Human empowerment in self-adaptive socio-technical systems

Nicolas Boltz
nicolas.boltz@kit.edu
Karlsruhe Institute of Technology
Germany

Sinem Getir Yaman
sinem.getir.yaman@york.ac.uk
University of York
United Kingdom

Paola Inverardi
paola.inverardi@gssi.it
Gran Sasso Science Institute
Italy

Rogério de Lemos
r.delemos@kent.ac.uk
University of Kent
United Kingdom

Dimitri Van Landuyt
dimitri.vanlanduyt@kuleuven.be
KU Leuven
Belgium

Andrea Zisman
andrea.zisman@open.ac.uk
The Open University
United Kingdom

ABSTRACT

Recent advances in generative AI and machine learning have stirred up fears about the unbridled adoption of autonomous, self-adaptive decision mechanisms in socio-technical systems. This vision paper explores the critical relationship between software-intensive systems and the empowerment of humans as individuals and society. We highlight the need for human empowerment within the context of self-adaptive socio-technical systems (SASTSs), which require mechanisms for balancing of diverse needs, values, and ethics on the individual, community, and societal levels. We propose an architecture comprised of *Connector* and *Mediator* elements, and *third-party auditing*, to support interactions and ensure preservation of human needs, values, and ethics. We use an example of Robot-Assisted A&E Triage system to motivate and illustrate our work and discuss some open challenges for future research.

ACM Reference Format:
Nicolas Boltz, Sinem Getir Yaman, Paola Inverardi, Rogério de Lemos, Dimitri Van Landuyt, and Andrea Zisman. 2024. Human empowerment in self-adaptive socio-technical systems. In *19th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS '24), April 15–16, 2024, Lisbon, AA, Portugal.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3643915.3644082

1 INTRODUCTION

The pervasive influence of technology in our daily lives and society cannot be overstated. In an era dominated by software-driven solutions, our actions and behaviors are increasingly shaped by the impact of digital systems, both positively and negatively.

In today's interconnected world, data flows freely across digital systems, transcending geographical and temporal boundaries, and, consequently, the behavior of individuals and society as a whole is linked to this pervasive connectivity. Within this software-intensive environment, it is imperative to consider the concept of human empowerment as a cornerstone in the design, engineering, and utilization of software systems. By *humans*, we encompass individuals, communities, and society at large. More specifically, we consider that humans' engagement in system interactions may be motivated by their specific and intrinsic needs, values, and ethical norms [5, 15, 26, 37, 38].

These needs, values, and ethical norms may range from deeply personal beliefs to broader societal principles, with some even being established due to regulations. This multi-tiered perspective acknowledges that nations, regions, and societies themselves constitute distinct types of communities; each with the authority to establish rules and regulations that reflect the ethical concerns shared by a broader societal consensus.

In [22] the notion of digital ethics as introduced by Floridi [15] is suggested as a way to help draw the line of system's autonomy with respect to human autonomy. Digital ethics encompasses two dimensions, namely: *hard ethics*, which are firmly rooted in legal frameworks and in established social norms (e.g., GDPR); and *soft ethics*, which encompass the moral preferences and ethical considerations of individuals and groups.

However, the march toward extensive automation, particularly in self-adaptive socio-technical systems (SASTSs), has raised concerns about the erosion of human autonomy. While humans, including moderators, decision-makers, data scientists, and operators, oversee these types of systems, three critical concerns have emerged:

- **Misaligned Objectives**: SASTSs may optimize goals that deviate from societal and human values, potentially leading to outcomes that conflict with hard ethics.
- **Algorithmic and Data Bias**: SASTSs may inadvertently perpetuate or amplify existing biases, potentially resulting in discriminatory or unfair outcomes [42].
- **Neglect of Human Values**: SASTSs may not consider and integrate soft ethics in their decision-making processes [11].

These concerns surrounding SASTS pose significant obstacles to their continued development, operation, evolution, and widespread acceptance. The introduction of regulatory measures, such as the GDPR and AI Act [1], reflects society's response to these concerns. However, these regulations often lag behind technological advancements, serving primarily as a reactive safeguard against excesses and problems that have already materialized. Furthermore, overbearing regulation can stifle innovation and hinder societal progress.

Our goal is to address the challenges associated with strengthening the role of human empowerment within SASTSs. We direct our attention to changes within systems and their environments, shifts

SEAMS '24, April 15–16, 2024, Lisbon, AA, Portugal

ACM ISBN 979-8-4007-0585-4/24/04.
https://doi.org/10.1145/3643915.3644082

[1] digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

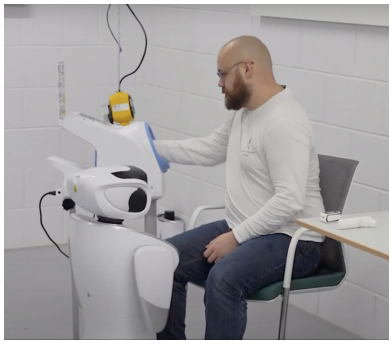

Figure 1: Patient-robot interaction in the triage system [34]

in soft ethics, and alterations in the regulatory landscape within which these systems operate (hard ethics).

The targeted problem poses several key challenges:
Complex Decision-Making: How can SASTSs navigate the complex landscape of diverse hard and soft ethics, while simultaneously achieving other critical system's objectives?
Capturing Stakeholder Goals: How can individuals effectively articulate their soft ethics in a manner that is understandable and actionable within the intricate decision-making processes of SASTSs?
Stakeholder Engagement: How can individuals remain informed about these processes of SASTSs, fostering transparency and awareness, and how can they exercise agency when necessary to safeguard their hard and soft ethics?
Evolution & Change: How can SASTSs adapt to the fluid nature of hard and soft ethics, and how can these systems evolve while preserving respect for these fundamental human aspects?

Motivating Example: To illustrate the significance of the above challenges, consider a real-world SASTS, the Diagnostic AI System for Robot-Assisted A&E Triage (DAiSY [34]). DAiSY is composed of a robot that assists humans (patients and clinicians) by collecting medically relevant information about a patient (e.g. the vital signs), and establishes the connection between humans and the planning system responsible for triage decisions, as shown in Figure 1. When patients arrive in the triage room of a hospital, they have the option to register themselves through the robot or to queue to be registered manually by a nurse. The robot can significantly expedite the process. However, the use of a robot creates concerns about soft and hard ethics, as described below:

- **Misaligned Objectives:** The assistive robots and triage system will collect sensitive patient information, yet it remains crucial to ensure that the requested information remains within boundaries and compatible with the purposes of the overall system (so as to not be construed as a form of surveillance or interrogation). Furthermore, the collected data should be handled securely and in compliance with *privacy* regulations (e.g., GDPR in Europe). *Transparency* in the robot's and triage system's decision-making processes is also essential to build *trust* and *accountability*. In addition, patients and clinicians must provide informed consent for any interaction or data collection involving the assistive robot and triage system. Ensuring that patients understand the robot's capabilities and purpose, and how their data will be used is essential to respect their autonomy and decision-making.
- **Algorithmic and Data Bias:** The algorithms and data used to train assistive robots and implement the triage system can introduce biases, which can result in unfair treatment of certain demographic groups [42]. Efforts should be made to identify and mitigate biases in decision-making processes to ensure *fairness* and *equity*.
- **Neglect of Human Values:** The triage system should be designed with *cultural sensitivity* in mind, respecting diverse cultural norms and practices. What is acceptable in one culture may not be in another, and robots and the triage system should be adaptive and respectful. Furthermore, triage situations can be emotionally charged, and both patients and clinicians may need emotional support. While robots can provide information and assistance, they lack the *emotional empathy* that humans can offer. Care must be taken not to dehumanize the patient experience.

This real-world application emphasizes the need for digital systems to transcend purely economic or technical objectives and incorporate social, legal, ethical, empathetic, and cultural norms [18, 45].

The remainder of this paper[2] is structured as follows. Section 2 outlines our vision, defines an architecture to empower humans, and argues how this architecture will lead to improved awareness and management of the diverse soft and hard ethics of humans.

2 VISION

We envision a scenario in which SASTSs and humans (individual, community, and society) are able to react to changes without requiring explicit awareness, formal approvals, or concerns about the consequences of evolution that may arise from these changes. In this complex scenario, it is essential to equip humans with the ability to deal with the consequences of dynamic changes. At the same time, it is essential to equip SASTSs with necessary mechanisms to address changes in soft and hard ethics.

A central element of our vision is to establish a boundary for the autonomy of SASTSs. This includes a clear separation between decisions that SASTSs can make autonomously and those that can be adapted or negotiated in response to human soft and hard ethics during interactions with individuals. A SASTS must adhere to the principles of hard ethics while accommodating the subtleties of soft ethics. This adaptation allows the system to facilitate decisions that involve ethical considerations.

However, humans interacting with SASTSs have soft ethics that serve as an expression of their values. It is necessary to create an environment that supports the expression of these soft ethics and allows them to influence the behavior of the system in a way that is consistent with the ethical orientation of the humans. Considering our example of the DAiSY system, a compassionate customer should be able to voluntarily give up his or her assigned triage priority in favor of others, thereby influencing the DAiSY system's default soft ethics.

Our vision aims to support humans interacting with SASTSs in an environment characterized by dynamic changes and inherent

[2]This is an extended version of the initial concept presented in Chapter 9 of the research agenda of the 2023 Bertinoro Meeting on Uncertainty in Self-Adaptive Systems [46].

uncertainties. In order to achieve this, we propose an architecture that empowers humans and distributes responsibility by including two main architectural elements between the Human and SASTS, namely, *Connector* and *Mediator*.

Human The *Human* that interacts with the *SASTS* may refer to an individual human stakeholder, but also to communities consisting of multiple individuals, which in turn can be informal, semi-formal, and society.

Connector An architectural element close to the *Human*. The *Connector* manages and represents the soft ethics of *Humans* and serves as an intermediary on their behalf when interacting with the *SASTS*. This element needs to be controlled entirely by the *Human*. It engages in negotiation and communication with the *SASTS* on the *Human*'s behalf.

Mediator The *Mediator* is an architectural element placed in between the *Connector* and the *SASTS*. It checks hard ethics and makes complex trade-offs. The *Mediator* engages in negotiations with the *SASTS* and *Connector* to achieve a trade-off. The negotiations and trade-offs take into account the *Human*'s soft ethics, the given hard ethics, and the requirements of the *SASTS* for providing a service requested by the *Human*. In our vision, multiple *Mediators*, each specialized in checking different hard ethics could co-exist and work together. Depending on the context of the *SASTS*, different *Mediators* can be combined. Changes in the *SASTS* could entail a change within a *Mediator* itself or the set of Mediators that are combined.

Third Party Auditor The *Third Party Auditor* is an element that is not directly involved in the communication between the *Human* and the *SASTS*. The task of the *Third Party Auditor* is to check whether the elements of our proposed architecture adhere to the given ethics and generally operate as defined. Specific focus is placed on the monitoring of the *Mediator*, as this element ensures general functionality and compliance with hard ethics, even if the system does not fully do so itself. However, the *SASTS* also needs to be audited, e.g., as it is currently done for GDPR compliance with data protection impact assessments (DPIA). The role of *Third Party Auditor* should be performed by a neutral institution (e.g. a body authorized by the state).

Human Models Models of the *Human* are created and maintained by the *Connector* and Mediator. The models contain information about the *Human*, including information about their soft ethics. Furthermore, our vision entails the creation of particular models for each unique system, including Connectors and Mediators. The *SASTS* does not store Human Models, instead, it engages with the Mediator to acquire information about and soft ethics of the *Human*. Depending on the implementation, the *Mediator* could also use (freely) available *Human Models* as information when checking compliance regarding hard ethics, e.g. whether data of the human can be considered anonymous or not.

SASTS A self-adaptive socio-technical system (SASTS) with which the Human wants to interact. The *SASTS* provides the *Human* with services or information. *SASTSs* might change and evolve depending on its current environment, context, and general business needs.

As shown in Figure 2, when interacting with *SASTSs*, the *Human* only interacts with the respective *Connector*. The *Connector* in turn interacts with the *Mediator*, which interacts with the *SASTS*. Through our envisioned architecture, the responsibility of aligning with the diverse ethics of humans is distributed across different architectural elements.

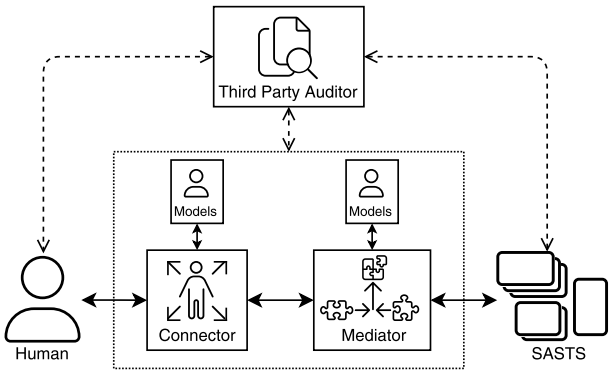

Figure 2: Vision of human empowerment with regards to systems.

In cases where the *SASTS* changes, the *Mediator* has to react and adapt to ensure compliance with these changes and appropriately negotiate with relevant *Connectors*. Changes in the *Mediator* might also entail adaptation of the *Connector*, as the *Mediator* might require previously undefined information or decisions of the *Human*. In case of changes, the *Connector* might require the *Human* to provide some input regarding characteristics of his/her individual values. Changes in hard ethics (e.g., new regulations) require the *Mediator*(s) to adapt accordingly. Such changes may require the adaptation to new interfaces or forms of negotiation. Changes propagate throughout all elements of our envisioned architecture and each element has to be capable of reactive self-adaptation.

In the following, we discuss how the proposed architecture addresses the four challenges of our problem statement (discussed above in Section 1):

- **Complex decision-making:** The *Mediator* plays a central role in decision-making as it actively tries to reconcile different and sometimes competing stakeholder goals. Access to the *Human Models* (which are expressions of soft and hard ethics) is essential to correctly assess acceptability and impact at the level of the *Humans* affected. The *Mediator* also actively orchestrates the interactions between the *Connectors* and the *Humans* in cases where additional information is required, bringing the human (indirectly) in the loop in the decision-making.
- **Capturing Stakeholder Goals:** The *Connector* is the main interface used by the *Humans* for expressing their stakeholder goals (soft ethics). These are reflected and persisted in the *Human Models* which can be consulted and taken into account during decision-making.
- **Stakeholder engagement:** As discussed above, the *Connector* is actively used by the *Human* to express ethics, but also to provide input and feedback upon request by the SASTS, for example when initiated by *Mediators*. In addition, the *Connector* provides access to decisional outcomes and explanations. Finally,

the *Connector* also provides means of intervention and control to the human stakeholder.

- **Evolution & change:** The *Connector* is responsible for co-evolving the *Human Models* with the human ethics. It monitors and keeps track of deviations between the *Human* and its model representation, and when this happens, an actualization activity is performed. The direct feedback provided by and interventions performed by *Humans* might be additional indicators of evolution. Independently, the *Third Party Auditor* assumes a similar role, evaluating at a higher level the overall system outcomes and evolution (dotted arrow on the right in Figure 2), taking heed of the soft and hard ethics of the affected *Humans* (dotted arrows at the center and on the left in Figure 2).

3 EXEMPLARY APPLICATION

To showcase the general applicability of our vision outlined in Section 2, we apply our architecture to the motivating example described in Section 1.

As shown in Figure 3, the assistive robot of DAiSY acts as an intermediary element between the humans and the *Triage System*. For our example, we focus on two types of humans interacting with DAiSY: *Patients* and *Clinicians*. In the case of triage decisions, a neutral *Ethics Committee* takes the role of Third Party Auditor. The *Patient* uses the robot to register with the *Triage System* and is guided through different automated examinations. Furthermore, the *Clinician* interacts with the *Triage System* as necessary, and upon the patient's admission, the clinician steps in to deliver care based on the *Patient*'s triage priority.

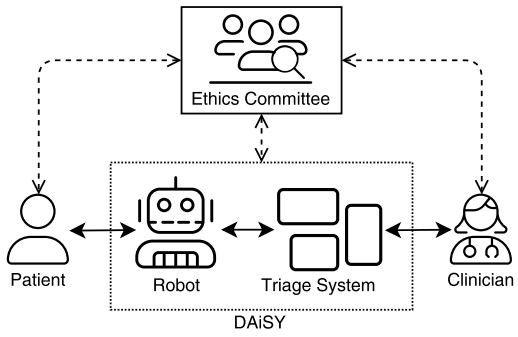

Figure 3: High-level architecture of DAiSY.

Figure 4 shows a more detailed description of the instantiation of our architecture in DAiSY, focusing on the empowerment of the *Patient* and *Clinician*. In this example, the robot is augmented with a *Patient Connector*, which empowers the patient to establish their own ethical choices which cover aspects of the triage decision-making process and also directives related to data privacy. The *Clinician Connector* enables the *Clinician* to establish their unique ethical guidelines. These guidelines can pertain to work-related matters, such as ethical considerations regarding overtime, or how productivity indicators are gathered and utilized for assessment or comparison with peers. These individualized ethical standards are stored in both the *Patient Model* and the *Clinician Model*.

The *Triage Mediator* communicates with the *Triage Back-end*, *Patient* and *Clinician Connector*. It ensures that the triage system adheres to the hard ethics (e.g., the applicable legal framework) and tries to make trade-off decisions based on the soft ethics that the connectors communicate to the *Triage Mediator*. An exemplary trade-off decision could involve a *Patient* willingly lowering their priority to favor elderly patients with a similar or lower priority. To make these kinds of trade-off decisions, the *Triage Mediator* might use other *Patient Models*, as shown in Figure 4.

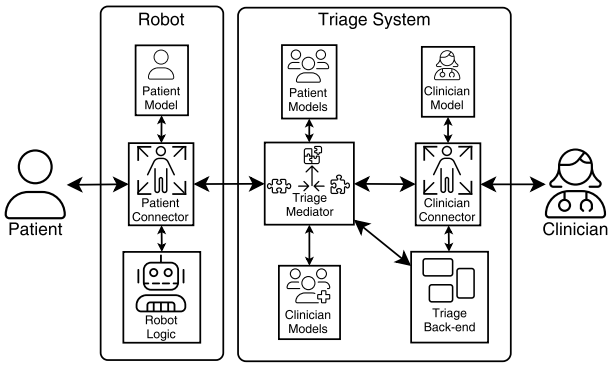

Figure 4: Illustration of the introduction of connectors and mediators in DAiSY.

The application of our architecture addresses the concerns stated in the motivating example in Section 1. The *Triage Mediator* plays a pivotal role in mitigating the concerns regarding the misaligned objectives and neglect of human soft ethics by enforcing hard ethics regarding triage. In the case of our exemplary DAiSY system, these hard ethics include the legal framework regarding triage situations, but also data privacy and security of the sensitive patient and clinician data. To ensure equitable allocation of scarce medical resources, including clinicians, the *Triage Mediator* can utilize the *Patient Models* and *Clinician Models* to examine and, if needed, engage in automated negotiations with the triage System.

The *Patient* and *Clinician Connector*, serving as the interface between the Human and the *Triage System*, also empower humans to express their preferences regarding soft ethics. The *Connectors* also serve to mitigate concerns regarding bias and the neglect of human soft ethics, by allowing the human to recognize that there is a bias through the Connector. The Connector could use the gained knowledge to trigger adaptation in the SASTS. In our example, the *Patient Connector*, being close to the patient helps to reduce the necessity of having to infer or learn specific details, such as demographic information from generic models. This is because the *Patient Connector* supplies all the required decisions made by the patient concerning their unique ethical guidelines. Normally, these would need to be learned by DAiSY through profiling. In addition, the connectors –being a close reflection of the soft ethics of the patients– can also ensure a certain degree of cultural sensitivity. The *Ethics Committee* as the role of Third Party Auditor also helps in mitigating all three concerns, by ensuring correct behavior of all architecture elements regarding triage decisions.

4 RELATED WORK

We investigate the state of the art in three categories in the following subsections.

4.1 Architecture Elements

To support the interaction between humans and SASTSs, the proposed architecture relies on two architectural elements: software connectors and mediators. In software architecture, connectors are architectural elements that model interactions among components and rules that govern such interactions. They are a design mean to conceptually separate a system's interactions from computations. Connectors can assume different roles depending on the logic that they embed to manage interactions: communication, coordination, adaptation, and mediation [43].

The latter are mediating connectors or mediators. The mediator concept was initially introduced to cope with the integration of heterogeneous data sources [48] and as design pattern. It was proposed to deal with components's protocol mismatches [23, 50] and in the field of software architecture as ad-hoc wrappers to address communication problems [41]. Mediators and automated mediation were proposed within the Web Services and Semantic Web contexts [28] as well as in ubiquitous environments to cope with components' behavioral diversities at run-time [6, 40].

Underlying the concept of mediator is the need to represent a centralized unit of control that can manage interactions to and from multiple sources. The idea of having connectors that are able to self-adapt at run-time for enabling functional and non-functional interoperability has been looked into by automating the synthesis of connectors either at the middleware level [31] or application level [13, 29].

More recently, connectors have been proposed as ethical mediators between the autonomous system and its users able to intercept their interactions and prevent/adapt/modify system behaviors that are not admissible by the user's ethical preferences [3].

4.2 Human in the Loop

Socio-technical systems (STSs) consist of human, hardware, and software agents that work in tandem to fulfill stakeholder requirements [39]. Self-adaptation comes into play when the agents that comprise an STS must continuously adapt their behaviors to take into account risks and opportunities that arise at run-time [32]. It was Peng et al. [32] that originally looked into this problem, and proposed a decision-theoretic self-adaptation framework that incorporates how changes should be handled when agents reconsider and renegotiate their commitments and plans with other agents. In the context of socio-cyber-physical systems (SCPSs), which are similar in concept to STSs, Calinescu et al. [9] identified as a challenge the incorporation of self-adaptive capabilities into SCPSs, in particular, the leveraging of human-interpretable input. As indicated in that paper, machine learning may provide a solution to that challenge, but before that, we need to understand how humans can be empowered when interacting with these types of systems.

Whether it is human on or in the loop, when framing the interaction of humans with self-adaptive systems (SASs), several approaches have addressed human empowerment in the process of decision-making [19]. One approach uses explanations to be provided to humans, which should be part of another feedback layer residing on top of the classical MAPE-K loop, for them to steer decision-making [30]. A similar approach relies on probabilistic model checking for providing explanations associated with adaptations, for supporting the human to select the appropriate adaptation [24]. A different approach to support decision-making combines digital twins for representing an SAS, and virtual reality for supporting an immersive and realistic human involvement in the self-adaptation loop [51]. A more intrusive approach considers a co-adaptive solution to operationalize the neural input into software systems [25].

4.3 Values and Ethics

Several practitioners and approaches have been advocating the need to align technologies with human values [15, 27, 47], norms, and ethics [35]. Some of these approaches are concerned with the different types of human values. Examples are found in [36] in which 36 different types of human values classified as instrumental and terminal have been proposed; in [37] where ten universal value categories known as the Theory as Basic values are suggested; and in [26] in which values are seen as mental representations that can be studied at system, abstract, and instantiation levels.

Other approaches have been proposed that study values in human-computer interaction [10] and in software engineering [7, 49]. Social values have been used in software design patterns [21], and the impact of values has also been measured in requirements engineering activities [33]. Moreover, in value-based requirements engineering, values are considered as personal attitudes and beliefs that influence requirements [44]. Values can also be treated as soft-goals or non-functional requirements [4]. Some approaches promote the need to reflect on values before attempting to operationalize them [14] and suggest a ranking mechanism to elicit values. Value-sensitive design [16] has also been used as an approach to identify values of ethical importance using scenarios and storyboarding. In [38] the authors present a comprehensive survey about considering values during the development of software systems.

Recent attempts were directed toward the modeling of users's soft ethics based on surveys. In [2] the survey is based on the correlations between ethics positions (idealism and relativism), personality traits (honesty/humility, conscientiousness, Machiavellianism and narcissism), and worldview (normativism), and then by using a clustering approach to create ethical profiles predictive of user's digital behaviors concerning privacy violation, copyright infringements, caution, and protection. In [1] a new methodology directed to gather data on the moral preferences of users was presented that takes the form of a questionnaire based on real-life scenarios in which the user's decision has a moral impact. The new questionnaire captures the choices that people make and their underlying motivations according to the agents themselves.

Existing approaches support value operationalization in the early stages of software development. An exception is found in Values@Runtime [5], in which the authors advocate the use of software to help users articulate, measure, and reflect on their values at run time, following the views that users better understand their values as they experience, reflect, and learn more about them [17]. The work described in this paper complements this view by providing an architecture in which users are empowered with respect

to their values and ethical considerations when interacting with socio-technical systems.

5 OPEN CHALLENGES

While our architectural vision successfully tackles some of the initial challenges we identified (as discussed in Section 2), we now encounter a new set of ongoing challenges. We group these new challenges into four different areas:

Realization of Architectural Elements. An obvious challenge lies in the realization of the proposed architecture and elements. Questions regarding the deployment, ownership, and operation of the Connector and Mediator exist in practice and will require interdisciplinary cooperation e.g. with legal experts. Clear interfaces need to be defined for Connectors and Mediators. The potential communication overhead, added by the Connector and Mediator needs to be feasible for different application scenarios. Also, each architectural element, including the SASTS, needs to provide a way for the Third Party Auditor to effectively assess and audit.

Expressing Ethics. We identify the operationalization of soft and hard ethics to be a significant challenge [5, 38]. Despite some advances in the literature (e.g. [18, 45]), the operationalization of soft ethics, as defined in [38] — '*the process of identifying human values and translating them to accessible and concrete concepts so that they can be implemented, validated, verified, and measured in software*', is still an open problem both during design time and run time of the systems [5, 45]. There are difficulties in modeling soft and hard ethics and accounting for context-awareness. Additionally, we see, that soft and hard ethics need to be made tangible, ensuring that all proposed architectural elements, including the SASTS and Human, understand them and can adapt according to their changes. Moreover, legal change, the evolution of regulations, and the representation of these hard ethics in a manner that is comprehensive to the SASTS need to be considered [8, 20].

Empowering Humans. Regarding humans, empowering individuals, groups, and society is a challenge [2, 3]. In addition to the current lack of means to empower people, there is uncertainty about the effectiveness of future methods and whether they will actually be utilized. Furthermore, we see that when providing means to empower humans, potential opportunities for abusing the system arise. This in turn produces the challenge of dealing with potential adversaries. It is necessary to support negotiation between humans and the SASTSs when interacting.

Negotiations and Trade-offs. In terms of the realization of our vision described in Section 2, the development and operationalization of the SASTSs, Connectors, and Mediators and their respective models, remain open challenges. Negotiation protocols that handle the previously described challenges need to be established. The negotiation protocols have to consider transactional properties of negotiation outcomes and enable re-negotiation in response to changes in the human, the SASTS, or the environment. In this process, the SASTS should be able to inform humans about its changes and be aware of possible reluctance. Additionally, it is crucial that these negotiation protocols can deal with conflicts that may arise due to incompatibilities between a human's soft and hard ethics and the SASTS's goals. Balancing the SASTS's involvement with Connectors and Mediators is also challenging. We highlight that there should always be a fallback policy that may involve minimal SASTS involvement or even require blocking it altogether. Therefore, the SASTS should remain open, aware, and adaptive to Mediators and not adopt a "take-it-or-leave-it" mentality when human soft and hard ethics and SASTS goals do not perfectly align.

6 CONCLUSION

We have proposed a vision for human empowerment in the development of self-adaptive socio-technical systems (SASTSs), which takes heed of the soft and hard ethics of humans – at the level of individuals, communities, and society at large.

Conway's law states that the fundamental structures of a system will typically mimic or reflect the structure of the organization that develops it [12]. In that sense, it should not be entirely surprising that the main structure of the proposed architecture likewise reflects some of the best practices in how a democratic society is organized. Citizens have been given rights and freedoms and means to enact these rights and freedoms in the form of *Connectors*. Society at large is constantly making non-trivial trade-off decisions between different and competing goals, carefully balancing the interests of all involved actors. In the proposed architecture, this role is played by the *Mediators*. Access to and awareness of the soft and hard ethics of the different involved and affected humans is essential to proper and informed decision-making – the reification of these are the *Human Models* in our proposed architecture. Finally, the overall functioning is monitored and audited by governmental and non-governmental organizations such as ethics committees, watchdogs, etc – these are the *Third-Party Auditors*.

The adoption of this architecture has a number of important preconditions and consequences: (i) by exclusively relying on *Connectors*, the involved humans should be able to come to an understanding about decisions and outcomes, about how their specific soft and hard ethics were taken into account (informed participants), and about how they themselves may contribute to more optimal overall outcomes (human-in-the-loop), (ii) the *Human Models* are not static and *Connectors* should constantly verify that these model representations are still in line with reality (co-evolution of human models), (iii) the decision-making should be proactively transparent and open to the *Third-Party Auditors* and impactful decisions should be explainable and independently verifiable (transparency and openness).

ACKNOWLEDGMENTS

We would like to thank the organizers and attendees of the 2023 Bertinoro Seminar on Uncertainty in Self-Adaptive Systems for the inspiring discussions that resulted in this publication.

This work was supported by the research project SofDCar (19S21002), which is funded by the German Federal Ministry for Economic Affairs and Climate Action, the EPSRC project EP/V026747/1 'UKRI Trustworthy Autonomous Systems Node in Resilience', the PRIN project 2022JKA4SL - HALO: etHical-aware AdjustabLe autOnomous systems, the MUR (Italy) Department of Excellence 2023 - 2027 for GSSI, the Research Fund KU Leuven, the H2020 ERATOSTHENES project (Grant Nb. 101020416), and the EPSRC Platform Grant on Secure Adaptable Usable Software Engineering (EP/R013144/1).

REFERENCES

[1] Costanza Alfieri, Donatella Donati, Simone Gozzano, Lorenzo Greco, and Marco Segala. 2023. Ethical Preferences in the Digital World: The EXOSOUL Questionnaire. In *International Conference on Hybrid Human-Artificial Intelligence (HHAI)*. 290–299. https://doi.org/10.3233/FAIA230092

[2] Costanza Alfieri, Paola Inverardi, Patrizio Migliarini, and Massimiliano Palmiero. 2022. Exosoul: ethical profiling in the digital world. In *International Conference on Hybrid Human-Artificial Intelligence (HHAI)*. https://doi.org/10.48550/arXiv.2204.01588

[3] Marco Autili, Davide Di Ruscio, Paola Inverardi, Patrizio Pelliccione, and Massimo Tivoli. 2019. A Software Exoskeleton to Protect and Support Citizen's Ethics and Privacy in the Digital World. *IEEE Access* 7 (2019), 62011–62021.

[4] Balbir S Barn. 2016. Do you own a Volkswagen? Values as non-functional requirements. In *Human-Centered and Error-Resilient Systems Development*. 151–162. https://doi.org/10.1007/978-3-319-44902-9_10

[5] Amel Bennaceur, Diane Hassett, Bashar Nuseibeh, and Andrea Zisman. 2023. Values@Runtime: An Adaptive Framework for Operationalising Values. In *International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*. https://doi.org/10.1109/ICSE-SEIS58686.2023.00024

[6] Amel Bennaceur and Valérie Issarny. 2015. Automated Synthesis of Mediators to Support Component Interoperability. *IEEE Trans. Software Eng.* 41, 3 (2015), 221–240. https://doi.org/10.1109/TSE.2014.2364844

[7] Marcello M. Bersani, Matteo Camilli, Livia Lestingi, Raffaela Mirandola, Matteo Rossi, and Patrizia Scandurra. 2023. Architecting Explainable Service Robots. In *Software Architecture*. 153–169. https://doi.org/10.1007/978-3-031-42592-9_11

[8] Nicolas Boltz, Leonie Sterz, Christopher Gerking, and Oliver Raabe. 2022. A Model-Based Framework for Simplified Collaboration of Legal and Software Experts in Data Protection Assessments. *INFORMATIK 2022* (2022).

[9] Radu Calinescu, Javier Cámara, and Colin Paterson. 2019. Socio-Cyber-Physical Systems: Models, Opportunities, Open Challenges. In *International Workshop on Software Engineering for Smart Cyber-Physical Systems (SEsCPS)*. 2–6. https://doi.org/10.1109/SEsCPS.2019.00008

[10] G Cockton. 2004. Value-centric HCI. In *NordHCI*.

[11] European Commission. 2018. European Group on Ethics in Science and New Technologies, Statement on artificial intelligence, robotics and 'autonomous' systems. *Publications Office* (2018). https://data.europa.eu/doi/10.2777/531856

[12] Melvin E Conway. 1968. How do committees invent. *Datamation* (1968), 28–31.

[13] Antinisca Di Marco, Paola Inverardi, and Romina Spalazzese. 2013. Synthesizing self-adaptive connectors meeting functional and performance concerns. In *International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS)*. 133–142. https://doi.org/10.1109/SEAMS.2013.6595500

[14] Maria Angela Ferrario, Will Simm, Stephen Forshaw, Adrian Gradinar, Marcia Tavares Smith, and Ian Smith. 2016. Values-first SE: research principles in practice. In *International Conference on Software Engineering Companion (ICSE-C)*. 553–562. https://doi.org/10.1145/2889160.2889219

[15] Luciano Floridi. 2018. Soft Ethics and the Governance of the Digital. *Philosophy & Technology* 31 (2018), 1–8. https://doi.org/10.1007/s13347-018-0303-9

[16] Batya Friedman. 1996. Value-sensitive design. *interactions* 3, 6 (1996), 16–23.

[17] Mary C Gentile. 2010. *Giving voice to values*. Yale University Press.

[18] Sinem Getir Yaman, Charlie Burholt, Maddie Jones, Radu Calinescu, and Ana Cavalcanti. 2023. Specification and Validation of Normative Rules for Autonomous Agents. In *Fundamental Approaches to Software Engineering (FASE)*. 241–248. https://doi.org/10.1007/978-3-031-30826-0_13

[19] Miriam Gil, Vicente Pelechano, Joan Fons, and Manoli Albert. 2016. Designing the Human in the Loop of Self-Adaptive Systems. In *Ubiquitous Computing and Ambient Intelligence*. 437–449. https://doi.org/10.1007/978-3-319-48746-5_45

[20] Clement Guitton, Aurelia Tamò-Larrieux, and Simon Mayer. 2022. Mapping the Issues of Automated Legal Systems: Why Worry About Automatically Processable Regulation? *Artifcial Intelligence and Law* (2022).

[21] W Hussain, D Mougouei, and J Whittle. 2018. Integrating social values into software design patterns. In *International Workshop on Software Fairness*. https://doi.org/10.1145/3194770.3194777

[22] Paola Inverardi. 2019. The European Perspective on Responsible Computing. *Commun. ACM* 62, 4 (2019), 64. https://doi.org/10.1145/3311783

[23] Paola Inverardi and Massimo Tivoli. 2001. Automatic synthesis of deadlock free connectors for COM/DCOM applications. In *European Software Engineering Conference (ESEC)*. 121–131. https://doi.org/10.1145/503209.503227

[24] Nianyu Li, Javier Cámara, David Garlan, and Bradley Schmerl. 2020. Reasoning about When to Provide Explanation for Human-involved Self-Adaptive Systems. In *International Conference on Autonomic Computing and Self-Organizing Systems (ACSOS)*. 195–204. https://doi.org/10.1109/ACSOS49614.2020.00042

[25] Eric Lloyd, Shihong Huang, and Emmanuelle Tognoli. 2017. Improving Human-in-the-Loop Adaptive Systems Using Brain-Computer Interaction. In *International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS)*. 163–174. https://doi.org/10.1109/SEAMS.2017.1

[26] Gregory R Maio. 2016. *The psychology of human values*. Routledge.

[27] D. Mougouei, H. Perera, W. Hussain, R. Shams, and J. Whittle. 2018. Operationalizing human values in software: A research roadmap. In *Joint Meeting on European Soft. Engineering Conference and Symp. on the Foundations of Software Engineering (ESEC/FSE)*. 780–784. https://doi.org/10.1145/3236024.3264843

[28] Hamid R. Motahari Nezhad, Boualem Benatallah, Axel Martens, Francisco Curbera, and Fabio Casati. 2007. Semi-automated adaptation of service interactions. In *International Conference on World Wide Web (WWW)*. 993–1002. https://doi.org/10.1145/1242572.1242706

[29] Nicola Nostro, Romina Spalazzese, Felicita Di Giandomenico, and Paola Inverardi. 2016. Achieving functional and non functional interoperability through synthesized connectors. *Journal of Systems and Software* 111 (2016), 185–199.

[30] Juan Parra-Ullauri, Antonio García-Domínguez, Nelly Bencomo, and Luis Garcia-Paucar. 2022. History-Aware Explanations: Towards Enabling Human-in-the-Loop in Self-Adaptive Systems. In *International Conference on Model Driven Engineering Languages and Systems Companion (MODELS-C)*. 286–295. https://doi.org/10.1145/3550356.3561538

[31] J.L. Pastrana, E. Pimentel, and M. Katrib. 2011. QoS-enabled and self-adaptive connectors for Web Services composition and coordination. *Computer Languages, Systems & Structures* 37, 1 (2011), 2–23. https://doi.org/10.1016/j.cl.2010.07.001

[32] Xin Peng, Yi Xie, Yijun Yu, John Mylopoulos, and Wenyun Zhao. 2014. Evolving Commitments for Self-Adaptive Socio-technical Systems. In *International Conference on Engineering of Complex Computer Systems (ICECCS)*. 98–107. https://doi.org/10.1109/ICECCS.2014.22

[33] H Perera, R Hoda, R Shams, A Nurwidyantoro, M Shahin, W Hussain, and J Whittle. 2021. The impact of considering human values during requirements engineering activities. *IEEE Transactions on Software Engineering* (2021).

[34] DAiSY project. 2023. Diagnostic AI System for Robot-Assisted A&E Triage. https://cs.york.ac.uk/research/projects/daisy-project/. [visited on 5-Oct-2023].

[35] Awais Rashid, John Weckert, and Richard Lucas. 2009. Software Engineering Ethics in a Digital World. *IEEE Computer* 42, 6 (2009), 34–41. https://doi.org/10.1109/MC.2009.200

[36] Milton Rokeach. 1973. *The nature of human values*. Free press.

[37] Shalom H Schwartz. 2012. An overview of the Schwartz theory of basic values. *Online readings in Psychology and Culture* 2, 1 (2012), 2307–0919.

[38] Mojtaba Shahin, Waqar Hussain, Arif Nurwidyantoro, Harsha Perera, Rifat Shams, John Grundy, and Jon Whittle. 2022. Operationalizing Human Values in Software Engineering: A Survey. *IEEE Access* 10 (2022), 75269–75295. https://doi.org/10.1109/ACCESS.2022.3190975

[39] Ian Sommerville, Dave Cliff, Radu Calinescu, Justin Keen, Tim Kelly, Marta Kwiatkowska, John Mcdermid, and Richard Paige. 2012. Large-scale complex IT systems. *Commun. ACM* 55 (2012), 71–77. https://doi.org/10.1145/2209249.2209268

[40] Romina Spalazzese and Paola Inverardi. 2010. Mediating Connector Patterns for Components Interoperability. In *European Conference on Software Architecture (ECSA)*. 335–343. https://doi.org/10.1007/978-3-642-15114-9_26

[41] Bridget Spitznagel and David Garlan. 2003. A Compositional Formalization of Connector Wrappers. In *International Conference on Software Engineering (ICSE)*. 374–384. https://doi.org/10.1109/ICSE.2003.1201216

[42] Savas Takan, Duygu Ergün, Sinem Getir Yaman, and Onur Kilinççeker. 2023. Bias in human data: A feedback from social sciences. *WIREs Data. Mining. Knowl. Discov.* 13, 4 (2023). https://doi.org/10.1002/widm.1498

[43] Richard N. Taylor, Nenad Medvidovic, and Eric M. Dashofy. 2010. *Software Architecture - Foundations, Theory, and Practice*. Wiley. http://eu.wiley.com/WileyCDA/WileyTitle/productCd-EHEP000180.html

[44] Sarah Thew and Alistair Sutcliffe. 2018. Value-based requirements engineering: method and experience. *Requirements Engineering* 23, 4 (2018), 443–464.

[45] Beverley Townsend, Colin Paterson, T. T. Arvind, Gabriel Nemirovsky, Radu Calinescu, Ana Cavalcanti, Ibrahim Habli, and Alan Thomas. 2022. From Pluralistic Normative Principles to Autonomous-Agent Rules. *Minds and Machines* 32 (2022), 683–715. https://doi.org/10.1007/s11023-022-09614-w

[46] Danny Weyns, Radu Calinescu, Raffaela Mirandola, Kenji Tei, et al. 2023. Towards a Research Agenda for Understanding and Managing Uncertainty in Self-Adaptive Systems. *SIGSOFT Softw. Eng. Notes* 48, 4 (2023), 20–36. https://doi.org/10.1145/3617946.3617951

[47] Jon Whittle. 2019. Is Your Software Valueless? *IEEE Software* (2019).

[48] Gio Wiederhold. 1992. Mediators in the Architecture of Future Information Systems. *Computer* 25, 3 (1992), 38–49. https://doi.org/10.1109/2.121508

[49] Emily Winter, Stephen Forshaw, Lucy Hunt, and Maria Angela Ferrario. 2019. Advancing the study of human values in software engineering. In *International Workshop on Cooperative and Human Aspects of Software Engineering (CHASE)*. 19–26. https://doi.org/10.1109/CHASE.2019.00012

[50] Daniel M. Yellin and Robert E. Strom. 1997. Protocol Specifications and Component Adaptors. *ACM Trans. Program. Lang. Syst.* 19, 2 (1997), 292–333.

[51] Enes Yigitbas, Kadiray Karakaya, Ivan Jovanovikj, and Gregor Engels. 2021. Enhancing Human-in-the-Loop Adaptive Systems through Digital Twins and VR Interfaces. In *International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS)*. 30–40. https://doi.org/10.1109/SEAMS51251.2021.00015