

# A Human-Centred Approach to Symbiotic AI: Questioning the Ethical and Conceptual Foundation

Antonio Carnevale <sup>a</sup>, Antonio Lombardi <sup>a</sup> and Francesca A. Lisi <sup>b,1\*</sup>

<sup>a</sup> *Department of Humanistic Research and Innovation, University of Bari Aldo Moro, Palazzo Ateneo, Piazza Umberto I, 70121, Bari, Italy*

<sup>b</sup> *Department of Informatics, University of Bari Aldo Moro, Via E. Orabona 4, 70125 Bari, Italy*

**Abstract.** This paper advocates for a constructivist approach to symbiosis to restore human-centredness in the governance of Symbiotic Artificial Intelligence (SAI). Challenging rigid, deterministic foundational methods warns against the risk of divorcing ethics from mere adherence to moral principles. Instead, it calls for a shift towards a distributed, contextual, relational, and dialectical structure to embody human-centredness. Through an analysis of the SAI landscape and its interplay between social and technological factors, the paper argues for a reconceptualisation of theoretical foundation and human responsibility within the socio-technical perspective. Chapter 2 delves into foundational issues of SAI, questioning the application of biological categories and proposing patterns of SAI based on definitions of intelligent life. Chapter 3 explores the potential of a constructivist approach, emphasising flexibility and context awareness, and presents a framework for understanding and evaluating SAI systems, components of an evolving methodology.

Keywords: Symbiotic Artificial Intelligence (SAI), Human-Centric AI, AI Ethics, Machine Ethics, Philosophy of Technology

## 1. Introduction

The European Commission (EC) has long advocated the certainty of a *human-centric approach* to Artificial Intelligence (AI), emphasising communication between societal stakeholders and technology developers to establish a tangible foundation of trust in AI. This approach found notable expression in the “Ethics Guidelines for Trustworthy AI” [30] promoted by the EC High-level Expert Group on Artificial Intelligence (hereafter referred to as HLEGAI). Explicit references in this regard are also present in other European documents. Notably, as early as April 2018, the Commission had outlined the cornerstones of its strategic direction by emphasising the centrality of humans in AI development. The Commission’s communication to the European Parliament, the Council, the European Economic and Social Committee, and the Committee of the Regions, titled

“Building Trust in Human-Centric Artificial Intelligence” [20], deserves mention. These communications have been incorporated not only in the guidelines but also in the “White Paper on AI” [21] and the proposed “Artificial Intelligence Act” [22].

Collectively, these documents underscore the European Union’s (EU) political agenda not only to coordinate an integrated approach for maximising AI opportunities and addressing associated challenges but, more importantly, to position itself as a global leader in the development and deployment of AI that *normatively* aligns with principles of sustainable technological innovation and respect for fundamental rights (already started with the GDPR [19], the so-called “Brussels effect”, see: [59]).

This approach underscores several pivotal strengths, notably encapsulated in the HLEGAI’s endeavour to leverage a dual normative function of fundamental rights. Indeed, these rights encompass both the legal

---

<sup>1</sup>\*Corresponding author. E-mail: francesca.lisi@uniba.it.

protections enshrined in the constitutional frameworks of nations (actual legal rights) and the inherent rights of individuals grounded in the intrinsic moral status of human beings (ethical values that may not be legally binding but are pivotal for ensuring the trustworthiness of AI on a global scale).

Another salient strength is recognising that every normative orientation constitutes a structured, albeit not always systematic, array of moral assertions interlinked in diverse manners. In essence, all these assertions can be traced back to a primary moral principle (monistic ethics) or can stem from multiple principles (pluralistic ethics) [57]. The EU's ethical framework for a "good AI society" [25], particularly as formulated within the HLEGAI, adopts a pluralistic ethical standpoint contingent upon interrelated principles for the betterment of society.

In this paper, we discuss the challenges posed by applying a human-centric normative orientation to *Symbiotic AI* (SAI) or human-AI symbiosis [28] [39] [51]. Human-computer symbiosis is a relatively long-time metaphor [29], initially proposed by J.C.R. Licklider in his historical 1960 article [39]. However, his vision describes a symbiotic relationship between humans and computers, where humans initiate goals and hypotheses, and computers process data to test these hypotheses, with the anticipation of a networked system for communication. Despite the initial vision from sixty years ago, contemporary AI systems now surpass this model by discerning patterns in large datasets, making decisions, and even creating the data themselves, suggesting a transition towards "technosymbiosis" [29] or what we call symbiosis as collaborative teaming framework. SAI today aims to enhance human-machine collaboration in various domains, such as healthcare and manufacturing, where humans and AI systems collaborate for improved outcomes [32]. Machines are becoming agents of technology's "in-betweenness" [24], and they can engineeringly (and not cognitively) generate informational meaning. Something dissimilar to human-like understanding and inferring human-like responsibility [26], but precisely because of this, it represents a new ethical challenge that highlights the *sociotechnical interdependence* between humans (and nonhumans) and computing devices, noting both benefits and risks. This is why fostering a human-AI symbiosis requires today's design of AI systems according to a *human-centred approach*.

Against this backdrop, research at the University of Bari, conducted within the NRPP-funded Future AI Research (FAIR) project, focuses on designing new

interaction paradigms to amplify human performance while ensuring system reliability, safety, and trustworthiness. Acceptability of SAI systems, particularly concerning value alignment between AI and humans, is a crucial focus of research within FAIR. This involves interdisciplinary studies to address issues such as privacy policies, security, and fundamental freedoms. Philosophical perspectives are also explored to understand the epistemological and ethical aspects of SAI.

In this paper—continuing the work presented at the workshop BEWARE@AIxIA 2023 [6]—we contribute to the understanding of human-centred symbiosis with AI by introducing a methodology for assessing the impact of SAI systems using a human-centred approach.

In Chapter 2, we address some foundational and philosophical challenges of SAI by analysing first the relationship between life and artefact (Section 2.1) and then the one between intelligence and symbiosis (Section 2.2).

Chapter 3 will explicitly address the potential of a constructivist approach in redefining the foundation of SAI and its alignment with human-centred ethics (Section 3.1), advocating for flexibility and context awareness. We will propose a constructivist framework for understanding (Section 3.2) and evaluating (Section 3.3) SAI systems, delineating the outline of an evolving methodology on which we are working in the project FAIR – constructivism is a social sciences theoretical perspective asserting that knowledge and meaning are constructed through social interactions and experiences, emphasising the role of culture and context in shaping human understanding [2] [6] [29] [36] [43] [53].

In Chapter 4, we conclude the paper with final remarks and outline directions for future research endeavours. A last clarification before proceeding: We use the terms 'human-centric' and 'human-centred' synonymously, referring both to design and research approaches that prioritise human needs, behaviours, and experiences as the central focus, ensuring that the resulting systems or solutions are optimally aligned with user requirements and usability principles.

## 2. Foundational and Philosophical Challenges of SAI

The concept of 'Symbiotic Artificial Intelligence' introduces an additional layer of complexity to the already problematic notion of 'Artificial Intelligence.' Incorporating the adjective 'symbiotic' further

complicates the scenario yet represents one of the newest boundaries in AI research. The term ‘symbiosis’ inherently suggests the coexistence of two or more entities that collaborate or, at the very least, derive mutual benefit from their association. Thus, a genuinely Symbiotic AI would foster an interaction between humans and machines that transcends the traditional dynamic of controller and control, evolving into a reciprocal relationship between two agents potentially equal in decision-making, capable of mutually influencing one another. These agents, though not identical, are both intelligent, with the former ‘controller’ – the human – entering a partnership where the machine may assume control by adjusting, amending, or directly intervening in decisions in real-time towards a shared objective, leveraging a vast and precise understanding of data and procedures necessary for the task. It’s important to acknowledge that despite the potential asymmetry in this interaction between intelligent entities, a degree of control always remains, exercised mutually, ensuring the symbiotic relationship’s continuity.

A visionary example could be Elon Musk’s Neuralink. Let’s imagine that one day, some of us, for both therapeutic and creative purposes, could equip ourselves with brain implants designed to allow us to use our neural signals to control external technologies. Naturally, the AI guiding such implants would be fitted with an *ethics by-design* that – at the very least – requires it to adhere to Asimov’s laws. However, it is also true that, despite the significant imbalance in the symbiotic relationship between humans and the implant, the AI would also have its capacity for increasing self-learning. Paradoxically, this tendency to autonomy would grow by feeding on our neural processes, thereby evidently limiting our absolute control over the machine and, consequently, over ourselves. This model of human-machine interaction represents an extreme and radical variant of what Floridi defines as a “third-order technology” [24]. In the Italian philosopher’s idea, a third-order technology removes or replaces humanity from the loop, eliminating the H2M interface in favour of M2M protocols. Delving into nanotechnologies and biotechnologies, AI is not simply reengineering the world with its protocols, but re-ontologizing it. Through the re-ontologising of our society, the digital realm is also redefining, from an epistemological standpoint, the rationality underlying our society, namely many of our established conceptions and ideas [26].

Against this backdrop, envisioning SAI raises questions about the nature of the technology required and whether current forms of human-machine interaction hint at a future symbiotic relationship. Kai-Fu Lee and Chen Qiufan [37] speculate about a near future where smartphone applications, through real-time data on our actions and behaviours, could suggest changes to our decisions or behaviours based on an intricate network of app interactions. Is this merely a matter of data availability and technological capability, or is something more profound at play in achieving a symbiosis between natural and artificial intelligence? To consider AI in this context is to imagine machines or software capable of not just ‘live’ learning and response, as seen with the IoT, but also of discerning behavioural regularities, intervening in real-time with a specific ‘authority’ to alter human conducts that would otherwise occur differently. Symbiosis, in this sense, implies a learning-driven agents’ relationship where both entities coexist and learn from one another: the machine from observing human actions and the human from insights provided by the machine based on its observations. In a recent project attempting to use symbiosis between machine and human beings to build AI systems capable of understanding the nuances of the real world, precisely this point is emphasised: “While the computer’s role has shifted from being a passive ‘executor’ to an active learner, the role of the human is still that of an actively involved teacher, because it is the humans that create clean, consumable data for the computer to analyse and hence train itself. However, what if there was a way in which the computer’s role remains that of an active learner, but we humans can passively sit back and go on with our lives while the computer learns from our actions?” [52, p. 4].

The shift here is from instructing machines to learning from them, not just acquiring new information but receiving guidance on enhancing our actions to achieve desired outcomes. This form of HCI is widely discussed in scientific literature, but the accuracy and implications of applying ‘symbiosis’ to AI warrant further examination and pose foundational philosophical questions. This section addresses these questions, exploring the definition and feasibility of SAI and its compatibility with a human-centric approach.

## 2.1. *Life + Artefact: How is Symbiosis Possible for Machines?*

At the heart of considering Symbiotic Artificial Intelligence (SAI) from a foundational perspective is the interplay between organic life and manufactured artefacts, traditionally viewed as diametric opposites. The essence of a living organism differs markedly from that of an artificial construct, yet in the context of SAI, these distinct entities are envisioned to coalesce. The historical backdrop for differentiating natural beings from artificial ones dates back to Aristotle's *Physics* and extends through Darwin's delineation between organic and inorganic realms [11], suggesting an initial clear division between two distinct realms of existence that, within the SAI acronym, are brought into conjunction.

According to this traditional paradigm of the life sciences that goes from Aristotle up to modern biology, an artefact or a machine (an entity that is a product of *techne*, according to Aristotle's own definition) is not something alive. What does it mean to be alive for Aristotle? In *De anima* Aristotle says that to live means to have the capacity to nourish, grow and decay [1, p. 22]. These functions are all united under the sign of movement, specifically spontaneous movement. For Aristotle and then throughout the long course of Western biology, what distinguishes a living thing from an artefact is the ability to move spontaneously. The machine is not endowed with life, it is not 'animated' (ensouled) because it does not possess an autonomous principle of movement. We move the machine, whereas a living thing, an insect, or an animal, such as plants themselves (which have a nonlocal motion but a motion of growth), move autonomously. We do not drive them.

The modern era, especially with the rise of the mechanistic model, will reject such a view of life by coming to conceive even living beings as machines whose movement is always the result of a complex chain of external impacts (think of Descartes, Hobbes, and Spinoza). But a revival of Aristotelian conceptuality will occur in the nineteenth century, with *Naturphilosophie* and the emergence of an 'organicistic' paradigm, which gave birth to modern biology.

This conception, whereby a living being is one that possesses an autonomous principle of movement, a self-finality (*entelechy*), has been somewhat challenged by the emergence of robotics and computer science in the twentieth century. These sciences study the ways in which it is possible to build machines that

can attain a certain level of autonomy. Obviously, it is not the autonomy that pertains to living beings in general and human beings less than ever. However, all artificial intelligence systems, especially Large Language Models (LLMs), have shown self-development capacity. Robotics and computer science constitute an exception to the old distinction between *physis* and *techne*, between life and artefact. Robots and instances of artificial intelligence are a special kind of machine: they confront us with 'mechanical' processes that nevertheless exhibit characteristics that somehow have to do with life and, thus, potentially also with symbiosis. Maturana and Varela proposed in the twentieth century the famous distinction between autopoietic and allopoietic machines: the former machines would be the living organisms, that is, machines whose process results in the preservation, reproduction, and development of the machines themselves, while allopoietic machines are machines that are operated and have something other than the machine itself because of their process [43].

One might ask: What distinction would remain if an artefact (a device or technology devoid of organic components) shared all characteristics typical to organisms composed of organic material? Would they both be deemed living entities? [44, p. 55; p. 56]. Is it possible for an artefact to transform into an autopoietic machine? This avenue has been explored since the mid-20th century (for instance, by Von Neumann with his self-replicating machines) and later during the 1980s in the field of artificial life (A-life). A-Life operates on the functionalist premise that life can also be digitally synthesised, meaning the composition of the substrate, whether atomic or digital (bits), is irrelevant. What matters is that this substrate demonstrates specific interrelations and characteristics, such as self-preservation, self-reproduction, and autonomous movement for its own sake, among others. Numerous examples exist (Conway's Game of Life, Polyworld, RepRap, Slugbot, etc.). These advancements appear to blur the traditional line between life and artefact, a matter of great significance for the question of founding SAI. In this respect, SAI seems to be an AI that potentially engages in a 'life-to-life' relationship. But accordingly, Are robots 'living' entities? Are artificial intelligence such as ChatGPT entities that, although not "material," exhibit some of the characteristics of life?

Addressing robots, Lévy [38] suggests that, according to some broadly accepted biological benchmarks (such as those proposed by Koshland [34]), it is not

entirely far-fetched to classify robots and, by extension, AI programs as life forms. De Collibus [13] expresses a contrary view regarding Large Language Models (LLMs). Several attributes typically associated with living beings are also identifiable in ChatGPT, like the ability to self-evolve and learn, a trait shared with humans, dogs, or fish. ChatGPT can capitalise on accumulated data like these organisms, enhancing its complexity and devising increasingly effective solutions over successive problem-solving attempts. However, these technologies lack a crucial aspect of life as defined by Spinoza's concept of *conatus* and Schopenhauer's *Wille zum Leben*, the instinct for self-preservation. None of these technologies operate driven by a desire to continue existing. Perhaps the self-purposefulness Aristotle initially identified as a primary feature of life is not fully present in AI: it is indeed capable of self-expansion or self-improvement, but to what end? It does not appear to do so in pursuit of its assertion or its own 'good'. This could represent an insurmountable gap between living beings and machines, posing a significant challenge to applying biological categories such as symbiosis to AI technologies.

## 2.2. *Intelligence and Symbiosis: Three Patterns* *Based on (Dis)continuity*

How do things stand with intelligent life? This question is essential for the foundation of an SAI since it is assumed that the two lives that should enter symbiosis are both intelligent lives. Indeed, in the case of the artificial symbiont, its life might coincide completely with its intelligence. SAI thus allows us to ask important questions about the relationship between life and intelligence.

The problem is tough since it depends on what kind of definition of intelligence we start with.

A spectrum of perspectives on defining intelligence, ranging from inclusive to exclusive, underpins discussions on the nature and potential of SAI. Starting with Schelling and Darwin, a minimal view emerges that intelligence essentially involves the capacity for organisation. This perspective posits that even simple biological entities, such as mushrooms or worms, demonstrate intelligence through their organised efforts to solve problems and adapt to their environment. This organisation principle extends to artificial intelligence, where software and programs exhibit a basic structure enabling them to reconfigure based on contextual inputs to meet specific goals [41,

p. 21]. From this perspective, which could be called 'homogeneous continue', an SAI is already at work. It has been for a long time [39]: our devices, through which we interact with generative AI systems, are already in some sense forms of intelligence with which we live in symbiosis. Here we are dealing with a definition of intelligence that flattens discontinuities and reduces them to a mere difference in degree: it is enough for an entity to exhibit a specific computational strategy to be considered intelligent (from viruses to humans and AI systems). At this level, a SAI is conceivable as a form of integration between human beings and digital devices [62] [35], in a relationship that provides for their increasing autonomy, seamlessness and self-development. A perspective that has begun to be investigated in the *White Papers* on Symbiotic Autonomous Systems (SAS) [49] [39] [3].

However, there are less inclusive definitions, such as those that restrict the field of intelligence to vertebrates with a minimum of brain activity [40]. In this case, calling that AI an 'intelligent' life would already become less noticeable. Floridi [26], for example, contends that the intelligence of current generative AI systems falls short of even that of a sheepdog, likening their cognitive capabilities to those of a dishwasher (see also [14], in which Dennett argues that AI is the result of "a slow, mindless process"). Accepting this premise implies that SAI cannot be based on our current interactions with technologies like smartphones or ChatGPT. Instead, envisioning SAI requires imagining a relationship between humans and a robot whose artificial intelligence parallels at least that of an animal. Under this scenario, true SAI would manifest primarily in an exosymbiotic, rather than endosymbiotic, framework, potentially side-lining wearable or prosthetic AI technologies from the SAI paradigm. However, the problem is that the realisation of 'animal' level AI remains highly problematic.

In opposition to continuism, there are various versions of intense or absolute discontinuism, whereby intelligence is a uniquely human capability. To define intelligence as any ability to solve problems or to calculate even the way animals do would be an equivocal way of talking about intelligence because proper intelligence is only our own. Indicative of 'proper' intelligence would only be capabilities such as intentionality, universalisation, creativity, spontaneity, self-consciousness, emotionality, etc.: all characteristics that are the exclusive preserve of human beings. In this case, to talk about SAI, we

would have to approach science fiction or post-humanism or very distant in time perspectives that contemplate the possibility of achieving an Artificial General Intelligence (AGI) and thus some “Singularity” [7].

### 3. Embracing Constructivism: Rethinking Symbiotic AI’s Foundation

Suppose we delve into the historical divide between organic life and artificial constructs, questioning whether machines can exhibit characteristics traditionally associated with living organisms. In that case, speaking and envisaging a foundational hypothesis of SAI takes a lot of work. While some argue for a continuum of intelligence across biological and artificial entities, others contend that true symbiosis with machines necessitates AI with capabilities rivalling at least those of animals. Accordingly, from the kinds of human-machine interaction known today, it is pretty impossible to find ‘symbiosis’ by reflecting on the nature of intelligence, the potential for symbiotic relationships between humans and machines, and the boundaries between living and non-living entities.

Based on these premises, in this part of the paper, we will argue in favour of a possible alternative pathway to the foundation of SAI. We will pursue a *constructivist* approach, a theoretical perspective asserting that knowledge and meaning are constructed through social interactions and experiences, emphasising the role of culture and context in shaping human understanding.

From a deterministic and theoretically firm perspective, we know that a constructivist approach may not appear as the most rationally justified path to speak of ‘foundation’, as nothing is constructed without first founding something. However, given the highly sociotechnical nature of SAI, there is no possibility of a foundation other than the ‘weak’ foundation of a constructivism that maintains a flexible, adaptable, cautious, and context-aware direction of thought. This position is supported not only by the foundational dilemmas and doubts we have shown in Chapter 2 but also by the argument that a constructivist approach is the only possible one if we want to reconcile an explainable justification of SAIs with ethics of AI based on the human-centred and trustworthy design of machines, as demanded by European institutions [21] [22] [30] and significant international agencies [48] [63]. To demonstrate this, in Section 3.1, we will highlight the critical points of

this discrepant opposition between the foundation of SAIs and the ethics of AI, such as human-centeredness. Having highlighted this discrepancy in Section 3.2, we will argue in favour of a constructivist approach as the only possible way to reconsider the relationship between SAI and human-centeredness in dialectical terms rather than opposition. Of this approach – which we are trying to refine in the FAIR project – in this paper, we present its methodological outline consisting of a theoretical onto-epistemological framework (3.2) and a preliminary evaluation framework (3.3) to identify the main sociotechnical characteristics of an SAI system and, from there, proceed to the subsequent steps – which we are still refining in our FAIR team – for the ethical assessment of the human-centeredness of SAI.

#### 3.1. SAI foundation vs. human-centred ethics

As previously discussed, the intricate nature of the ‘symbiosis’ concept encompasses various definitions and foundational viewpoints regarding the complex relationship between beings in general, with particular emphasis on humans and artificial agents. This relationship poses a significant challenge due to its dual investigative nature, where symbiosis is interpreted as a connection between life forms at one level and as a relationship between forms of intelligence at another [28].

However, what implications arise from such foundational variety when transitioning from speculation to ethical evaluation and normative orientation? Do we genuinely necessitate an ethics of SAIs? And if so, given the absence of a unified foundation, what type of normative orientation should we adopt?

In ethics, specifically, the normative ethics associated with AI [26], obligations are not necessarily imposed, but behavioural normative orientations are expressed. A normative orientation indicates a ‘should be’, reflecting a structured yet not always systematic framework containing justifiable and coherent moral normative statements [12]. However, it’s essential not to misconstrue the ‘should be’ as a forced adherence to a value or ideal; instead, adherence may be motivated by various factors, such as a consideration of consequences (*consequentialist ethics*), one’s character or behavioural disposition (*virtue ethics*), or a principle transcending consequential calculations (*deontological ethics*) [57]. Moreover, as observed in *eudaimonic ethics*, individuals may follow a rule to

align with a personal or collective ‘should be’, even for the sake of happiness [12].

Therefore, the challenge lies in establishing a foundation for being and determining the normative orientation to govern the diverse aspects of moral ‘should-be’. Consequently, the question arises: to what normative orientation can SAI, already complicated by excessive variability at its foundation, be traced back?

The human-centric approach is the indispensable cornerstone of normative guidance within the FAIR project framework. It forms the bedrock of the European Commission’s ethos on AI ethics [21] [22] [30] and resonates throughout various international normative doctrines [48] [63]. Yet, harmonising the intricate foundation of SAI with the imperative of AI human-centeredness presents formidable hurdles. Our inquiry reveals this reconciliation to be somewhat problematic. Upon scrutiny, a growing dichotomy emerges between these two paradigms, primarily due to irreconcilable disparities.

Beginning with SAI, the quandary lies in its excessive foundational variability, which defies attempts to embed it within the realm of human understanding. Consequently, its origins need a distinctly human-centric focus.

A second challenge arises concerning the ethical dimension of AI human-centeredness, revealing inherent vulnerabilities. Even before delving into specific technology interactions, criticism may be levelled at the human-centric approach for its latent speciesism [60]. Critics contend that human superiority is a flawed concept, advocating instead for acknowledging intrinsic value in non-human entities. The application of anthropocentrism to AI ethics thus risks being perceived as a biased global guiding principle, prompting scrutiny over its relevance in matters of international and environmental justice [8] [9] [10].

These challenges underscore the apparent discordance between an SAI needing a more explicit foundation and a notion of human-centeredness that may verge on abstraction or even AI colonialism [10]. If left unresolved, this dichotomy is poised to escalate. On the one hand, “techno symbiosis” is on the rise, albeit not exclusively in a biological context but also in a metaphorical and meta-semantic sense [29]. Conversely, substantive proposals to transcend mere principle-based definitions and actualise normative ideals into tangible, democratic social and political frameworks – such as navigating human existence in

the algorithmic age – remain conspicuously absent [27].

### 3.2. *Constructivism as an onto-epistemological framework of SAI*

Could we transcend *this* apparent tension of irreconcilability? Such an avenue seems feasible only through an approach rooted in constructivism, emerging as the foremost plausible option for reinstating a human-centred focus in the governance of SAI.

Symbiosis, in essence, does not exist. It is not merely constructed *within* interactions between humans and machines – whether these interactions occur at a biological level (prosthetics) or within the realms of imagination (the metaverse and digital twins), creativity (ChatGPT, DALL-E), or as support for determinations, judgments, and courses of action in an organisation or business (DSS). Instead, it is constructed *through* interactions between humans and machines. In other words, symbiosis lies not solely in the algorithm and its learning model (supervised, unsupervised, reinforcement learning) but in the *chain of choices, moral reasoning, and responsibility* along the sociotechnical axes that comprise an SAI system. The sociotechnical nature of symbiosis [3] [45] primarily imparts the symbolic form to this concept [29], resembling what the German sociologist and philosopher Max Weber referred to as an “ideal type” [33] [21]. An ideal type functions as an abstract and hypothetical construct, not detached from reality but used in social sciences to dissect and comprehend complex social phenomena. It captures essential characteristics while transcending strict realism, as a conceptual scaffold for understanding the intricate social networking of technosystems [33] [21].

An ideal type's comprehensive and explanatory power lies in the richness of the constructivist approach to the SAI foundation. It is not a proper foundation because, as we have said, symbiosis per se does not exist. Instead, symbiosis manifests as a set of processes rather than a brute fact in the interaction between humans and machines. Therefore, theoretical attention should shift toward understanding the dynamism of the processes through which such condensations exhibit some degree of symbiosis [8] [9].

New approaches are emerging around similar needs, challenging the link between machines and ethics understood solely in normative terms (compliance with laws and ethical principles) or engineering terms (machine ethics). The problem does not lie in the

ontological definition of the type of symbiosis with machines but in extracting rational and social value from these definitions, interrogating them both as technological devices and as metaphors of “socio-material practices” [49] or “techno-scientific practices” [53].

Finally, such a repositioning of symbiosis in the domain of possibility and dynamics would render dialectical the old opposition between conceptual foundation and ethical evaluation of human-centeredness. In doing so, the foundational excess of SAI would not be a conceptual problem but an epistemic opportunity, a potential strength that constructivism would help to understand by extracting value and insights from how it can be defined. In this regard, it is precisely, *onto-epistemic* to use the terminology used by physicist-philosopher Karen Barad to point at the inseparability of ethics, ontology and epistemology when engaging in scientific knowledge production, with social and technological practices and with the world itself and its inhabitants – human and non-human beings that intra-actively co-constitute the world [2]. Our constructivist approach reshapes this agential realist perspective that underlines the intra-activity between the world and its forms of symbiosis because we foresee here different methodological advantaging premises:

**Flexibility in definition.** Artificial symbiosis evolves alongside technological advancements and societal shifts. A socio-technical constructivist approach recognises that the definition of artificial symbiosis can be problematic and should not be fixed in stone. It acknowledges that our understanding of this symbiosis depends on current knowledge, technology, and society. This approach allows for ongoing adaptation and refinement of the definition based on empirical testing and real-world implementation scenarios, ensuring it remains relevant and adaptable to changing circumstances.

**Anthropocentric value review.** The socio-technical constructivist approach acknowledges that the anthropocentric value of artificial intelligence is not absolute and specist but rather an ideal-typing perspective rooted in the social and historical dimensions of any realism, including values. This approach facilitates a more inclusive and open discussion about the roles and regulations of human-machine symbiosis. It allows us to incorporate diverse, intersectional perspectives within a framework that does not privilege any single viewpoint but recognises the value of coexistence and coevolution between humans and machines [50].

#### **Adaptability to emerging AI-based technologies.**

As AI evolves rapidly, a socio-technical constructivist approach can accommodate the emergence of new technologies and their impact on the symbiotic relationship between humans and machines. It enables us to consider and adapt to unforeseen developments, ensuring our understanding and regulations remain relevant and practical. This flexibility is essential in an era where technology is advancing at an unprecedented pace.

#### **A cautious approach to unbridled techno-ideology.**

Contemporary understanding underscores that technological potency is anything but neutral – it assumes a role in shaping and co-determining societal structures [18] [33] [58] [65]. Thus, embarking on a constructive exploration of the conditions underpinning human-machine symbiosis facilitates a precautionary and context-aware [56] stance against unchecked techno-ideological fervour, preventing succumbing to facile exuberance.

Evaluating a socio-technical system is notably more challenging than assessing an individual technology due to numerous variables, contextual variations, scenarios, and changing user dynamics. In our ongoing research as part of the FAIR project, we are developing a methodology for the human-centred impact assessment of SAI, which comprises several steps, each executable through diverse techniques and approaches.

The initial step of this methodology entails electing socio-technical constructivism, which we have delineated as the onto-epistemological foundation [53] upon which to ground our method. Naturally, this foundation is ‘non-foundational’, as it is determined by the levels of abstraction and participation implicated in techno-scientific practices.

### *3.3. Sociotechnical features for a preliminary ethical evaluation of human-centred SAI*

How can we effectively evaluate the human-centredness of a SAI system? What criteria should we consider when identifying the level of symbiosis within the system?

Indeed, establishing an evaluation framework is the second step of the methodology we are working on. This framework scrutinises and weighs the ethical robustness concerning the design and implementation of SAI systems in various contexts. Its primary purpose is to describe the application of an SAI system deployed in a specific project and sociotechnical

context. This step allows the method to initialise and serve: (a) to map out the general socio-technical features of the SAI system under analysis; (b) to provide a set of dimensions to contextualise the application of the SAI system under examination; (c) (For each dimension) to offer a series of questions, the answers to which facilitate the screening of the respective SAI system.

To be succinct, this article will not delve into the questions but rather aims to account for the five dimensions along which screening occurs in our method.

**AI model description.** First and foremost, it's essential to define the main features that identify the AI system being designed and implemented or that we want to consider in our analysis. In this initial impact assessment stage, developers are asked to describe the AI system under evaluation and specify the symbiosis's operating context. The entire developers' team should collaborate on completing this passage, explaining (a) the scope and users of the AI system, (b) the dependency on other AI systems, and (c) the *algorithm model description* [48] [63].

**Symbiosis description.** The elucidation of symbiotic levels holds pivotal significance in constructing the framework for the analysed SAI system. We have categorised the description into criteria and sub-criteria to address a diverse spectrum of symbiotic levels and dimensions comprehensively: (a) *biological and corporeal symbiosis*; (b) *symbiosis on task(s) to perform*; (c) *symbiosis on combining tasks and actions into multi-task, composite systems*; *symbiosis as autonomy level*.

**Data description.** Data encompasses the information and expert insights an AI model utilises to construct a representation of the context or environment. Expert input typically involves codifying human knowledge into rules. Key characteristics within this domain encompass the origin of data and information, the method of collection by either machine or human, data structure and format, and inherent data properties. These characteristics are pertinent to the data employed in training an AI system ('in the lab') and the data utilised in production ('in the field'). The following table provides a summary of these distinct criteria and sub-criteria: (a) *Data collection*; (b) *rights and identifiability*; (c) *structure and format*; (d) *quality and appropriateness* [48].

**Proportionality description.** In this framework descriptive phase, it is also crucial to comprehend the degree of proportionality between the AI system's promoted (or promised) services and utilities and

whether the intended purpose justifies its utilisation, considering the technology's risks, uncertainties, and drawbacks. This examination allows practitioners and experts to balance the means and the intended goal. It is a process to justify the necessity of utilising a specific method or AI system and to demonstrate its suitability. This ensures that procedures related to, or a part of, the AI system do not exceed what is necessary to achieve legitimate objectives [63].

**Governance description.** Conducting this screening is pivotal as it allows for clearly identifying critical roles and responsibilities, ensuring transparency and accountability within development teams. Indeed, in our method, framing the governance means paying attention to two intertwined aspects. On the one hand, the *identification of the responsibilities* (top-down). Given the inherent risks associated with AI, it becomes imperative to delineate clear lines of responsibility for each facet of the AI system. Establishing this governance framework not only safeguards against ambiguity in accountability but also enhances trust, ethical adherence, and effective decision-making within the development and deployment processes of SAI.

On the other hand, it is identifying *stakeholders' value chain* (bottom-up). Engaging diverse stakeholders is imperative for comprehensively evaluating the impacts associated with this SAI system. Consequently, the developers' team should devise a detailed stakeholder engagement strategy in the initial phases of system design. This strategy will enable the team to establish their engagement goals, which should be periodically reviewed. This ensures that stakeholder involvement is not merely a procedural task but becomes an integral and transformative element of the decision-making process.

#### 4. Conclusions and Future Work

A constructivist approach to symbiosis presents itself as the most viable alternative for reinstating human-centredness in the governance of SAI. Should we persist in adhering to rigid, deterministic foundational methods, we risk gradually divorcing ethics from being a mere adherence to moral principles (such as autonomy, beneficence, non-maleficence, justice, etc. [21] [30]) to being a compelling aspect of machine design, and more broadly, in shaping the digital society we aspire to inhabit. Indeed, the notion of sustainable, fair, and trustworthy algorithms – AI ethics – cannot be solely relegated to compliance with

ethical principles. Ethics, in this vein, remains overly abstract and unnecessary [9], fails to serve stakeholders by providing clarity or guidance in development and commercialisation [46], and is at risk of being exploited maliciously [26] or even weaponised as a tool of biopower and colonisation [10].

We must relinquish the grip of determinism – the persistent desire to control centrally – and instead embrace a distributed, contextual, relational, and dialectical structure to embody the ecosystemic essence of human-centeredness truly. While a constructivist approach may initially appear to shift responsibility from humans to machines, it, in fact, addresses the issue of AI responsibility not by fixating on who holds control (humans or machines) but by methodically analysing and assessing the *chain of responsibility*. In essence, transitioning from the question of “what” to an examination of the technology in action [49] entails grappling with the question of “how” [46]. Given its sociotechnical nature, symbiosis with machines prompts a re-evaluation of responsibility, focusing less on its ontological core – who bears this faculty? – and more on its inherent interplay between the social and the technological within everyday organisational life – what has also been called “socio-material practices” [49] or “techno-scientific practices” [53].

In Chapter 2, we highlighted the main issues concerning the possible foundation of SAI. Crucial philosophical questions were raised about the possibility of thinking about a form of AI capable of entering symbiosis with humans, starting from the specificity that such a technology should exhibit compared to more traditional AI. In Section 2.1, we questioned the application of biological categories – such as symbiosis – to artefacts. While robotics and AI challenge the conventional distinction between living and non-living based on the capacity of organisms to move autonomously, self-develop and reproduce, some differences remain that currently appear insurmountable. Also, in Section 2.2, we offered a scan into three patterns of SAI based on as many definitions of intelligent life: the kind of SAI we can conceive of strictly depends on how we think about the continuity between organisation and intelligence. The less continuist our position, the less chance we have of putting down a foundation of SAI in a strong sense, namely from a biological or ontological point of view.

Chapter 3 explicitly addresses the potential of a constructivist approach in redefining the foundation of

SAI and its alignment with human-centred ethics. We advocate for adopting a constructivist perspective to navigate the intricate landscape of SAI, placing particular emphasis on attributes such as flexibility, adaptability, and context awareness. The discourse extensively explores the multifaceted challenges arising from the diverse foundational perspectives of SAI, proposing a constructivist framework aimed at comprehensively understanding and evaluating SAI systems. This section delineates the fundamental aspects of a theoretical (onto-epistemological) framework and an evolving evaluative methodology designed to assess the sociotechnical characteristics and human-centred attributes of SAI systems.

This method that we are working on will require further refinements as the FAIR project progresses. For this reason, in the future, the research will involve jurists to address the theme of acceptability from the ethical and legal viewpoints. Then, the legal and ethical acceptability of SAI will need to go through an operationalisation process to be of practical relevance for the design and implementation of SAI applications. High-level principles will be, therefore, turned into operational definitions that pave the way to technical solutions, *e.g.*, for (partially) automated compliance testing. Operationalisation will be accompanied by appropriate modelling of the SAI application in hand. Notably, socio-technical systems (STS) are widely recognised as a valuable approach to complex organisational work design that stresses the interaction between people and technology in workplaces [3]. Also, multi-agent systems (MAS) [15] are promising from the practical viewpoint since they enable the simulation of possible scenarios and the experimentation of different operational definitions of the legal and ethical acceptability of SAI in a controlled environment. A starting point might be the MAS prototype presented in [16] for the moral evaluation and monitoring of dialogue systems. Finally, compliance tests might be reformulated as problems that can be addressed with automated reasoning techniques and formal methods. Overall, along the direction already explored, *e.g.*, [17], logic will play a prominent role in implementing the computational solutions for many of the problems in our research on SAI within FAIR.

## 5. Acknowledgements

This work was partially supported by the project FAIR—Future AI Research (PE0000013), which is

part of the NRRP MUR program funded by NextGenerationEU.

## References

- [1] Aristotle, *De Anima*, Clarendon and Oxford University Press, Oxford, 2016.
- [2] K. Barad, *Meeting the Universe Halfway. Quantum Physics and the Entanglement of Matter and Meaning*, Durham, Duke University Press, 2007.
- [3] G. Baxter and I. Sommerville, Socio-technical systems: From design methods to systems engineering, *Interacting with computers* **23**, 1 (2011) 4-17. DOI: <https://doi.org/10.1016/j.intcom.2010.07.003>.
- [4] S. Boschert, T. Coughlin, M. Ferraris, F. Flammini, J. Gonzalez Florido, A. Cadenas Gonzalez, P. Henz, D. de Kerckhove, R. Rosen, R. Saracco, A. Singh, A. Vitillo, M. Yousif, *Symbiotic Autonomous Systems*. A FDC Initiative, White Paper III, IEEE, 2019. URL: [https://digitalreality.ieee.org/images/files/pdf/ISAS\\_WP3\\_Nov2019.pdf](https://digitalreality.ieee.org/images/files/pdf/ISAS_WP3_Nov2019.pdf).
- [5] T. Brouwer, R. Ferrario, and D. Porello, Hybrid Collective Intentionality, *Synthese* **199** (2021), 3367-3403. DOI: <https://doi.org/10.1007/s11229-020-02938-z>.
- [6] A. Carnevale, A. Lombardi, F. A. Lisi: Exploring Ethical and Conceptual Foundations of Human-Centred Symbiosis with Artificial Intelligence. In G. Boella, F. A. D'Asaro, A. Dyoub, L. Gorrieri, F. A. Lisi, C. Manganini, G. Primiero: Proceedings of the 2nd Workshop on Bias, Ethical AI, Explainability and the role of Logic and Logic Programming co-located with the 22nd International Conference of the Italian Association for Artificial Intelligence (AI\*IA 2023), Rome, Italy, November 6, 2023. CEUR Workshop Proceedings 3615, CEUR-WS.org 2023: 30-43. URL: <https://ceur-ws.org/Vol-3615/paper3.pdf>
- [7] D. Chalmers, Singularity, *Journal of Consciousness Studies*, **17** (2010), 7-65.
- [8] M. Coeckelbergh, *AI Ethics*, MIT Press, Boston, 2020.
- [9] K. Crawford and R. Calo, There is a blind spot in AI research, *Nature* **538** (2016), 311-313. DOI: 10.1038/538311a.
- [10] K. Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, Yale University Press, Yale, 2022.
- [11] C.R. Darwin, *Old & useless notes about the moral sense & some metaphysical points*, J. van Wyhe, ed., CUL-DAR91.4-55, 1838-1840. URL: <http://darwin-online.org.uk/content/frameset?pageseq=1&itemID=CUL-DAR91.4-55&viewtype=text>.
- [12] G. De Anna, et al. *Filosofia morale. Fondamenti, metodi, sfide pratiche*, Le Monnier, Firenze, 2019.
- [13] F. M. De Collibus, *Are Large Language Models "alive"?*, 2023. URL: <https://philpapers.org/archive/DECALL.pdf>.
- [14] D.C. Dennett, *From Bacteria to Bach and Back. The Evolution of Minds*, Penguin Books, London, 2018.
- [15] A. Dorri, S.S. Kanhere and R. Jurdak, Multi-Agent Systems: A Survey, *IEEE Access* **6** (2018) 28573-28593. DOI: 10.1109/ACCESS.2018.2831228.
- [16] A. Dyoub, S. Costantini, I. Letteri, F.A. Lisi, A logic-based multi-agent system for ethical monitoring and evaluation of dialogues, in: A. Formisano, Y.A. Liu, B. Bogaerts, A. Brik, V. Dahl, C. Dodaro, P. Fodor, G.L. Pozzato, J. Vennekens, N. Zhou (Eds.), Proceedings 37th International Conference on Logic Programming (Technical Communications), ICLP Technical Communications 2021, Porto (virtual event), 20-27th September 2021, volume 345 of EPTCS, 2021, pp. 182-188. URL: <https://doi.org/10.4204/EPTCS.345.32>.
- [17] A. Dyoub, S. Costantini, F.A. Lisi. Learning Domain Ethical Principles from Interactions with Users, *Digital Society* **1**, 28 (2022). <https://doi.org/10.1007/s44206-022-00026-y>.
- [18] L. Edwards and M. Veale (2017). Slave to the algorithm? Why a right to explanation is probably not the remedy you are looking for. *16 Duke Law & Technology Review* **18** (2017), Available at SSRN. DOI: <https://doi.org/10.2139/ssrn.2972855>.
- [19] European Commission, *General Data Protection Regulation* (GDPR), Brussels, 2016. URL: <http://data.europa.eu/eli/reg/2016/679/oj>.
- [20] European Commission, *Building Trust in Human-Centric Artificial Intelligence*. COM/2019/168 final, Brussels, 2019. URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52019DC0168>
- [21] European Commission, *White Paper on Artificial Intelligence: A European approach to excellence and trust*, Brussels, 2020. URL: [https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust\\_en](https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en)
- [22] European Commission, EU AI Act. 2021/0106(COD), Corrigendum, 19 April 2024, 2024. URL: [https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01\\_EN.pdf](https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138-FNL-COR01_EN.pdf)
- [23] A. Feenberg, *Technosystem: The Social Life of Reason*, Harvard University Press, 2017.
- [24] L. Floridi, *The 4th Revolution: How the Infosphere is Reshaping Human Reality*, Oxford, Oxford University Press, 2014.
- [25] L. Floridi, et al., AI4People – An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations, *Minds and Machines* **28**, 4 (2018), 689-707. DOI: <https://doi.org/10.1007/s11023-018-9482-5>.
- [26] L. Floridi, *The Ethics of Artificial Intelligence*. Oxford University Press, Oxford, 2023.
- [27] H. Fry, *Hello World: Being Human in the Age of Algorithms*, W. W. Norton & Company, New York, 2018.
- [28] S.S. Grigsby, Artificial Intelligence for Advanced Human-Machine Symbiosis, in: *Augmented Cognition: Intelligent Technologies*, D. Schmorow, C. Fidopastis, eds., Lecture Notes in Computer Science, vol 10915, Springer, Cham, 2018, pp 255-266. URL: [https://doi.org/10.1007/978-3-319-91470-1\\_22](https://doi.org/10.1007/978-3-319-91470-1_22).
- [29] N. K. Hayles, Technosymbiosis: Figuring (Out) Our Relations to AI, in: *Feminist AI*, J. Browne, S. Cave, E. Drage, and K. McInerney, eds., Oxford University Press, Oxford, 2023, pp. 1-18. DOI: <https://doi.org/10.1093/oso/9780192889898.003.0001>.
- [30] HLEGAI, *Ethics guidelines for trustworthy AI*, Brussels, 2018-2019. URL: <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>.
- [31] D. Ihde, *Technology and the lifeworld: From garden to earth*, Indiana University Press, 1990.
- [32] H. James Wilson and P.R. Daugherty, Creating the Symbiotic AI Workforce of the Future, *MIT Sloan Management Review*, 21 October (2019). URL: <https://sloanreview.mit.edu/article/creating-the-symbiotic-ai-workforce-of-the-future/>.
- [33] M. Katell, et al., Toward situated interventions for algorithmic equity: Lessons from the field. Proceedings of the Conference on Fairness, Accountability, and Transparency, Fat\* 19, ACM Press, 2020, pp. 45-55. DOI: <https://doi.org/10.1145/3351095.3372874>.
- [34] D.E. Koshland, The seven pillars of life, *Science* **295**, 5563 (2002), 2215-2216. DOI: 10.1126/science.1068489.

- [35] P. Kotipalli, Symbiotic Artificial Intelligence, 22 June, 2019. URL: <https://p13i.io/posts/2019/06/symbiotic-ai/>.
- [36] B. Latour, *Reassembling the social: An introduction to actor-network-theory*, Oxford University Press, 2005.
- [37] K.F. Lee and C. Qiufan, *AI 2041: Ten Visions For Our Future*, Penguin Random House, New York, 2021.
- [38] D. Lévy, Are Robots Alive?, in: *Human-Robot Intimate Relationships*, A.D. Cheok and E.Y. Zhang, eds., Human-Computer Interaction Series, Springer Nature Switzerland AG, 2019, pp. 155-188. DOI: [https://doi.org/10.1007/978-3-319-94730-3\\_8](https://doi.org/10.1007/978-3-319-94730-3_8).
- [39] J.C.R. Licklider, Man-Computer Symbiosis, *IRE Transactions on Human Factors in Electronics HFE-1* (1960), 4-11.
- [40] E.M. Macphail and H.B. Barlow, Vertebrate Intelligence: The Null Hypothesis, *Philosophical Transactions of the Royal Society B*, Biological Sciences **308**, 1135 (1985), 37-51, DOI: <https://doi.org/10.1098/rstb.1985.0008>.
- [41] R. Manzotti and S. Rossi, *IO & IA*, Rubbettino, Soveria Mannelli, 2023.
- [42] S. Mason Dambrot, D. de Kerchove, F. Flammini, W. Kinsner, L. MacDonald Glenn, R. Saracco, *Symbiotic Autonomous Systems*. A FDC Initiative, White Paper II, IEEE, 2018. URL: <https://digitalreality.ieee.org/images/files/pdf/SAS-WP-II-2018-Finalv3.2.pdf>.
- [43] H.R. Maturana, F.J. Varela, *Autopoiesis and Cognition. The Realisation of the Living*, Reidel Publishing Company, Dordrecht, 1980.
- [44] E. Mayr, *The Growth of Biological Thought*, Harvard University Press and Belknap, Cambridge, 1982.
- [45] G. Mota-Valtierra, J. Rodríguez-Reséndiz and G. Herrera-Ruiz, Constructivism-Based Methodology for Teaching Artificial Intelligence Topics Focused on Sustainable Development, *Sustainability* **11**, 17 (2019), 4642. DOI: <https://doi.org/10.3390/su11174642>.
- [46] J. Morley, L. Floridi, L. Kinsey and A. Elhalal, From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Science and Engineering Ethics*, **26**, 4, 2141-2168. DOI: <https://doi.org/10.1007/s11948-019-00165-5>.
- [47] S. Natale, *Deceitful Media: Artificial Intelligence and Social Life after the Turing*, Oxford University Press, Oxford, 2021.
- [48] OECD, Framework for classifying AI systems, *OECD Digital Economy Papers* **323** (2022), OECD Publishing, Paris. DOI: <https://doi.org/10.1787/cb6d9eca-en>.
- [49] W.J. Orlikowski, Sociomaterial Practices: Exploring Technology at Work, *Organization Studies*, **28**, 9 (2007), 1435-1448. DOI: <https://doi.org/10.1177/0170840607081138>.
- [50] A. Ovalle, et al., Factoring the Matrix of Domination: A Critical Review and Reimagination of Intersectionality in AI Fairness, *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '23, Association for Computing Machinery, New York, 2023, pp. 496-511. DOI: <https://doi.org/10.1145/3600211.3604705>.
- [51] M. Paluch, Laying The Foundation For Symbiosis Between Humans And Machines, *Forbes*, 18 December (2019). URL: <https://www.forbes.com/sites/forbestechcouncil/2019/12/18/laying-the-foundation-for-symbiosis-between-humans-and-machines/?sh=3f4ab8682997>.
- [52] J.P. Ponda, *Artificial Intelligence (AI) through Symbiosis*, Undergraduate Thesis, Georgia Institute of Technology, 2022.
- [53] F. Russo, *Techno-Scientific Practices: An Informational Approach*, Rowman & Littlefield, Lanham, 2022.
- [54] R. Saracco, R. Madhavan, S. Mason Dambrot, D. de Kerchove and T. Coughlin, *Symbiotic Autonomous Systems. A FDC Initiative*, White Paper, IEEE, 2017. URL: <https://digitalreality.ieee.org/images/files/pdf/sas-white-paper-final-nov12-2017.pdf>.
- [55] G. Sartor, *L'informatica giuridica e le tecnologie dell'informazione*, 3th edition, Giappichelli, Torino, 2016.
- [56] A.D. Selbst, et al., Fairness and abstraction in sociotechnical systems, Proceedings of the Conference on Fairness, Accountability, and Transparency, Fat\* 19, ACM Press, 2019, pp. 59-68. DOI: <https://doi.org/10.1145/3287560.3287598>.
- [57] R. Shafer-Landau, *The Fundamentals of Ethics*, Oxford University Press, Oxford, 2017.
- [58] D. Shin, and Y.J. Park, Role of fairness, accountability, and transparency in algorithmic affordance, *Computers in Human Behavior* **98** (2019), 277-284. DOI: <https://doi.org/10.1016/j.chb.2019.04.019>.
- [59] C. Siegmann and M. Anderljung, *The Brussels Effect and Artificial Intelligence*, Centre for the Governance of AI, Oxford, 2022. URL: [https://cdn.governance.ai/Brussels\\_Effect\\_GovAI.pdf](https://cdn.governance.ai/Brussels_Effect_GovAI.pdf)
- [60] P. Singer (ed.), *In defence of animals*, Basil Blackwell, Oxford, 1985.
- [61] E. Sober, Learning from Functionalism: Prospects for Strong Artificial Life, in: *Artificial Life II*, C. Langton, C. Taylor, J.D. Farmer and S. Rasmussen, eds., SFI Studies in the Sciences of Complexity, Proceedings Vol. X, Addison-Wesley Publishing Company, Redwood City, 1991, pp. 749-765. DOI:10.1017/CBO9780511730191.021.
- [62] T. Stamer, *Using Wearable Devices to Teach Computers*, AWE Europe, 2017. URL: <https://www.youtube.com/watch?v=hi9RYBaPVPI>.
- [63] UNESCO, *Ethical impact assessment: a tool of the Recommendation on the Ethics of Artificial Intelligence*, Paris, 2023. URL: <https://unesdoc.unesco.org/ark:/48223/pf0000386276>.
- [64] P.-P. Verbeek, *Moralising technology: Understanding and designing the morality of things*, University of Chicago Press, Chicago, 2011.
- [65] P.-H. Wong, Democratising algorithmic fairness, *Philosophy & Technology* **33** (2019), 225-244. DOI: <https://doi.org/10.1007/s13347-019-00355-w>.