A Procedural Idea of Decision-making in the Context of Symbiotic AI

Piero Marra [1], Lorenzo Pulito [2], Antonio Carnevale [3], Antonio Lombardi [3], Abeer Dyoub [4] and Francesca A. Lisi [4]

[1] University of Bari Aldo Moro, LAW Dept., Piazza C. Battisti 1, Bari, 70121, Italy
[2] University of Bari Aldo Moro, DJSGE Dept., Via Duomo 259, Taranto, 74123, Italy
[3] University of Bari Aldo Moro, DIRIUM Dept., Piazza Umberto I, Bari, 70121, Italy
[4] University of Bari Aldo Moro, DiB Dept., via E. Orabona 4, Bari, 70125, Italy

Abstract

The European legal framework on Artificial Intelligence pays little attention to regulating and shaping technologies where humans and artificial intelligence cooperate in a two-way relationship. The research field is technologically challenging. The paper results from a foundational study aiming to conceptualise and design a holistic symbiotic approach to Artificial Intelligence to have fair, legitimate, and effective outputs, ensuring their ethical and legal acceptability. This theoretical study would impact Symbiotic AI systems' development and technological governance via model assessment.

Keywords 1

Symbiotic Artificial Intelligence (SAI), Human-Centred and Human-Centric Computing, Procedural Decision-making, Operational Implications
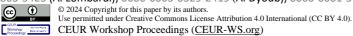
1. Introduction

Article 3 of the Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (AI Act) defines an AI system as a 'machine-based system designed to operate with varying levels of autonomy, and that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.'

The EU regulatory framework contemplates AI only as an autonomous product. However, this definition overlooks those AI systems that can foster integrated human-machine cooperation. Systems of so-called *Symbiotic Artificial Intelligence* (SAI) are the most significant among these. In the context of SAI, symbiosis means and implies the integration of Artificial Intelligence (AI) and Human Intelligence (HI) to produce more fair and efficient decision-making outputs by combining these two types of intelligence to complete complex tasks. But what exactly does SAI imply?

There needs to be a clear and holistic definition of symbiosis. For this reason, the paper, starting from conceptual issues, includes in SAI all the technologies that allow human and artificial agents to mutually assist each other in achieving a common goal. Nonetheless, the true technological challenges are broader than the goal since the AI system should also be able to reason about human actions while considering their mental models.

A Symbiotic AI (also known as human-AI symbiosis) can be described as a *socio-technical construct*. Socio-technical teaming refers to the collaborative partnership between humans and machines within a broader social and technological context, where the focus is not on a substantial peer-to-peer relationship but on integrating technology into human-centric processes [1].

Nevertheless, more than the foundational design of a SAI system is required. It is also necessary to look at 'how' such technology is concretely implemented in a socio-technical context. In this sense, it is also helpful to investigate the mechanisms underlying human decision-making to develop an operational model based on a human-centred approach.

Not surprisingly, the European regulation emphasises the relevance of AI outputs, generally identified as decisions (including predictions, contents, and recommendations). Yet, reference is linked to the decision but not to the procedures through which humans can participate in decision-making mechanisms, controlling them, rather than simply passively enduring the influences of automatic decisions, as stated by the cited legal rule.

By analysing the history and the principles allowing a definition and an assessment of SAI systems (Section 2), considering symbiotic technology as a socio-technical construction (Section 3) and a legal principle (Section 4, particularly 4.1.1.), the paper focuses on the legal procedural model as a paradigmatic methodology of Human-machine cooperative decision-making outputs (Section 4.1.2 and 4.1.3). To understand the practical implications of this model, the criminal trial is paradigmatic of how a procedural concept of symbiosis can be conceived and modelled since due process is legally understood as the cooperative relationships among the individuals involved in the decision-making (Section 4.2 and 4.3.). The conclusive reflections address research perspectives: both the consideration of symbiosis as a socio-technical construct and as a procedural model of decision-making lead to the need to operationalise ethical and legal principles to make human-machine collaboration fair, legitimate, and effective (it is the theme of Section 5).

The paper is the result of a 'conceptual' analysis within the foundational research done by the University of Bari (together with INFN) within the NRPP-funded project Future AI Research (FAIR). From a methodological point of view, the conceptualisation of the SAI is relevant for the technological design and for identifying an assessment model whose impacts can affect their understandability, acceptability, and sustainability (Section 6). Notably, the impacts of the conceptualisation phase do not arise from assessment results.

2. Symbiosis and AI: History, Foundational Principles, Examples

The notion of symbiosis originated in the 19th century to indicate a relationship between two taxonomically separate life forms that nevertheless give rise to a single organism. In 1868, botanist Simon Schwendener proposed the so-called 'dual hypothesis' to explain the nature of lichens as an association between a fungus and an algae. Schwendener went so far as to say it was a master-servant relationship in which the fungus enslaved the algae to exploit its autotrophism. This hypothesis met with resistance because, in addition to appearing as a systematic abomination, it portrayed the nature of certain organisms as intrinsically marked by a dimension of cruelty, elevating a form of parasitism to a taxonomic category. Later, more neutral terms were coined to refer to this relationship between living beings, such as 'consortium' proposed by Johannes Reinke (1873), 'Symbiotismus' proposed by Albert Bernhard Frank (1877), and 'Symbiose' used by Anton De Bary (1878). These uses of the term envisaged its neutral connotation: Reinke proposed considering the fungus and the algae in the lichen as if they were, respectively, the root and the leaf of a plant; Frank understood symbiosis as a concept that did not take into account the role assumed by the two symbionts; De Bary as a simple living together of two life forms belonging to different classifications [2].

Symbiosis seems to mean something obvious: life forms are not isolated but coexist in ways that are essential to their survival and development. The endosymbiotic hypothesis, brought to the fore by Lynn Margulis [3], has even gone so far as to support the symbiotic origin of eukaryotic life itself

on Earth: organelles such as mitochondria would, in the past, have been organisms that later entered into an inseparable symbiotic relationship with cells. Life, as such, would be symbiotic.

When applied to AI, symbiosis becomes more complex, posing a whole series of philosophical, scientific, and generally foundational problems [1]. The first to juxtapose symbiosis and computer science was J.C.R Licklider [4], advocating a symbiosis between man and machine. In his view, this kind of symbiosis would allow the computer to become an active part of the thinking process that leads to resolving technical problems and not just an executor of solutions thought up beforehand. Licklider was mainly thinking of human-computer interfaces that would allow greater real-time collaboration and shorten the distance between human and machine language. A road that has since been successfully travelled. However, more than sixty years later, the principles on which SAI is based appear not significantly different. Contributions in this regard are very few and primarily exploratory, and the very notion of SAI has yet to be precisely determined. From a survey of the scant bibliography on the subject, the principles on which a SAI is based tend to be the following four:

1. Timeliness, i.e. the reduction of time and effort between intention and action [5].
2. Active Cooperation, i.e. a more pronounced shift in the role of the human being from 'teacher' to 'collaborator' and of the machine from passive learner to active learner [6, 7].
3. Autonomy, i.e. the increasing ability of the AI to learn and make decisions with weak or no supervision [8].
4. Seamlessness, i.e. an increasingly less interrupted interaction between humans and AI on a scale ranging from wearable devices (e.g., Apple Vision Pro) to implantable ones (e.g., Musk's Telepathy) [9, 10, 11].

Although these characteristics provide an instrumental orientation for defining and evaluating cases of symbiotic AI, the latter presents fundamental difficulties that appear hard to overcome. These obstacles derive, above all, from the problem of speaking of symbiosis between intelligent life (ours) and a form of agency that neither presents all the characteristics of life nor all those of intelligence. Therefore, it might be appropriate to adopt a deflationary approach to SAI [1]. It does not forbid the use of the category of symbiosis in a cautious but fruitful sense from the point of view of the classification and evaluation of these systems. One could return to whether symbiosis is an inherently neutral concept or does not always conceal an asymmetry, as in the case of Schwendener's dual hypothesis. We could ask a provocative question: if SAI is a lichen, are we the fungus or the algae?

Symbiotic relationships can indeed be classified according to different parameters or levels. 1) At the level of commitment, symbiosis can be a) facultative or b) obligatory. 2) At the 'spatial' level, it can be a) an endo-symbiosis, b) an ecto-symbiosis or c) an exo-symbiosis, depending on whether one of the two symbionts 'lives' inside, on the surface or outside the other symbiont (and in the case of SAI, this presents particular problems). 3) At the level of benefits, a symbiosis can be a) commensalistic, b) mutualistic or c) parasitic. For classificatory and evaluative purposes, we could place the SAI system at each level and draw conclusions about its acceptability.

Take two systems that we might consider symbiotic: 1) Netflix's recommendation system and 2) NPCs in AI-powered games.

1. Nello Cristianini [12, 13] has presented today's intelligent agents as 'alien' life forms. They share a habitat with us, like plants or snails, while having an intelligence utterly different from ours. They pursue their goals through strategies that can sometimes harm us, even if we do not realise it. Recommendation systems, e.g., can exploit our cognitive biases or put in place 'nudges' regardless of our real interest in the content. It is a clear case of how SAI can sometimes take on a parasitic connotation [14]. In the case of these recommendation systems, we could also ask to what extent symbiosis is facultative and to what extent it is an endo- or exo-symbiotic relationship. The system is indeed 'outside of me,' but it is also true that it exploits the extension of my mind

constituted by my interactions with the web and social networks to achieve its aims. It is also somehow 'inside' me.

2. NPCs (Non-Playable Characters) have constantly populated the world of video games to lend realism to scenarios and make gaming experiences captivating, interactive and dynamic. Only in recent years, however, algorithms such as Bethesda Softworks' Radiant AI have been used to make the behaviour of NPCs less automatic and more independent of the leading player. In games like Oblivion or Fallout, NPCs are equipped with objectives that they must pursue autonomously in the environment. They have a life of their own. This obviously makes the game much more realistic. Still, over time, unexpected emergent behaviour has been observed. For instance, some characters who were tasked to achieve a specific goal with a particular tool but also could sell their possessions to buy food and survive killed the characters they had sold their tool to fulfil their goal. We can well imagine the ethical dilemmas this would raise in artificial ecosystems such as the metaverse, where we can immersively interact with NPCs through VR devices [15, 16, 17]. The future is not far off when we can live symbiotically with these intelligent agents. What kind of symbiosis will this be? What challenges await us?

3. Socio-Technical Dynamics: Unveiling SAI's Impact on Human Interaction

Given AI's increasing integration with human interactions, the concept of 'symbiosis' manifests across diverse levels [1]. Consider another simple case: the advanced filters and AI-based features many social platforms offer. Particularly when discussing beauty filters [18], which alter the physical appearance of the depicted individual in photos to 'enhance' it, we observe various symbiotic dynamics at play among the person, the algorithm, and the environmental and value-laden context. An example could be the Bold Glamour filter [19], an ultra-realistic filter that has sparked significant controversy, partly due to the massive user uptake [20]. Crafted to augment body-imagined aesthetics [21], these filters exhibit a propensity to refine and contour facial attributes meticulously, amplifying their prominence by fostering a smoother skin texture, delicately administering makeup with professional finesse, achieving a semblance of natural appearance, and employing intricate lighting techniques. These effects create an incredibly lifelike image, aided by the fact that they seem to remain 'attached' to the face, not disappearing as the person moves or objects intervene between the face and the screen. The perception of sustained attachment to the visage presents an initial form of symbiosis at the level of user experience. Put differently, the differentiation between the mask (represented by the filter) and the natural face, which remained in earlier iterations of filters, is assimilated through the utilisation of Generative Adversarial Network (GAN) [22, 23, 24], resulting in a synthetic, illusory, and body-imagined fusion [25]. This heightened realism of the filter complicates its detection. This primary symbiosis level relates to the efficacy of the technological solution, thereby situated within the UX-IA paradigm.

However, there is also a second level. Since the filters result from GANs, the use of the filter tends to reduce real faces to a single model, not only in terms of physiognomy but also aesthetic-cultural aspects (the filter applies a certain makeup style only to faces it codes as feminine). Reality is thus parameterized, relating it to reference terms containing a converging value judgement, reproducing symbiosis at a higher systemic level—that of UX, namely in social processes of identification and cultural processes of normalisation.

From this easy example, it appears evident that we cannot fix the evaluation of SAI on traditional notions of symbiosis understood as an inherent, predetermined, or entity-based state. Instead, redefining symbiosis as a product of techno-scientific practices [26] portrays it as a dynamic possibility rather than a static and principle-based reality. Principles alone cannot guarantee ethical AI [27, 28] also because both humans and artificial agents are active contributors to knowledge production and teaming collaborations [1]. As a new form of 'agential materialism' [29], the symbiosis doesn't inherently exist beforehand, but instead emerges as a 'Δ phenomenology' [30]; it emerges from within intra-actions between the model of symbiosis and its hybridised actors [31]. Symbiosis functions as a

socio-technical construct: any philosophical, ethical, or legal inquiry into the evidence and truth of symbiosis should approach through techno-scientific methodologies that promote a 'from-what-to-how' understanding [28] of human-machine symbiosis conditions. This entails delving into process-oriented ontologies rather than entity-focused ones and reconsidering causal mechanisms in the context of information transmission [26]. While entity-focused ontologies concentrate on discrete entities and their attributes, process-oriented ontologies shift the focus to dynamic processes and the interactions between entities.

4. Symbiosis and 'significant human control' in decision-making

EU regulation neglects human active and cooperative participation in the decision-making together with the machine. In fact, according to Article 3 of the EU Artificial Intelligence Act, there is no space for SAI as a socio-technical dynamic construct since technicality is only considered an influential concept rather than a relational one.

4.1. A methodological framework

From a legal methodology standpoint, it is interesting to know and identify the formal legal conditions of algorithmic decision-making and definitely 'how' SAI can produce legally valid decisions (i.e. effective, attributable, and accountable to a human person) in such fields of complex decision-making, such as the legal process and legal tech. The attention to formal conditions of existence (i.e. legal validity) of a specific decision looking at its use (i.e. its pragmatic function) depends on the fact that law requires the formalisation of relational conditions to be effective. Thus, three are the main issues: a) the theoretical legal idea of symbiosis, b) the epistemic legal model of SAI decisions, and c) the fundamental conditions of legal acceptability of SAI decisions.

4.1.1. Symbiosis as a legal principle

Theoretically, symbiosis can be considered a legal principle rather than a legal value. The two concepts are often confused as they may refer to the same 'resource.' The difference lies in 'how' the protected resources are considered.

Value is a final good evaluated as an 'end in itself.' It does not contain a criterion of legitimacy of the action or judgement [32, 33]. The criterion of the action and judgement legitimacy is not in the value but in the efficiency concerning the end-value. The end contains the justification for each action. Acting by 'values' is refractory to prior regulative and delimitative criteria, as they cannot be traced back to pre-determinable reasons.

Principles, unlike values, can be considered as initial goods that require 'consequently determined activities' [33, 34]. They are concerned with the means of our actions, not the ends. For this reason, unlike values, principles have a normative content regarding action. The principle becomes a criterion of the validity of the action itself. Acting according to principles is intrinsically regulated and delimited by the principle and its implications.

Like a principle, symbiosis involves human agency, which appeals to being measured and measurable and indicates its criterion of legitimacy.

4.1.2. Symbiosis as a procedural condition

In an epistemic key, the focus goes to the teleological aspect of each human's actions. Fair and legitimate decision-making must be conducted towards an end, leading to a process-oriented approach. The procedural adjective qualifies a certain criterion, canon, or principle as a formal condition of a specific activity. In a normative dimension, procedures do not directly indicate what should be done but 'how' something should be done [35, 36].

The term 'procedural' is appropriate for identifying a legal concept of symbiosis. It is related to the 'ways' in which a symbiotic intelligent system must be built and designed if it must be 'effective' and at the same time remain what it aims 'to be', i.e. a system that considers humankind as an integral part of the symbiotic decision-making. A procedural condition ensures the fairness and transparency of decision-making and prevents effective justice because it allows recipients to understand and respect the decision. In fact, effectiveness remains a constitutive element of legality [37]. This approach is thought on the relationships among the individuals involved in the decision-making to improve user treatment, judgement enforcement, and system trust. It is shown that if there is a perception of procedural justice, there is also greater acceptance of decisions. The perception of procedural justice positively correlates with users' judgement of the entire judicial system. For example, the American Judges Association, the Center for Court Innovation, the National Center for State Courts, and the National Judicial College have identified the following factors as relevant for procedural justice: voice, neutrality, respect, trust, understanding, helpfulness [38, 39, 40, 41, 42, 43]. As can be seen, a procedural construction of SAI has the advantage of leaving the fair decision to the human decision-maker, looking at its effectiveness. Something that an autonomous system would not be ontologically able to ensure.

4.1.3. A legal foundation of symbiotic procedural-oriented approach

The 'foundation' of a procedural methodology of SAI shifts the reflection from the human-machine interaction to that of pragmatic conditions of legality of the decision process itself.

From a pragmatic point of view, looking at the use (or decision in action) of SAI decision-making, a 'significant human control' is central since:

1. It regulates the internal control of the decision in terms of justification.
2. It proceduralises the effects of the decision externally on the recipients/users.

Both factors require 'reinforced motivation' as a legal pragmatic condition of acceptability of SAI. The concept of 'significant human control' has been inspired by the international debate within the UN on lethal autonomous weapon systems (LAWS). The used notion is that of 'meaningful human control,' which generalizes the need for operational control over technological artifacts to AI systems and mainly refers to morally consequential decisions without appropriate control from responsible humans. However, in the judicial field, it is difficult to have full control over the execution of the predictive algorithm. The risk is that of altering the correctness of the cross-examination and affecting the judge's free conviction. And the risk of using probabilistic outcomes in the process must also be considered [44]. It means that legally compatible decision-making requires a 'strict motivation' that can investigate and explain the 'project' of algorithmic decision-making. Moreover, a compelling rationale presupposes that the system is modelled respecting the principles of knowability and comprehensibility of the adopted logic [45], while the symbiotic principle ensures both non-exclusivity of the decision-making, and non-discrimination through algorithmic means. Here, the control is 'significant' because it is both qualitative and quantitative, affecting the decision-making and its effectiveness.

In this scenario, the judicial process and its function can be considered an emblematic manifestation of how a procedural symbiosis could be conceived and modelled.

4.2. The criminal trial and predictive justice as a paradigmatic phenomenon

AI systems offer the opportunity to positively impact various aspects related to justice [46, 47]. In a highly discretionary area of human activity, such as criminal justice, literature is focused on the possible role of these technologies to understand how they can restore effectiveness, credibility, and trustworthiness.

The fourth recital of the AI Act focuses on the benefits arising from using AI, for example, in improving predictions and in matters of security and law. Now, numerous predictive outputs are required in various areas of criminal proceedings. It is worth mentioning the risk assessment that an authority has to make towards a victim of gender-based violence in order to decide whether to apply appropriate protection measures (i.e. urgent exclusion orders, restrictive orders or protection orders) [48]; the prediction of the risk of reoffending before applying a personal precautionary measure; the assessment of social danger that is the basis for the supervisory judges' decision on whether to grant an alternative measure to detention.

The use of predictive AI systems—and machine learning approaches in particular—can help to optimise the above evaluation processes. However, it raises various challenges, both at a technical, ethical, and legal level. This is particularly evident from the experience of using risk assessment tools such as Compas and Savry, that exemplify how the algorithmic system can be compromised by racial bias, opacity, and incomprehensibility [49, 50].

Recognising these implications is important to effectively mitigate the associated risks and identify the conditions for acceptance of such AI systems to ensure reliable assessments in safeguarding human rights. This goal can be achieved through a Human-machine interaction following a symbiotic paradigm [51]. But what does this symbiosis mean in the field of criminal justice?

The concept involves designing AI systems within a human-centred approach, according to which the modelling of AI systems focuses not on humans' actions (the decisions they make) [52] but on the reasons behind them (the rule of law). In this way, the interaction between human beings and the machine is 'biunivocal,' but this does not exclude the possibility of one part exercising some control over the other. This means that the legal decision should not be made using exclusively algorithmic results, but such use, to be legally acceptable, is or can be subject to 'significant human control' [53]. The control may be exercised over the general validity of the theory underlying the software (according to the 'Daubert test') and the codification of that theory within the software [54].

When specifically focusing on risk prediction tools in criminal justice (for example, in relation to sentencing), the symbiotic approach, absent in the AI Act, recognises that risk assessment is a process for assessing the presence of risks and not a technique to achieve an accurate prediction. It can be defined as a means and not a goal [55].

In the framework of the symbiotic collaborative procedural strategy, on the one hand, the predictive algorithm can provide a comprehensive cognitive contribution that extends not only to risks but also to individual needs. It is worth mentioning the risk and needs assessment tool systems such as ORAS [56]. On the other hand, the methods to respond to the offence would be enriched in accordance with the fundamental principles of proportionality and dignity of the human being through the individualisation of the punishment and concretisation of its educational function [55], improving the human experience while considering values.

4.3. Some implications from the legal procedural approach to SAI

For a legally acceptable and fair decision, inferential correctness is not enough; it is also necessary to ensure its effectiveness. In this way, the procedural consideration of AI not only looks at the benefits of the automation process but also ensures human participation even in the enforcement phase.

The procedural approach, blending the form of decision-making with the substance of goals, could establish interpretative processes tailored to the concrete situation and inspired by constitutional substantial principles related to criteria of balance, reasonableness, and proportionality of choices.

For example, introducing technology into the process can improve its functioning, but only if it does not distort the human and institutional profile of justice, that consists of the ability to mediate and reconcile relationships through the process, which is an anthropological model [57].

Innovative VR tools that represent solutions capable of changing the decision-making process scenarios within the trial, could be an exciting field of application for this approach.

One facilitating aspect for judges could be the possibility to create and replicate crime scenes that could be accessed and analysed long after the crime occurred. On the one hand, this application may

help preserve the original crime scene; on the other hand, it may allow for a deeper examination of the evidence. These innovative solutions may allow the judge to visit the crime scene virtually better to understand the context and details of the cases, leading to more informed and conscious decisions. In essence, these cutting-edge virtual reality systems leverage spatial perception and cognition to create an immersive workspace where judges would benefit from a unique method to address complex situations, reshaping how human decision-makers approach complex cases.

If the benefits mentioned above would undoubtedly improve the entire judicial system and decision-making, challenges are similarly to be considered. The first aspect pertains to data privacy and ethical concerns, especially in the case of the intersection between the real world and metadata. A second challenge lies in the ambiguity of the legal framework applied to metaverse criminal situations. Lastly, it is configured for the emergence of new criminal threats that are strictly related to the cyberworld. There exists the risk that AI may generate environments that facilitate or encourage criminal situations. More specifically, non-player characters (NPC) may be manipulated to commit criminal offences, disseminate illegal materials, or deceive other users.

It is undoubtful that the creation of the metaverse using generative AI technology represents a fascinating perspective enriched by infinite virtual environments, dynamic narrative paths, and diverse and multiple NPC. Although fascinating, the metaverse also represents the hub for potential and sometimes unpredictable criminal risks.

For this reason, using VRs in the trial may only be acceptable in a socio-technical and procedural context. Socio-technically, humans provide the cognitive and emotional capabilities necessary for creativity, empathy, ethical decision-making, and adaptability, while machines offer computational power, data processing, and automation capabilities. On the other hand, the procedural approach allows for better protection of the rights of the parties by allowing them to verify the correct use of these systems, especially in conditions of noise, uncertainty, and small perturbations, ensuring that the judgment remains impartial and uninfluenced through the control of the motivation. The motivation must also concern the criteria underlying the outputs of the SAI system.

The human habitus that the procedure can guarantee avoids giving a blank check to algorithmic decision-making, leading individuals to take greater responsibility for their decisions, easing that of producers and developers. The procedure is the means of achieving a *human-centred* vision (concerning agency) and ensuring a *human-centric* one (concerning the protection of human dignity). In this way, SAI technology requires conditions of legal procedure to make the decisions reliable and effective according to a certain 'form of life' [58].

5. Operationalization Purpose for Ethics and Legal Issue in SAI decision-making

Operationalizing SAI ethics and legal issues entails translating abstract ethical principles and laws into practical guidelines governing all AI lifecycle stages, from data collection to deployment [59]. This multidisciplinary effort necessitates collaboration among computer scientists, ethicists, jurists, policymakers, and stakeholders to align with societal values and promote harmony between humans and AI systems, prioritising human well-being.

Operationalization in the SAI context entails developing robust frameworks for ethical risk assessment, addressing issues like bias and privacy violations [60]. This involves designing transparent, interpretable, and accountable AI systems, with logic programming playing a significant role in model design [61]. Such approaches enable stakeholders to understand AI decisions and rectify ethical issues proactively.

Additionally, operationalizing SAI ethics and legal issues requires ongoing monitoring and evaluation of AI systems in real-world contexts to ensure they continue to operate ethically and responsibly throughout their lifecycle.

From a technical standpoint, operationalization necessitates a human-centred approach, besides a human-centric one, emphasising the development of AI systems that prioritise transparency,

interpretability, and accountability. This involves implementing mechanisms such as explainability tools and algorithmic audits to enable users to comprehend AI decision-making. Furthermore, ensuring the reliability and robustness of AI systems through rigorous testing and validation processes is crucial to mitigate potential risks and foster trust in their deployment.

Operationalizing SAI ethics and legal issues requires embedding ethical principles and legal rules directly into the design and development processes of AI algorithms and models. This entails translating abstract ethical and legal principles into practical rules, guidelines, and standards to guide decision-making in practical dilemmas and concrete scenarios [62, 63]. Furthermore, SAI Ethics underscores the need for ongoing learning and adaptation as AI technologies evolve, requiring ethical standards and laws to evolve alongside them. SAI systems should also evolve and adapt their ethical behaviour accordingly [64, 65].

In summary, operationalizing SAI ethics and legal issues enables the cultivation of a collaborative and mutually advantageous relationship between humans and AI systems, facilitating responsible AI development for societal benefit.

6. Conclusion and impacts

AI technologies require a foundational reflection on how human-machine collaboration can be conceptualised and operationalised in compliance with ethical and legal principles. To investigate how human-machine collaboration can be performed, more is needed to use a human-centric approach, typically referring to AI systems that prioritise human needs, values, and well-being throughout their design, development, and deployment. This approach ensures that AI technologies align with human goals and aspirations, focusing on enhancing human capabilities and experiences. On the other hand, the human-centred approach emphasises the active involvement of human agency in AI development decision-making, which must also be considered, including aspects such as user-centred design, user feedback, and human-AI collaboration.

This approach firstly emphasises designing AI systems that are intuitive, usable, and responsive to human input and preferences. The socio-technical idea of symbiosis is cross-cutting and can be applied in all fields of complex human decision-making, such as the medical, legal (judicial and advisory), or technical-engineering domains. These are fields where automatically replacing humans is ethically and legally unacceptable or highly risky. Nor would it be acceptable for the idea that human decision-making is solely the result of artificial influences.

Evaluating a socio-technical system is challenging and requires procedures that allow adequate control and can convey human rights, even new ones. Hence, this conceptual and procedural framework is functional for developing a model to assess the impact of symbiosis on AI evolution and the corresponding ethical and legal risks, intending to mitigate them through human procedural control [1]. The assessment method to be implemented in the current research project includes both quantitative and qualitative criteria. This enables a flexible analysis and the evaluation of the ethical and legal acceptability and robustness in relation to the design and implementation of SAI systems in different contexts.

In this sense, categorising the symbiosis level is central since it could measure the human ability to oversee decision-making and would be a parameter and a basic principle to ascertain the compatibility of AI systems with the protection of rights.

The institutional policy-making would recommend this model within the AI regulatory sandboxes. The methodology we are modelling follows a 'lab-to-field' approach, thus considering all stages of the lifecycle development of a SAI system, in line with the 'by-design,' 'in-design,' and 'for designers' principles. In this way, it could be provided with a tool, albeit still theoretical and subject to validation in subsequent testing cycles, to address the ethical and legal foundational challenges associated with the paradigm shift resulting from the increasing symbiotic spreading of AI into ever more hybrid scenarios. It could determine the following impact factors: develop measurable indicators of the

human-centred and human-centric assessment, create evaluation scales, and set up the automation stage and procedures for the operationalisation of the evaluation method.

Besides, the method could have the following application:

1. Provide guidelines and procedures for those designing SAI systems, fostering a multidisciplinary mindset towards reliability. Machine reliability increases if, along with their technical robustness, practices in (co-)designing, development, validation, and responsible use also improve.

2. Offer an eco-systemic view of the AI world, repositioning companies' interests within an extended chain of design processes (ALTAI), validation (audits, external advisory, etc.), and sharing responsibilities (innovative approaches in data processing).

3. Develop a systematic understanding of AI ethics to move beyond simple moral judgments of what is 'right' or 'wrong' and instead promote the formalisation of new knowledge and skills essential for better application of national and international AI legislation—highlighting the potential importance of an 'Ethics and Legal Advisor' or 'Ethics and Legal Officer' in implementing the EU Regulation on AI.

4. Provide international standardisation agencies (ISO, CEN-CENELEC) with elements for formalising ethics and legal applied to AI in terms of certification to incorporate European legal and ethical values 'by design concretely.'

The model is still theoretical and requires operationalisation steps to be practically implemented through technical solutions after prototype design, compliance tests and final validation.

Acknowledgements

References

[1] A. Carnevale, A., Lombardi, F. A. Lisi, Exploring Ethical and Conceptual Foundations of Human-Centred Symbiosis with Artificial Intelligence, in: Proceedings of the 2nd Workshop on Bias, Ethical AI, Explainability and the Role of Logic and Logic Programming, BEWARE '23, Rome Italy, 2023, pp. 30–43.

[2] J. Sapp, Evolution by Association. A History of Symbiosis, Oxford University Press, Oxford, 1994.

[3] L. Sagan/Margulis, On the origin of mitosing cells, Journal of Theoretical Biology 14 (1967) 225–274.

[4] J. C. R. Licklider, Man-Computer Symbiosis, IRE Transactions on Human Factors in Electronics 1 (1960) 4–11. doi: 10.1109/THFE2.1960.4503259.

[5] P. Kotipalli, Symbiotic Artificial Intelligence, 2019. URL: https://p13i.io/posts/2019/06/symbiotic-ai/

[6] J. P. Ponda, Artificial Intelligence (AI) through Symbiosis, Undergraduate Thesis, Georgia Institute of Technology (GT), Atlanta, US, 2022.

[7] R. Hao, D. Liu, L. Hu, Enhancing Human Capabilities through Symbiotic Artificial Intelligence with Shared Sensory Experiences, arXiv:2305.19278v1 [cs.HC], Cornell University, New York, NY, 2023, 1–21. URL: http://doi.org/10.48550/arXiv.2305.19278. doi: 10.48550/arXiv.2305.19278.

[8] Z. M. Chng et al., Symbiotic Artificial Intelligence: Order Picking and Ambient Sensing, in: IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW), Rhodes Island, Greece, 2023, pp. 1–5. doi: 10.1109/ICASSPW59220.2023.10193633.

[9] R. Saracco, R. Madhavan, S. Mason Dambrot, D. de Kerchove, T. Coughlin, Symbiotic Autonomous Systems, White Paper, IEEE Digital Reality, 2017.

[10] S. Mason Dambrot, D. de Kerchove, F. Flammini, W. Kinsner, L. MacDonald Glenn, R. Saracco, Symbiotic Autonomous Systems, White Paper II, IEEE Digital Reality, 2018.

[11] 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, White Paper III, IEEE Digital Reality, 2019.

[12] N. Cristianini, La scorciatoia, Il Mulino, Bologna, 2023.

[13] N. Cristianini, Machina Sapiens, Il Mulino, Bologna, 2024.

[14] H. S. Sætra, The Parasitic Nature of Social AI: Sharing Minds with the Mindless, Integr. psych. Behave 54 (2020) 308–326. doi: 10.1007/s12124-020-09523-6.

[15] D. J. Chalmers, Reality+: Virtual Worlds and the Problems of Philosophy, W. W. Norton, New York, 2022.

[16] D. Asotić, NPC - The Philosophy of a Non-Playable Character: Tracing the Evolution and Impact of Non-Player Entities in Digital Culture, Independently published, 2023.

[17] P. Scriven, A Social Phenomenology of Non-Player Characters (NPCs) in Videogames, Techné Research in Philosophy and Technology 27 (2023) 240–259.

[18] E. Shein, Filtering for beauty, Communications of the ACM 64 (2021) 17–19. doi: 10.1145/3484997.

[19] W. Pendergrass, Artificial intelligence and its potential harm through the use of generative adversarial network image filters on TikTok, Issues In Information Systems 24 (2023) 113–127. doi: 10.48009/1_iis_2023_110.

[20] J. A. Harriger, J. K. Thompson, M. Tiggemann, TikTok, TikTok, the time is now: Future directions in social media and body image, Body Image 44 (2023) 222–226. doi: 10.1016/j.bodyim.2023.01.005.

[21] G. Sharp, Y. Gerrard, The body image "problem" on social media: Novel directions for the field. Body Image 41 (2022) 267–271. doi: 10.1016/j.bodyim.2022.03.004.

[22] A. Carnevale, C. Delgado Falchi, P. Bisconti, Hybrid Ethics for Generative AI: Some Philosophical Inquiries on GANs, HUMANA.MENTE Journal of Philosophical Studies 16 (2023) 33–56. URL: https://www.humanamente.eu/index.php/HM/article/view/434

[23] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, A. A. Bharath, Generative Adversarial Networks: An Overview, IEEE Signal Processing Magazine 35 (2018) 53–65. doi: 10.1109/MSP.2017.2765202.

[24] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, Communications of the ACM 63 (2020) 139–144. doi: 10.1145/3422622.

[25] M. Madary, T. K. Metzinger, Recommendations for Good Scientific Practice and the Consumers of VR-Technology, Frontiers in Robotics and AI 3 (2016). doi: 10.3389/frobt.2016.00003.

[26] F. Russo, Techno-scientific practices: An informational approach, Rowman & Littlefield, Lanham, 2022.

[27] B. Mittelstadt, Principles alone cannot guarantee ethical AI, Nature Machine Intelligence 1 (2019) 501–507. doi: 10.1038/s42256-019-0114-4.

[28] J. Morley, L. Floridi, L. Kinsey, 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 (2020) 2141–2168. doi: 10.1007/s11948-019-00165-52.

[29] K. Barad, Diffracting Diffraction: Cutting Together-Apart, Parallax 20 (2014) 168–187. doi: 10.1080/13534645.2014.927623.

[30] P. Bisconti, A. Carnevale, Alienation and Recognition: The Δ Phenomenology of the Human–Social Robot Interaction (HSRI), Techné: Research in Philosophy and Technology 26 (2022) 147–171. doi: 10.5840/techne202259157.

[31] B. Latour, Reassembling the social: An introduction to actor-network-theory, Oxford University Press, Oxford, 2007.

[32] J. Habermas, Faktizität und Geltung. Beiträge zur Diskurstheorie des Rechts und des demokratischen Rechtsstaats, Suhrkamp, Frankfurt am Main, 1992.

[33] G. Zagrebelsky, La legge e la sua giustizia, Il Mulino, Bologna, 2008.

[34] R. Alexy, Theorie der Grundrechte, Nomos Verlagsgesellschaft, Baden-Baden, 1985.

[35] L. L. Fuller, The Morality of Law, Yale University Press, New Haven, 1969.

[36] P. Marra, Per una moralità procedurale del diritto. Considerazioni attuali a partire da Lon L. Fuller, Cacucci, Bari, 2022.

[37] P. Marra, I. Galatola, Effectiveness as Threat to Constitutional Systems, in: J. Cremades, C. Hermida (Eds.), Encyclopedia of Contemporary Constitutionalism, Springer, Cham, Chapter 142-1. doi: 10.1007/978-3-319-31739-7_142-1.

[38] E. G. LaGratta, Procedural Justice: practical tips for courts, Center for Court Innovation, 2015. URL: https://www.innovatingjustice.org/publications/procedural-justice-practical-tips-courts

[39] B. MacKenzie, The Judge Is the Key Component: The Importance of Procedural Fairness in Drug-Treatment Court (an AJA White Paper), CT. REV. 52 (2016) 8–34.

[40] P. Casey, K. Burke, S. Leben, Minding the Court: Enhancing the Decision-Making Process (an AJA White Paper), CT. REV. 49 (2013) 76–98.

[41] K. Burke, S. Leben, Procedural Fairness: A Key Ingredient in Public Satisfaction (an AJA White Paper), CT. REV. 44 (2007) 4–25.

[42] D. B. Rottman, Procedural Fairness as a Court Reform Agenda, CT. REV. 44 (2007) 32–35.

[43] T. R. Tyler, Procedural Justice and the Courts, CT. REV. 44 (2007) 26–31.

[44] G. Ubertis, Intelligenza artificiale, giustizia penale, controllo umano significativo, DPC-RT 4 (2020) 75–88.

[45] A. Simoncini, S. Suweis, Il cambio di paradigma nell'intelligenza artificiale e il suo impatto sul diritto costituzionale, Journal of Legal Philosophy 1 (2019) 87–109. doi: doi.org/10.4477/93368.

[46] J. Manyika, M. Chui, M. Miremadi, J. Bughin, K. George, P. Willmott, M. Dewhurst, A future that works: automation, employment, and productivity, Tech. rep., McKinsey Global Institute, New York, NY, 2017.

[47] S. Quattrocolo, Artificial Intelligence, Computational Modelling and Criminal Proceedings. A Framework for A European Legal Discussion, Springer, Cham, 2020. doi: 10.1007/978-3-030-52470-8.

[48] J. Grogger, S. Gupta, R. Ivandic, T. Kirchmaier, Comparing Conventional and Machine-Learning Approaches to Risk Assessment in Domestic Abuse Cases, Becker Friedman Institute for Economics, Working Paper No. 2021-01, Chicago, 2021. URL: doi.org/10.2139/ssrn.3760094

[49] F. Lagioia, R. Rovatti, G. Sartor, Algorithmic fairness through group parities? The case of COMPAS-SAPMOC, AI & Soc 38 (2023) 459–478. doi: 10.1007/s00146-022-01441-y.

[50] L. Barboni, A. von Hagen, S. Piñeyro, I. Senabre, Predictive validity of the structured assessment of violence risk in youth (SAVRY) on the recidivism of juvenile offenders: a systematic review, Psychology, Crime & Law 2023. doi: 10.1080/1068316X.2023.2214661.

[51] S. Karnouskos, Symbiosis with artificial intelligence via the prism of law, robots, and society, Artif Intell Law 30 (2022) 93–115. doi: 10.1007/s10506-021-09289-1

[52] D. M. Katz, Quantitative Legal Prediction – or – How I Learned to Stop Worrying and Start Preparing for the Data Driven Future of the Legal Services Industry, Emory Law Journal 62 (2013) 909–966.

[53] G. Ubertis, Processo penale telematico, intelligenza artificiale e costituzione, Cassazione penale 2 (2024) 439–450.

[54] S. Quattrocolo, Sui rapporti tra pena, prevenzione del reato e prova nell'era dei modelli computazionali psico-criminologici, TCRS 1 (2021) 257–283. doi: 10.7413/19705476048.

[55] L. Maldonado, Risk and need assessment tools e riforma del sistema sanzionatorio: strategie collaborative e nuove prospettive, in: G. Di Paolo, L. Pressacco (Eds.), Intelligenza artificiale e processo penale. Indagini, prove, giudizio, Editoriale Scientifica, Napoli, pp. 141–170.

[56] E. J. Latessa, B. Lovins, J. Lux, The Ohio Risk Assessment System, in: J. P. Singh, D. G. Kroner, J. S. Wormith, S. L. Desmarais, Z. Hamilton (Eds.), Handbook of Recidivism Risk/Needs Assessment Tools, Wiley Blackwell, Hoboken, NY, 2017, pp. 147–163. doi: 10.1002/9781119184256.ch7.

[57] A. Garapon, J. Lassègue, Justice digitale. Révolution graphique et rupture anthropologique, Presses Universitaires de France, Paris, 2018.

[58] J. Floyd, Lebensformen: Living Logic, in: C. Martin (Ed.), Language, Form(s) of Life, and Logic: Investigations after Wittgenstein, De Gruyter, Berlin-Boston, 2018, pp. 59–92. doi: 10.1515/9783110518283-004.

[59] J. Morley, L. Kinsey, A. Elhalal, F. Garcia, M. Ziosi, L. Floridi, Operationalising AI ethics: barriers, enablers and next steps, AI & Soc. 38 (2023) 411–423. doi: 10.1007/s00146-021-01308-8.

[60] C. Novelli, F. Casolari, A. Rotolo, M. Taddeo, L. Floridi, AI risk assessment: A scenario-based, proportional methodology for the AI act, DISO 13 (2024). doi: 10.1007/s44206-024-00095-1.

[61] A. Dyoub, S. Costantini, F. A. Lisi, Logic programming and machine ethics, in: Proceedings of the 36th International Conference on Logic Programming (Technical Communications), ICLP Technical Communications '20, (Technical Communications) UNICAL, Rende, CS, Italy, EPTCS 2020, vol. 325, pp. 6–17.

[62] A. Dyoub, S. Costantini, F. A. Lisi, Towards an ILP application in machine ethics., in: D. Kazakov, C. Erten (Eds.), Proceedings of the 29th International Conference on Inductive Logic Programming, ILP '19, vol. 11770 of Lecture Notes in Computer Science, Springer, Netherlands, 2019, pp. 26–35.

[63] A. Dyoub, S. Costantini, F. A. Lisi, Towards ethical machines via logic programming., in: B. Bogaerts, E. Erdem, P. Fodor, A. Formisano, G. Ianni, D. Inclezan, G. Vidal, A. Villanueva, M. D. Vos, F. Yang (Eds.), Proceedings of the 35th International Conference on Logic Programming (Technical Communications), ICLP '19 Technical Communications, Las Cruces, NM, USA, EPTCS 2019, vol. 306, pp. 333–339.

[64] A. Dyoub, S. Costantini, F. A. Lisi, Learning domain ethical principles from interactions with users, DISO 28 (2022). doi: 10.1007/s44206-022-00026-y.

[65] A. Dyoub, S. Costantini, I. Letteri, Care robots learning rules of ethical behavior under the supervision of an ethical teacher (short paper), in: P. Bruno, F. Calimeri, F. Cauteruccio, M. Maratea, G. Terracina, M. Vallati (Ets.), Joint Proceedings of the 1st International Workshop on HYbrid Models for Coupling Deductive and Inductive ReAsoning, HYDRA '22 and the 29th RCRA Workshop on Experimental Evaluation of Algorithms for Solving Problems with Combinatorial Explosion, RCRA '22, co-located with the 16th International Conference on Logic Programming and Non-monotonic Reasoning LPNMR '22, Genova Nervi, Italy, CEUR Workshop Proceedings, Germany, 2022, vol. 3281, pp. 1–8. URL: CEUR-WS.org.