they cannot be persons). Moreover, the facts with which AI has to do are abstracted from the contexts in which they made sense, and so, they are "neutral data".[106] Thus, in Dreyfus' thinking, the machine-human interaction can't, in any sense, be interpersonal. The personhood of computers doesn't even come into consideration here.

By contrast, the personhood of humans as end users *does* raise questions. Let me recall Ong's statement that tools of communication restructure consciousness. In the case of computers, this restructuring starts with the users' perception of space and time.[107] Dreyfus is, however, interested in something else: how does the computer shape human behaviour?

From Archer's point of view, the question is whether personhood is a privilege of humans or it can be attributed to other beings, too (including beings with inorganic bodies). She asserts[108] that belonging to the human species is, in itself, a necessary condition of selfhood, not of personhood; nonetheless, every human being is to be "treated as possessing worth and dignity" because of their "being made in the image of God".[109] At the same time, she allows for the personhood of other (e.g. inorganic) kinds of beings: "... personhood is not in principle confined to those with a human body".[110] Her goal with this is to avoid anthropocentricism and speciesism.

As I have mentioned above, concerning the relationship between computers (more precisely, the internet) and humans, Dreyfus' question is how this relationship transforms human behaviour: "What if the Net became central in our lives?"[111] He sees the very humanity of human beings at stake here, for two reasons. 1. The "promise of the Net is that each of us will be able to transcend the limits imposed on us by our body";[112] our Platonic philosophical tradition praises the soul dispensing the body (and, according to Nietzsche's famous dictum, Christianity is Platonism for the people). Thus, he reformulates the question the following way: "Is the body just a remnant of our descent from the animal... or does the body play a crucial role even in our spiritual and intellectual life"?[113] The answer to the question formulated as

[102] Cf. Turing 1950: 447ff.
[103] In the history of philosophy, this ability has been attributed either to God alone ('uno ictu mentis', Boëthius, *de consolatione philosophiae,* 5.4.33; 'uno obtutu', Kant: *De mundi sensibilis...,* I. §1, n 2.), or humans as well ('uno obtutu', Leibniz*, De totae cogitabilium varietatis uno obtutu complexione*).
[104] Landgrebe and Smith (2022: x) emphasise that Dreyfus' considerations about AI are inspired by Heideggerian thinking. Still, his conclusions are the same as theirs, although their arguments are grounded "on the mathematical implications of the theory of complex systems".
[105] Cf. Dreyfus 1967.
[106] Cf. Dreyfus 1992: 281. "[i]nformation must not be confused with meaning", formulated the same phenomenon Weaver 1949: 99.
[107] Bolter 1984: chs. 6-7.
[108] Archer 2011: 283.
[109] To be precise, she writes about every person's being made in the image of God. She also states that dignity is not a feature of *personhood.* It is a relational category; dignity is 'conferred' on somebody by others (Archer 2019: 23). With this view, Archer differs from e.g. Spaemann (1996), who regards belonging to the human species as a necessary and sufficient condition of personhood.
[110] Archer 2019: 28.
[111] Dreyfus 2008: 6.
[112] Ibid.: 4.
[113] Ibid.: 6.

https://doi.org/10.33621/jdsr.v7i341920

either/or seems evident. In this case, the greatest promise of the Net is at the same time its greatest danger (and *das Rettende* doesn't appear on the horizon). If it is true that the Cartesian (or Platonic) dualism is false; if it is true that computers deprive users of their bodies (or at least body awareness, creating the illusion that we are no longer tied to a definite place or time); and if to be possibly regarded as a person, someone must 'have' a body identifiable in space, then the conclusion must be that computers deprive their users of their personhoods. Or put in another way: "What is most seductive about the virtual world, ... the promise of freedom from finitude".[114] All this holds if and only if users can't or won't make the necessary and evident distinction between the real and virtual world (supposing that our world believed to be real is not a simulation). However, a crucial virtue of the virtual world is just the seductive power with which it invites users to immerse in it.[115]

From this perspective, the answer to how the internet affects our personhood is simple. If the possibility of human experience is inseparable from the body, and if access to reality is possible only through everyday practice, telepresence is absence. As Dreyfus asserts, "telepresence, both of objects and people, is parasitical on a robust sense of the presence of the real correlative with the body's set to cope with things and people".[116] If this is so, immersion into virtual reality – if (and insofar as) it becomes everyday practice and the user loses their ability to tell the virtual from the real – threatens to annihilate the possibility of conceptual thinking (including the decision about relevance) as well as the capability of commitment and making value judgements.

For Dreyfus, a vexing question concerning the internet is whether the "World Wide Web is improving or diminishing the quality of our lives".[117] In light of what was earlier said, his answer seems simple. Given the importance of embodied (and embedded) practices and the disembodied nature of the internet, quality of life can't but diminish. This answer presupposes the existential viewpoint, according to which life is worth living if and only if it has meaning. For our life to have sense, we must be committed to something while being aware of our finitude and vulnerability. Thus, if "one is already committed to a real-world cause, the World Wide Web can increase one's power to act".[118] Therefore, primarily those who are exposed to a danger – namely, that they become unable to be committed to a cause and attached to a real community –, are they who became addicted to the internet's virtual world before being committed to a real-world cause. Inversely, those who became addicted to the virtual world before having committed themselves to a real-world cause are especially exposed to the danger that they become unable to find attachment to a real community. In this respect, Dreyfus regards Kierkegaard as his predecessor for whom "the public sphere itself [was] .... a new and dangerous cultural phenomenon". (Or from another perspective: the danger is not that the public space of a virtual community fosters the 'tyranny of public opinion' feared by Mill or Tocqueville).[119] In Kierkegaard's view, the press produces nihilism ('anything goes'); it is the source of levelling, ultimately due to the Enlightenment's idea of the detached observer. The public sphere, which from one perspective could seem 'the triumph of democratisation' (since everyone can develop an opinion about just anything), from the other, was "destined to become a detached world in which everyone had an opinion about and commented on all public matters without needing any first-hand experience and without having or wanting any responsibility".[120] While McLuhan believed (or perhaps *chose* to believe) that the electronic global village could provide a remedy for barbaric tribalism,

[114] Ibid.: 105.
[115] Dreyfus' conception of AI is inseparably linked with his old-fashioned cultural criticism, as is evident from his references to Pascal concerning *diversion* (Dreyfus 2008: 97). Immersion in a virtual world (like that of *Second Life*) is a kind of *divertissement pascalien*. Heim (1993: 154) identifies immersion as one key factor of virtual reality. ("...the illusion is immersion", ibid.: 112). Bolter writes in the same vein about "immersion in popular film, television, and fiction" and its rejection (Bolter 2019: 93f., 112, 117).
[116] Dreyfus 2008: 123.
[117] Ibid.: 136. Note that posing the question in this way implies the availability of a standard with the help of which one can give an answer which is not to be doubted. (In the earlier chapters, Dreyfus analyses the prospects of search engines or the then-new phenomenon of distance learning. His discussions seem to retain only some historical relevance today.)
[118] Ibid.: 137.
[119] Ibid.: 74
[120] Ibid.: 75.

https://doi.org/10.33621/jdsr.v7i341920

according to Dreyfus, the promising prospect of "a worldwide electronic agora precisely misses the Kierkegaardian point that the people talking to each other in the Athenian agora were members of a direct democracy who were directly affected by the issues they were discussing, and, most importantly, the point of the discussion was for them to *take the responsibility and risk of voting publicly* on the questions they were debating. For Kierkegaard, a worldwide electronic agora is an oxymoron".[121] The public sphere provided by virtual space (cyberspace) is thus *a virtual community* offering the possibility of immersion without engagement or commitment. The contrast of anonymous immersion and commitment recalls Kierkegaard's distinction between the ethical and the aesthetic way of life (*Either/Or*), which since then has become a cornerstone for critics of the Enlightenment (a product of which is the public sphere; the third, the religious way of life, is beyond the scope of the phenomena discussed here). MacIntyre characterises[122] the aesthetic way of life as that of masks (the individual is not able to be committed to somebody or something); the life of those choosing the first way of life lacks unity (it "[is] dissolved into a series of separate present moments"). In contrast, "in the ethical life the commitments and responsibilities to the future springing from past episodes in which obligations were conceived and debts assumed unite the present to past and to future in such a way as to make of a human life a unity". The paradigm of the aesthetic expression is the romantic lover, while that of the ethical is the marriage.[123]

Dreyfus applies Kierkegaardian opposition to highlight the difference between the virtual and the real world. While the internet is a medium which favours the aesthetic way of life, the prerequisite of living an ethical way of life is the *small community* in the *real world* because such a milieu enables and promotes responsible communication. He formulates his own either/or: either "disembodied nihilism" or "embodied meaningful [individual] differences".[124] He suggests that we bear an ethical responsibility to choose between the two and that the choice depends on someone's worldview: the Enlightenment shows affinity with the aesthetic, tradition with the ethical way of life imbued with religiosity: "If we remain the kind of beings that Kierkegaard understood us to be, we will despair if all meaningful distinctions are levelled, and since Judeo-Christian meaningful distinctions require commitment and vulnerability, which require our embodied finitude...".[125]

Unlike Dreyfus, Archer does take into consideration the possibility of AI personhood. She can do so because she attempts to avoid anthropocentricism and speciesism, insofar as she doesn't regard belonging to the human species as a necessary condition of personhood. For her, the emergence of personhood from selfhood results from a morphogenetic process in which she attaches central importance to "two emergent capacities ... our *reflexivity* and our *concerns*". Archer explicitly states that these capacities are "dual conditions for personhood".[126] Therefore, if she can show that these features can be attributed to non-human agents as well, then, theoretically, nothing prevents them from the possibility of being regarded as persons.

Archer enumerates[127] three regular objections to the possibility of AI personhood.[128] The first barrier is that of normativity, the robots' alleged lack of ability to tell right from wrong. The counterargument is very simple and plausible. Alluding to MacIntyre's questions (*Whose justice? Which morality?*), she asks: Whose and which morality should be programmed into robots? These questions are relevant for her because of the crisis of normativity experienced in our everyday life: "With the shift from Law to Bureaucratic Regulation, the social need for shared normativity diminishes". The first barrier is closely related to the second, the emotional one: AI robots lack emotions or feelings. Archer considers this

[121] Ibid.: 138f.
[122] MacIntyre 2007 (1989): 242.
[123] Ibid.: 40.
[124] Dreyfus 2008: 123.
[125] Ibid. The choice between the two ways of life, according to MacIntyre (2007 [1989]: 40) doesn't amount to a choice between good and evil; "it is the choice whether or not to choose in terms of good and evil".
[126] Archer 2019: 16, 23.
[127] Archer 2020: 17.
[128] These objections are significant also because of what they betray about the robophobiacs' concept of a person.

https://doi.org/10.33621/jdsr.v7i341920

objection irrelevant, referring to the central importance of *concerns*. In this way, she intends to avoid, besides anthropocentricism,[129] emotivism.[130] She emphasises that there are "matters we human beings cannot help but care about" (e.g. preventing imminent danger); the emotions are "commentaries upon our concerns in the three orders of natural reality ...,"[131] and as such, they are not essential for our personhood. (Moreover, their goals may be completely unethical.[132]) The alleged third barrier is of little importance: it concerns "the absence of Qualia, a 'subjective feel'."[133] She reminds us that it is also a barrier between humans.[134] If this barrier can be set aside by learning the rules of language use (cf. Wittgenstein's private language argument), for what kinds of beings would this be a more appropriate task than the robots?

If AI robots can fulfil the criteria of personhood, and the barriers separating them from humans are not unsurmountable, then friendship between humans and robots is, perhaps, not impossible (we can "join in friendship. ... with a non-organic body but not with an anonymous human subject on a life-support machine").[135]

7. Conclusion

In this paper, I have attempted to compare Hubert Dreyfus' and Margaret Archer's views of personhood in Turing's universe. I have argued that a comparison between the two thinkers is possible because they both are 'realist' thinkers, and the comparison is also fruitful because they represent markedly different positions concerning the possibility of AI. For Dreyfus, the question is how the use of computers affects *us* – since the question of AI personhood is irrelevant to him. Archer targets two main questions: Are we justified in ascribing the possibility of personhood to robots? (can robots be regarded as agents?) and: can humans and non-humans together form a society based on the principle of solidarity?

Before directly comparing them, I hinted at how the development of technology questions the traditional definition of the human being as a rational animal since the very idea of AI challenges the view that personhood can be ascribed to the human species only. As a next step, I sketched the current scenarios of cohabitation between human and non-human agents, from that of civil war to that of peaceful and mutually profitable symbiosis. I pointed out that the emergence of AI and robots inaugurates the end of anthropocentricism: human beings must be ready to abandon the idea of their privileged status in the cosmos for the sake of survival. Finally, the disclosing of some aspects of the various kinds of "humanism" and sketching the "standard concept" of a person made up the theoretical background for giving an overview of the most important thoughts of the two thinkers. I pointed out that despite some common features of their thinking, they represent quite different standpoints concerning AI. Perhaps further investigation could shed light on the causes of this profound difference, which may have roots in their differing worldviews (given the similarities in their philosophical outlook).

My second step was to highlight their shared common theoretical background. Since they both are social realist thinkers, it does make sense to compare their views concerning AI. They represent different versions of 'realism' which leads them to different evaluations of the role of AI and robotics. As a third step, I tried to show that our understanding of the possibilities of human-machine cohabitation largely depends on whether we take an anthropocentric or a non-anthropocentric view of the world.

[129] In this context, anthropocentricism amounts to "conflating worth with being", i.e. our concern with our subjective judgement.
[130] Archer 2004: 328. Archer's argument here is very similar to that explained by MacIntyre. In his analysis (2007 [1989]: 23f.), emotivism "entails the obliteration of any genuine distinction between manipulative and non-manipulative social relations" making it impossible to appeal to impersonal criteria in a moral debate. This is the cause of the "interminable and unsettlable character of so much contemporary moral debate" (ibid.: 226).
[131] Archer 2021:180.
[132] Archer 2000: 225.
[133] Archer 2020: 17.
[134] "I can only believe that someone else is in pain, but I know it if I am"; "Another person can't have my pains", so Wittgenstein in *Philosophical Investigations* (§253, §302), before asking the question: "In what sense are my sensations private?" (§ 246).
[135] Cf. Archer 2019: 27.

https://doi.org/10.33621/jdsr.v7i341920

References

Adorno, T. W. (1954). How to Look at Television. *The Quarterly of Film Radio and Television*, 8(3), 213-235. https://doi.org/10.2307/1209731

Archer, M. S. (1995). *Realist social theory: the morphogenetic approach*. Cambridge University Press. https://doi.org/10.1017/CBO9780511557675

Archer, M. S. (2000). *Being human the problem of agency*. Cambridge University Press. https://doi.org/10.1017/CBO9780511488733

Archer, M. S. (2004). Emotions as commentaries on human concerns. In J. H. Turner (Ed.), *Theory and Research on Human Emotions* (Vol. 21, 327-356). Emerald Group Publishing Limited. https://doi.org/10.1016/S0882-6145(04)21013-X

Archer, M. S. (2011). 'Caritas in Veritate' and Social Love. *International Journal of Public Theology*, 5(3), 273-295. https://doi.org/https://doi.org/10.1163/156973211X581542

Archer, M. S. (2015). The Relational Subject and the person: self, agent, and actor. In M. S. Archer & P. Donati (Eds.), *The Relational Subject,* 85-122. Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9781316226780.004

Archer, M. S. (2019). Bodies, persons and human enhancement. Why these distinctions matter. In I. Al-Amoudi & J. Morgan (Eds.), *Realist responses to post-human society: ex machina,* 10-32. Routledge. https://doi.org/10.4324/9781351233705

Archer, M. S. (2021). Friendship Between Human Beings and AI Robots? In J. von Braun, M. S. Archer, G. M. Reichberg, & M. Sánchez Sorondo (Eds.), *Robotics, AI, and Humanity: Science, Ethics, and Policy,* 177-189. Springer. https://doi.org/10.1007/978-3-030-54173-6_15

Archer, M. S. (2023). Can humans and AI robots be friends? In M. Carrigan & D. V. Porpora (Eds.), *Post-Human Futures Human Enhancement, Artificial Intelligence and Social Theory,* 132-152. Taylor & Francis. https://doi.org/10.4324/9781351189958-7

Archer, M. S., & Donati, P. (2015). The Plural Subject versus the Relational Subject. In M. S. Archer & P. Donati, *The Relational Subject,* 33-76. Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9781316226780.002

Archer, M. S., & Morgan, J. (2020). Contributions to realist social theory: an interview with Margaret S. Archer. *Journal of Critical Realism*, 19(2), 179-200. https://doi.org/10.1080/14767430.2020.1732760

Arendt, H. (1978). *The life of the mind* (One-vol. ed). Harcourt Brace Jovanovich.

Arendt, H. (1998). *The human condition* (2nd ed.). University of Chicago Press.

Balkin, J. (2017). *The Three Laws of Robotics in the Age of Big Data. Information Technology & Systems eJournal.* https://openyls.law.yale.edu/bitstream/handle/20.500.13051/4697/78_Ohio_St._L.J._1217__2017_.pdf?sequence=2&isAllowed=y

Bolter, J. D. (1984). *Turing's man: western culture in the computer age*. University of North Carolina.

Bostrom, N. (2019). The Vulnerable World Hypothesis. *Global Policy*, 10(4), 455-476. https://doi.org/https://doi.org/10.1111/1758-5899.12718

Burnside, J. (2011). *God, justice and society: aspects of law and legality in the Bible*. Oxford University Press.

Chalmers, D. (2022). *Reality+: virtual worlds and the problems of philosophy*. W. W. Norton & Company.

Coeckelbergh, M. (2015). The tragedy of the master: automation, vulnerability, and distance. *Ethics and Information Technology*, 17(3), 219-229. https://doi.org/10.1007/s10676-015-9377-6

Dewey, J. (2018). *Democracy and education: an introduction to the philosophy of education*. Myers Education Press.

Dinello, D. (2016). The Borg as Contagious Collectivist Techno-Totalitarian Transhumanists. In K. S. Decker. & J. T. Eberl (Eds.), *The Ultimate Star Trek and Philosophy,* 83-94. https://doi.org/https://doi.org/10.1002/9781119146032.ch8

Donati, P. (2019). Transcending Human: Why, where, and how? In I. Al-Amoudi & J. Morgan (Eds.), *Realist responses to post-human society: ex machina,* 53-81. Routledge.

Dreyfus, H. L. (1967). Why Computers Must Have Bodies in Order to Be Intelligent. *The Review of Metaphysics*, 21(1), 13-32. http://www.jstor.org/stable/20124494

Dreyfus, H. L. (1992). *What computers still can't do: a critique of artificial reason*. MIT Press.

Dreyfus, H. L. (2008). *On the internet* (2nd ed.). Routledge.

Dreyfus, H. L., & Taylor, C. (2015). *Retrieving realism*. Harvard University Press.

Frank, M. (2022). In Defence of Pre-Reflective Self-Consciousness: The Heidelberg View. *Review of Philosophy and Psychology*, 13(2), 277-293. https://doi.org/10.1007/s13164-022-00619-z

Frankfurt, H. G. (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy*, 68(1), 5-20. https://doi.org/10.2307/2024717

Frankfurt, H. G. (1988). Identification and wholeheartedness. In H. G. Frankfurt (Ed.), *The Importance of What We Care About: Philosophical Essays,* 159-176. Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9780511818172.013

Freud, S. (1917). *Eine Schwierigkeit der Psychoanalyse*. Imago(V), 1-7. https://doi.org/https://doi.org/10.11588/diglit.25679.1

Fuchs C. (2022). *Digital capitalism*. Routledge.

Gabriel, M. (2021). Could a Robot Be Conscious? Some Lessons from Philosophy. In J. von Braun, M. S. Archer, G. M. Reichberg, & M. Sánchez Sorondo (Eds.), *Robotics, AI, and Humanity: Science, Ethics, and Policy,* 57-68. Springer International Publishing. https://doi.org/10.1007/978-3-030-54173-6_5

Haraway, D. J. (2004). A Manifesto for Cyborgs: Science, Technology, and Socialist Feminism in the 1980s. In *The Haraway Reader,* 7-45. Routledge.
Hegel, G. W. F. (2018). *The phenomenology of spirit* (T. P. Pinkard, Trans.). Cambridge University Press.
Heim, M. (1993). *The metaphysics of virtual reality*. Oxford University Press.
Heim, M. (1998). *Virtual realism*. Oxford University Press.
Jaki, S. L. [1969]. *Brain, mind and computers*. Herder & Herder.
Jonas, H. (1984) *The imperative of responsibility: in search of ethics for the technological age.* Chicago University Press.
Jorion, P. (2022). *Humanism and its discontents: the rise of transhumanism and posthumanism.* Palgrave Macmillan Cham. https://doi.org/10.1007/978-3-030-67004-7
Kantorowicz, E. H. (2016). *The King's two bodies: a study in mediaeval political theology.* Princeton University Press. https://doi.org/10.2307/j.ctvcszz1c
Landgrebe, J., & Smith, B. (2022). Why Machines Will Never Rule the World: Artificial Intelligence without Fear (1st ed.). Routledge. https://doi.org/10.4324/9781003310105
MacIntyre, A. C. (2007). *After virtue: a study in moral theory* (3rd ed.). University of Notre Dame Press.
Masco, J. (2019). Ubiquitous Surveillance. In C. Besteman & H. Gusterson (Eds.), *Life by Algorithms: How Roboprocesses Are Remaking Our World,* 125-144. University of Chicago Press. https://doi.org/doi:10.7208/9780226627731-008
Nagel, C. (1998). Intersubjectivity and the Internet. In A.-T. Tymieniecka (Ed.), *Ontopoietic Expansion in Human Self-Interpretation-in-Existence: The I and the Other in their Creative Spacing of the Societal Circuits of Life Phenomenology of Life and the Human Creative Condition,* Book III, 179-197). Springer. https://doi.org/10.1007/978-94-011-5800-8_11
Nyíri, K. (2016). Conservatism and Common-Sense Realism. *The Monist*, 99(4), 441-456. https://www.jstor.org/stable/26370770
Pellegrino, G. (2008). Convergence and Saturation. Ecologies of Artefacts in Mobile and Ubiquitous Interaction. In J. K. Nyíri (Ed.), *Integration and ubiquity: towards a philosophy of telecommunications convergence,* 75-82. Passagen.
Rubin, C. T. (2014). *Eclipse of Man: Human Extinction and the Meaning of Progress*. Encounter Books.
Rushkoff, D. (2010). *Program or be programmed: ten commands for a digital age*. OR Books.
Rushkoff, D. (2014). *Present shock: when everything happens now*. Current.
Schuetz, A. (1945). On Multiple Realities. *Philosophy and Phenomenological Research*, *5*(4), 533-576. https://doi.org/10.2307/2102818
Shields, R. (2003). *The virtual.* Routledge.
Sorgner, S. L. (2023). *We have always been cyborgs: digital data, gene technologies, and an ethics of transhumanism*. Bristol University Press.
Spaemann, R. (1996). Is Every Human Being a Person? *The Thomist: A Speculative Quarterly Review*, 60 no. 3, 463-474. https://doi.org/https://doi.org/10.1353/tho.1996.0013
Strawson, P. F. (1959). *Individuals: an essay in descriptive metaphysics.* Routledge.
Taylor, C. (1975). *Hegel.* Cambridge University Press.
Turing, A. M. (1950). *Computing machinery and intelligence.* Mind, 59(October), 433-460. https://doi.org/10.1093/mind/LIX.236.433
Turkle, S. (2005). *The second self: computers and the human spirit* (20th-anniversary ed.). MIT Press.
Weaver, W. (1949). Recent contributions to the mathematical theory of communication. In C. E. Shannon & W. Weaver, The mathematical theory of communication, 93-117. The University of Illinois Press.

