The Supportive AI Framework: From recommending to supporting

Toni Waefler[1], Samira Hamouche[1] and Andrina Eisenegger[1]

[1] University of Applied Sciences and Arts Northwestern Switzerland (FHNW), School of Applied Psychology (APS), 4600 Olten, Switzerland.
toni.waefler@fhnw.ch

Abstract. This paper presents the Supportive AI Framework, a conceptual framework for the design of human-AI collaboration to augment human cognition. AI-based decision support systems that are recommendation-driven (i.e. the AI makes a recommendation, and the human must decide whether to accept or reject it) often overstrain humans. The reason for this is the problem known as the 'ironies of automation', which occurs when humans are expected to supervise a technology that exceeds human capabilities. In terms of recommendation-driven AI, this is an impossible task for humans, as they must decide on AI-generated recommendations that take into account far more data and factors than humans are able to consider. Against this background, the Supportive AI Framework aims to go beyond recommendation-driven AI towards AI that explicitly supports cognitive processes of human decision-making, human learning, human trusting, and human motivation. This as a complement to providing comprehensibility through explainable AI and interpretable models. The Supportive AI Framework is theory-based and includes theories from the areas of natural decision making, experiential learning, intrinsic motivation, socio-technical system design and complementary function allocation.

Keywords: Human-AI Collaboration, Augmented Cognition, Decision-Making, Critical Infrastructure, Complementary Function Allocation.

1 Introduction

This is a conceptual, theory-based paper that introduces the Supportive AI Framework. The purpose of the framework is to conceptualize human-AI collaboration for true augmented cognition in demanding decision-making scenarios. This is intended to complement the usual approach to AI-supported decision-making, which is mostly recommendation-driven, i.e. the AI provides recommendations with or without explanations, and the human must decide whether to accept or reject them. Recommendation-driven approaches are also mostly used for AI that is explicitly designed for human-AI teaming, as noted by Dubey et al. (2020).

In contrast, the Supportive AI Framework does not aim to use AI for providing recommendations based on the AI's problem-solving capabilities. Rather, the function of

AI is to explicitly support the problem-solving capacities of humans by explicitly complementing corresponding cognitive processes of humans. In relation to these cognitive processes, the Supportive AI Framework helps to identify opportunities for AI support.

As work psychology is the basis of the framework, it therefore does not provide details for technical design of AI. However, it is integrated into an overall framework for AI development (Bessa et al., 2024) including technical aspects as it was partly developed in the project AI4REALNET, which aims to develop AI-based solutions for critical networks that are traditionally operated by humans, and where AI systems complement and augment human abilities (cf. ai4realnet.eu).

It is also not the aim of the Supportive AI Framework to provide operationalized criteria for the design of human-AI collaboration as other frameworks already do (e.g. Amershi et al, 2019). A very recent of these frameworks is from Kirwan (2025), which sets out detailed criteria for the design of human-AI teaming explicitly taking into account human factors in aviation, i.e. in safety-critical domains. However, he concludes that there is still an important research gap regarding "human-AI teamworking arrangements" (p. 29). Adressing this gap is precisely the aim of the Supportive AI Framework presented in this paper, as it focuses on human cognitive processes in critical decisions in order to explicitly support them through AI. Thereby, our framework goes beyond human cognitive processes of decision-making in the narrower sense and includes the cognitive processes of human learning, human trusting and human motivation.

The Supportive AI Framework was developed in applied research projects that focus on decision-making in knowledge-intensive tasks involving experienced human experts and where the stakes are high. It takes a complementary approach to human-machine function allocation and therefore aims to empower humans through AI rather than replace them. As a consequence, the framework is not suitable for projects aiming at full technical autonomy or support for non-experts, such as consumers.

The next section justifies the need for the Supportive AI Framework. This is followed by an overview of the framework and more detailed descriptions of the different types of AI support for cognitive processes of human decision-making, human learning, human trusting and human motivation. The paper concludes with a brief discussion.

2 Why the Supportive AI Framework is required

Bainbridge (1983) described a major challenge in human-machine interaction as "Ironies of Automation", referring to the fact that the design of technology often leads to an unrealistic task for humans. It is mainly the task of supervisory control (Sheridan, 1987) that humans are not able to take. In this task, humans are expected to monitor and, if necessary, intervene in processes that are automated by a technology whose capabilities exceed those of humans. This exceedance mainly refers to the quickness of information processing and the amount of variables considered in computer controlled processes, which both overstrain human capabilities in real-time monitoring. Furthermore, humans lose skills due to the automation of processes, as they are no longer trained. These are skills that are essential for supervisory control to detect situations that require intervention and to choose the appropriate measures. However, because of

both, the exceedance of human information processing capabilities by computers as well as the deskilling due to automation, supervisory control assigns humans an impossible task, which Bainbridge (1983) calls irony.

While Bainbridge (1983) was referring to automation through programmed computer control, Endsley (2023) builds on her work and discusses the "Ironies of AI". According to her, models trained by machine learning not only pose similar problems to human-machine interaction, but create even further challenges, especially when used to support high stakes decision-making. Some of these challenges arise from the opaqueness and biases of machine-learned models and the possibility that they hallucinate. Furthermore, like any technology, decision support based on machine learning influences human behavior, and not necessarily for the better. Biases in human decision-making, for example, can be exacerbated by AI-generated decision recommendations (Endsley, 2023). This is because providing recommendations can trigger anchoring and confirmation biases in human decision-making.

Making AI comprehensible by the means of interpretability and explainability is the main approach to mitigating the challenges described so far (Schmid, 2024). While the former refers to the models trained by machine learning, the latter refers to the recommendations generated by these models. However, in their literature review Buçinca et al. (2021; 2024) found that humans frequently over-rely on AI and that approaches to make AI comprehensible do not substantially reduce this overreliance. Their reasoning is, that contrary to the expectations of AI developers, humans do not tend to engage analytically with the means that are supposed to make AI comprehensible. Rather, humans switch to what Kahneman (2011) describes as "System 1 Thinking" when using AI, i.e. fast and unconscious thinking based on heuristics rather than conscious analysis. When humans do not engage with AI-generated functions and do not question them, performance decreases, as Dell'Acqua et al. (2023) found in their studies. They therefore raise the question of whether AI is suitable for high-stakes decision-making at all.

To overcome the shortcomings as described above, functions of cognitive forcing are implemented. These functions aim at forcing humans to analytically engage with recommendations generated by AI and hence to switch to "System 2 Thinking" (Kahneman, 2011) when using AI, i.e. slow, logical and conscious thinking. The functions of cognitive forcing ensure that the human remains in the loop, for example by prompting them to make a decision before receiving a recommendation generated by the AI, or by delaying the presentation of a recommendation generated by the AI. Buçinca et al. (2021) found in their study that cognitive forcing significantly reduces overreliance on AI. However, overreliance did not disappear completely. Despite cognitive forcing, the test persons tended to accept incorrect AI recommendations, even if they would have made a better decision without AI. Furthermore, the test persons who over-relied less on AI due to cognitive forcing also liked their tasks less. These results indicate that the improvement through cognitive forcing is accompanied by a poorer user experience.

Buçinca et al. (2021; 2024) as well as Dell'Acqua et al. (2023) conclude that AI-based decision support must go beyond leaving it to humans to accept or reject recommendations generated by AI. Rather, the design of human-AI collaboration must consider the specific knowledge flow required to accomplish a task (Dell'Acqua et al., 2023), and thus not only support the knowledge and mental models of human decision

makers, but even aim to improve the corresponding human decision-making capabilities (Buçinca et al., 2024). This is in line with Endsley's (2023) recommendation that the "ironies of AI" are partly mitigated when AI supports human cognitive decision-making processes rather than just providing decision recommendations.

With his concept of evaluative AI, Miller (2023) takes this claim. Evaluative AI goes beyond cognitive forcing and hence beyond human-in-the-loop towards machine-in-the-loop. Thereby, control over the decision-making process remains with the human, who first formulates a hypothesis, while the AI then provides data-based evidence for and against it. Miller (2023) argues that this approach explicitly supports important steps in human decision-making, such as identifying options and possible outcomes, assessing outcomes, and evaluating trade-offs. This may be considered a paradigm shift, as evaluative AI does not make (comprehensible) recommendations based on its own computational process but rather provides evidence to support the humans in developing their own arguments and making up their own mind.

With our framework for collaboration between humans and AI we extend Miller's (2023) focus on decision-making to include cognitive processes of continuous learning, trusting and intrinsic motivation. Thereby, we aim at conceptualizing for an intensified human-AI collaboration (Waefler, 2021), in which AI explicitly supports these human cognitive processes. Our approach is in the tradition of the complementary design of human-machine systems (Jordan, 1963), which regards humans and machines as complementary, i.e. as qualitatively different, each with different strengths and weaknesses. Accordingly, complementary system design aims at a function allocation that enables both mutual support of strengths and mutual compensation of weaknesses (e.g. Grote et al, 1996; Waefler et al, 2003).

3 The Supportive-AI Framework

Fig. 1 shows an overview of the Supportive-AI Framework. The main purpose of the framework is to conceptualize the collaboration between humans and AI in such a way that opportunities for AI to support human cognitive processes can be identified. Consequently, the framework maps these cognitive processes and the various possibilities of AI support in detail, while the inner workings of AI are not the subject of the framework. Generally, the framework contains four interconnected elements:

- Human agent including cognitive processes of decision-making, continuous learning, trusting and intrinsic motivation.
- Environment representing the subject matter of the decision-making process.
- Human-machine interface (HMI) through which the AI agent supports the human cognitive processes.
- The AI agent, which is not further differentiated.

Although in human-AI collaboration humans and AI normally interact, the arrows from the AI agent via the HMI to the human agent are unidirectional, as they represent the support of human cognitive processes by the AI. Similarly, the arrows originating

from the environment are unidirectional as well, as they represent the flow of information from the environment to the human agent and the AI agent, even though the agents normally interact with the environment.

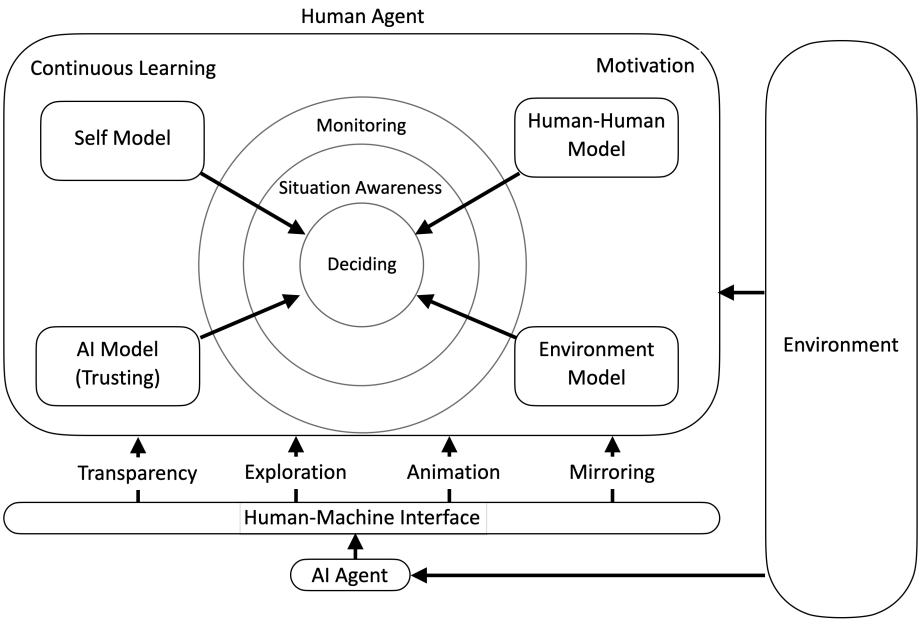

Fig. 1. shows the Supportive AI Framework with its elements Human Agent, Environment, Human-Machine Interface (HMI), and AI Agent. Deciding as a human cognitive process is in the core of the Human Agent. It is embedded in the human cognitive processes Situation Awareness and Monitoring as well as in four different areas of the Human Agent's knowledge required for decision-making, which are represented as mental models (i.e. Environment Model, Human-Human Model, Self Model, and AI Model). In addition, Continuous Learning as well as Motivation are also part of the Human Agent's cognitive processes.

Following, the Supportive AI Framework is described in more detail.

3.1 Different ways of AI support

The Supportive AI Framework emphasizes that AI can support human cognitive processes in different ways. Schmid (2024) provides an overview of the various options for providing transparency and comprehensibility. Although these possibilities of explainability and interpretability make the development and arguments for an AI-generated recommendation transparent, they do not explicitly support the human's own cognitive processes of decision-making. Therefore, the Supportive AI Framework adds three further possibilities of AI support to transparency: exploration, animation and mirroring. However, this list is not exhaustive, as further possibilities for AI support may be identified. Below, the different ways of AI support incorporated in the Supportive AI Framework are described.

Transparency ensures that AI is comprehensible for the humans through explainability and interpretability.

Exploration enables humans to explore the subject matter of their cognitive processes, e.g. the environment or the AI. Its purpose is to increase the humans' knowledge and it can be supported in different ways. Miller's (2023) concept of evaluative AI is a way of AI support for exploring the subject of a decision as it provides humans with evidence for and against their own hypothesis. This explicitly supports explorative human learning. Other forms of exploration may include AI-supported simulation of decision options so that humans can examine their different impacts and their trade-offs. Furthermore, the possibility to explore the AI itself is very important as knowledge about its capabilities and limitations is a prerequisite for developing adequate trust into the AI (e.g. Hoffmann et al., 2018).

Animation has the purpose to trigger cognitive processes and hence animates humans to think. This can take the form of an AI that alerts humans to unusual behavioral patterns in the environment that might be worth taking a closer look at. The AI might also observe the humans' behavior and ask questions about it. This might be especially helpful for experts as they tend to take decisions often intuitively without conscious control (Klein, 1998, 2008). Becoming aware of your own intuitive decisions and reflecting on them can support the learning process, e.g. by recognizing false assumptions.

Mirroring is related to animation, but focuses the humans' self-reflection. Human behavior in general, and also at work, is highly variable (e.g. Hollnagel, 2009). An AI can observe this variability and mirror it to the human. This allows the human to recognize their personal decision-making style, which may involve more risk at the end of the working day than at the beginning. In this way, mirroring can help humans to learn about themselves.

These are just a few ideas of how supportive AI might explicitly support human cognitive processes. However, corresponding AI functionalities still need to be developed. For animation and mirroring, for example, it would be helpful to have an AI that is inquisitive (cf. Wahde & Virgolin, 2022), and thus actively seeks out new insights.

3.2 AI to support decision-making

Since supportive AI aims at supporting human cognitive processes of decision-making, the framework considers the decision as the core element of the human agent (see Fig. 1). To understand the cognitive processes involved in human decision-making, we propose to adopt the concept of "Natural Decision Making" (NDM) (Klein, 1999; 2008), as it explains expert decision-making in real-life situations where the stakes are potentially high. However, the Supportive AI Framework does not limit the decision-making theories to be considered.

NDM is based on studies of experienced professionals who have to make quick decisions in potentially risky situations (e.g. firefighters). These studies provide several important insights into how experts make decisions under time pressure. They are not looking for the optimal but for a satisficing decision. Neither are they evaluating different options for decisions. Rather, based on their experience they know relevant cues, expect certain developments, recognize anomalies, know plausible objectives as well as typical actions required to pursue these objectives. The corresponding cognitive processes are not necessarily conscious, but often unconscious. Their basis is the human ability to recognize patterns in complex situations. This requires a great deal of experience with these situations that makes up the human experts' tacit knowledge. The quickness and accuracy of NDM are its advantage as compared with analytical procedures. However, the experts' tacit knowledge also includes all their erroneous assumptions and biases. AI that helps human experts to uncover such deficiencies in their own decision-making is supportive.

With the concept of macrocognition Klein and colleagues (Klein et al., 2003; Klein, 2018) further differentiated the cognitive processes and functions involved in decision-making. These basic processes include developing mental models, mentally simulating and storybuilding, managing uncertainty and risk, identifying leverage points, managing attention as well as maintaining common ground. Based on these fundamental cognitive processes are the macrocognitive functions, which include sensemaking, (re-)planning, adapting, detecting problems, coordinating as well as deciding.

For the graphical representation of the supporting AI framework (see Fig. 1), we have somewhat simplified the human cognitive processes involved in decision-making by focusing on monitoring, building situation awareness and the actual deciding. While monitoring involves the observation of the current situation, situation awareness comprises the three levels described by Endsely (1995) (i.e. Level 1: Perception of elements in current situation; Level 2: Comprehension of current situation; Level 3: Projection of the future state). Finally, deciding is the selection of a concrete decision. Of course, this simplified representation includes both NDM and the macrocognitive processes and functions in their entirety, as described above.

In the following, examples are described of how AI can support human cognitive decision-making processes.

- Exploration to support situation awareness: AI can support the building of level 3 situation awareness by allowing the human to explore and test their assumptions about possible future states of the current situation.
- Exploration to support deciding: AI can support the actual deciding by giving the human the possibility to explore the implications and trade-offs of using different leverage points to influence the environment.
- Animation to support monitoring: AI can support monitoring by alerting the human decision-maker to situational developments that show unusual patterns.

These examples have in common that AI does more than just make recommendations to humans. Rather, AI supports humans in various cognitive decision-making processes. Identifying corresponding possibilities is the actual purpose of the supporting

AI framework. However, developing concrete options for AI support in a project requires in-depth cognitive task analysis (e.g. Hollnagel, 2003).

3.3 AI to support learning and knowledge building

In the Supportive AI Framework monitoring and situation awareness provide the immediate input for decision-making. Both refer to up-to-date information on the current state of the environment or how it is required in the timeframe of the decision. Information is therefore volatile and must be constantly updated. The human agent's knowledge on the other hand, as conceptualized in the Supportive AI Framework, is more enduring. It incorporates for instance insights about the environment's specific characteristics, about its elements and their systemic interconnections, about its relevant behavioral patterns, and the like. With such content, knowledge is a basis for monitoring, as it tells for instance which indicators need to be monitored and which are not important, as well as for situation awareness, as it provides ground for perceiving, interpreting and projecting.

Like the cognitive processes of decision-making, knowledge is a complex construct. However, it is not the intention of this paper to discuss it in detail. For the purpose of the supporting AI framework, only basic aspects of knowledge and relevant content areas are outlined here.

With the realization that we all know more than we can say, Polanyi (1962) in his fundamental work made the distinction between explicit and tacit knowledge. While humans can communicate their explicit knowledge, this is not possible with tacit knowledge. The latter remains, at least to a large extent, implicit. It is estimated that most of an organization's knowledge is tacit (e.g. embrained or embodied in individuals) and only a minor part is explicit (Faust, 2007). Consequently, NDM assumes that human experts base a substantial part of their decisions on tacit knowledge (Klein, 1998; 2008). The distinction of explicit and tacit knowledge is important for the Supportive AI Framework not only because tacit knowledge is a major source for human experts' decision-making. It is also important because tacit knowledge is not primarily learned through knowledge transfer, but rather by gaining experience through training, practicing, observing and trying things out. Therefore, AI that supports learning processes must go beyond providing explanations and give humans the opportunity to explore and gain experience so that they can expand their tacit knowledge.

Moreover, in the real world of work, where people work in different forms of labor division, relevant knowledge is distributed among many humans rather than owned by individuals. The literature therefore distinguishes between individuals and collectives (or social systems) as entities bearing knowledge (e.g Lam, 2000). Collectives can be formal teams or other forms of organizational units, but they can also be anonymous socio-technical systems (Waefler & Rack, 2021) in which the work of individuals is interconnected without those affected being aware of it. With growing networking as a result of increasing digitalization, there will certainly be even more of these hidden connections and dependencies in the future. There is therefore great potential for supportive AI to provide humans with more transparency about hidden connections and to enable the exchange of knowledge.

Against this background, the Supportive AI Framework distinguishes four domains of human knowledge content, which are depicted as models in the graphical representation of the framework (see Fig. 1), and which are described below.

The environment model contains knowledge regarding the subject-matter of decision-making. This includes knowledge about its characteristics, its behavioral patterns, its elements and their interconnections as well as its typical problem areas with the corresponding leverage points, and the like. More generally spoken, it includes system knowledge and control knowledge (e.g. Kluwe, 2006) regarding the relevant environment or the subject of decision-making.

The human-human model contains knowledge about interrelations of one's own work with the work of others. It takes especially into account, that knowledge about the environment is distributed and that decisions taken by an individual are normally interrelated with decisions of other individuals. The former offers the opportunity to learn from the experiences of others. The latter on the other hand entails the risk that decisions taken individually are sensible from a local perspective, but suboptimal from a global perspective, according to the motto "my solution, your problem". We assume that an increased awareness of these interrelations can support better tuned decision-making. This is in line with Hutchins' (1995) as well as Stanton's et al. (2006) fundamental work on distributed cognition.

The AI model (trusting) contains knowledge about the AI as a tool. This knowledge is not about the algorithms and the inner workings of the AI, but rather about its capabilities and limitations as a basis for obtaining an accurate mental model of the AI (Bansal et al., 2019; Endsley, 2023). Corresponding awareness is prerequisite for developing adequate trust into a specific AI and hence for relying on it when appropriate while not relying on it when inappropriate (e.g. Hoffmann et al., 2018).

The self model contains the human decision-makers' knowledge about themselves, their own strengths and weaknesses, decision biases, behavioral patterns (e.g. the tendency to make riskier decisions at the end of a work shift), and the like. Hence the self model contains what humans learn about themselves to gain a more comprehensive understanding of their behavior (Jelodari et al., 2023; Pronin, 2007). Therefore, the self model is prerequisite for self reflection and metacognition and hence supports continuous improvement.

AI is supportive regarding these domains of knowledge if it empowers corresponding human cognitive processes of learning. Learning is a complex process. It influences the learners' perceptions of the world and their interactions with it. Learning bases on an ongoing, interactive relationship between the learners' characteristics and the learning content, all situated within the specific environment (Alexander et al., 2009).

One suitable conceptualization of these processes is provided by Kolb's (1984) "Experiential Learning Theory". However, as with decision-making (see above), the supporting AI framework does not limit the learning theories that can be considered. Kolb (1984) suggests a cyclic process of experiential learning, consisting of the four phases (i) concrete experience, (ii) reflective observation, (iii) abstract conceptualization, and (iv) active experimentation:

- Concrete experience: This initial phase involves making new experiences within the relevant domain of knowledge or transcending existing ones.
- Reflective observation: In the second phase, humans reflect on their experiences and consider what was successful or where there is room for improvement. This is prerequisite for the internalization of learning outcomes.
- Abstract conceptualization: In the third phase, humans conceptualize their thoughts, adapt existing ideas or develop new ones. In this phase, the abstract understanding materializes and enables the construction of new mental models or conceptual frameworks.
- Active experimentation: Finally, in the fourth phase, before iterating into the next learning cycle, humans test the cognitive representations acquired in the previous phases to observe the outcomes. Feedback from practice, based on active experimentation, is crucial for refining the knowledge acquired.

Kolb (1984) emphasizes the role of experience in human learning, and although it is related to conscious reflection, it applies not only to the learning of explicit knowledge, but also to tacit knowledge. This is because, for example, humans may consciously know and hence develop explicit strategies that are suitable for coping with certain situations based on their experience, although they are not able to justify this explicitly because the corresponding background knowledge is tacit.

The following examples concretizes how AI can support human learning:

- Exploration to learn about the environment: AI can support the human in refining their personal model of the environment by providing them the opportunity to make (simulation-based) new experiences and to reflect on them.
- Animation to refine tacit knowledge about leverage points: AI can support identifying biases in tacit knowledge by alerting human experts to actions they perform intuitively, asking them about the assumptions behind these actions, and helping them to reflect on these assumptions.
- Animation to support human-human knowledge sharing: AI can support learning about the environment by providing humans with experiences (e.g. ratings, likes) of other human experts with strategies for coping with a particular problem in the environment.
- Mirroring to improve the self model: AI can support self-reflection by mirroring to the humans their personal style of decision-making.

These examples are intended to illustrate how the Supportive AI Framework can help to identify multiple ways in which AI can support processes for both explicit and

tacit knowledge learning in the different knowledge areas. Of course, a much more in-depth analysis and ideation process is required for corresponding projects.

3.4 AI to support building adequate trust

As briefly mentioned in the section above, the Supportive AI Framework considers trust to be related with the human's knowledge about the AI, i.e. with their mental AI model (see Fig. 1). In line with the perspective of work psychology on human-machine collaboration, trust is seen as a dynamic process that is influenced by various aspects (Kaplan et al., 2023), including personal experience with a particular AI (Hoffman et al., 2018). While trust can be built on trustworthiness, it does not derive directly from trustworthiness (Hoffman, 2017), which is more of an attribute of a particular AI. Rather than being an AI's attribute, trust is the degree of confidence a particular human has in the automated system's ability to perform accurately in various contexts (Cahour & Forzy, 2009). With increasing experience with a certain AI, a human learns to trust or distrust the AI for certain tasks in certain contexts. Trust therefore does not develop by gradually rising to a more advanced state. Rather, it morphs between over-trust and under-trust in response to concrete experiences and ideally towards appropriate trust. When an appropriate trust is established, humans know when to confidently rely on the AI and when not to rely on it (Hoffman et al., 2018). In the context of the Supportive AI Framework, this knowledge is part of the human's AI model.

There is a huge body of literature on effects of inappropriate trust into automation (for an overview see e.g. Parasuraman & Manzey, 2010). Over-trust on the one hand can result in automation complacency, where humans accept information from the system without checking it or searching for additional information. Consequently, AI errors are not recognized, which leads to errors of commission (where the human blindly follows a recommendation provided by the automation) and errors of omission (where the human does not react in a critical situation because the automation does not prompt them to do so). On the other hand, under-trust can lead to humans not using the AI - or not using it as intended by the developers - and thus not having the opportunity to develop appropriate trust based on experience.

The complex processes in which trust continuously emerges from experience and thus from the specific way in which a technology is used make it clear that successful human-AI teaming requires a differentiated view of the many dimensions of human-AI interaction. This goes far beyond designing the technology. With the supporting AI framework, we want to contribute to this differentiated view.

3.5. AI to support intrinsic motivation

Motivation is considered in the Supportive AI Framework due to two main reasons. On the one hand, there is a tendency for algorithm aversion (Schaap et al., 2023). On the other hand, as already mentioned, even when using an AI, humans tend not to engage analytically with the explanations provided by the AI, but to over-rely on it (Buçinca et al., 2021). Both phenomena make it clear how important motivation is. However, motivation as well is a complex construct with multiple influencing factors (cf. e.g. Ulich,

2011). Many of these, such as personality or extrinsic motivators, are not influenced by the AI design. Nevertheless, AI has an impact on the human's task and therefore on their intrinsic motivation (Parker & Grote, 2022). In the tradition of socio-technical system design (e.g. Clegg, 2000), the term "task orientation" has been important for many decades. It describes a mental state of interest and commitment to the task, which is caused by certain characteristics of the task (Emery, 1959). According to Hackman and Oldham (1976) these task characteristics must support the following three critical mental states as a prerequisite for intrinsic motivation, which can be considered similar to task orientation:

- Experienced meaningfulness: Work needs to be experienced as intrinsically meaningful by the worker, i.e. workers need to directly see why they do what they do. Task characteristics that provide to experienced meaningfulness are (i) skill variety (tasks that require several different skills rather than simple routine), (ii) task identity (tasks that produces an outcome recognizable for the worker, rather than tasks without visible link to the product), and (iii) task significance (tasks that matter).
- Experienced responsibility: Workers need to feel responsible for the outcome of their work. The task characteristic that evokes this feeling is autonomy. If the work processes are fully controlled externally without the workers being able to influence them, they do not feel responsible for the outcome (even if they are held accountable for it by job descriptions or managers).
- Knowledge of the work results: Feedback is the task characteristic that provides this knowledge. For this reason, pedometers for instance motivate people to exercise. Without knowing what they achieved or what they could change to improve their performance, workers cannot be motivated.

These findings of Hackman and Oldham (1976) make it obvious that the integration of an AI into work processes can significantly influence the task characteristics for humans and thus their intrinsic motivation both for good work and for using the AI. Regarding the latter, it can be considered a prerequisite for the developers' intended use of AI that the human user understands the reasons for the AI's behavior and has some control over it. However, the Supportive AI Framework does not limit theories of intrinsic motivation to be taken into account when designing human-AI interaction. Buçinca et al. (2024) for instance suggest to consider the Self-Determination Theory (SDT), which postulates that the three psychological needs of competence, autonomy and relatedness must be met to promote intrinsic motivation.

The following examples describe, how AI can support intrinsic motivation according to Hackman and Oldham (1976):

- Transparency to support experienced meaningfulness: AI can support experienced meaningfulness if it makes its behavior comprehensible.
- Exploration to support experienced meaningfulness: AI can also support experienced meaningfulness if it supports the human to explore causal relations in the environment. This enables humans to understand the why of phenomena they observe in their environment.

- Exploration for experienced responsibility: AI can support human autonomy and hence experienced responsibility if it allows humans to explore their hunches about weak signals that could indicate emerging problems in the environment. This increases the humans' scope of action.
- Transparency for knowledge of the work results: AI can support the feedback and thus the knowledge of the work results if it makes the effects of a decision on the environment transparent in comparison to the effects of other options.

As these examples show, there are many ways in which AI can support intrinsic motivation both for supporting task orientation and thus the human endeavor to do a good job, as well as for the use of AI.

4 Discussion

A large proportion of AI-supported decision-making aids focus on the objective of the decision and aim to make recommendations to humans. As the AI capabilities of data processing far exceed the corresponding human capabilities, it often becomes difficult or even impossible for humans to take the final decision and responsibility for these recommendations. The Supportive AI Framework presented in this paper addresses this problem by shifting the focus of decision support to the human cognitive processes of decision-making. In this way AI is supporting processes of human decision-making rather than recommending decisions. This alternative approach is intended to complement recommendation-based approaches.

Against this background, the aim of the Supportive AI Framework is to enable the identification of various possibilities for supporting human decision-making processes through AI as a basis to derive requirements for AI design. To this end, the framework proposes to take into account a variety of human cognitive processes involved in decision-making. In addition to the actual deciding, these human cognitive processes include monitoring and building situation awareness, but also learning in different knowledge domains such as the environment, the contexts of distributed decision making, the person of the decision maker and the AI. The latter is also important so that humans can build appropriate trust in a particular AI, which requires experience-based knowledge of the AI's capabilities and limitations. Furthermore, human cognitive processes of intrinsic motivation are included in the framework regarding both, the use of the AI and hence to avoid algorithm aversion as well as task orientation. All these human cognitive processes may be supported by AI not only through transparency (i.e. interpretability and explainability), but also through exploration, animation, and mirroring.

The Supportive AI Framework is conceptual and theory-based. Its application to AI research or development projects requires an in-depth analysis of the relevant tasks and the content of the associated cognitive processes. Further research is needed to develop corresponding methods and suitable AI solutions.

Acknowledgments. AI4REALNET has received funding from European Union's Horizon Europe Research and Innovation Programme under the Grant Agreement No 101119527 and from

the Swiss State Secretariat for Education, Research and lnnovation (SERI). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union and SERI. Neither the European Union nor the granting authority can be held responsible for them.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Alexander, P. A., Schallert, D. L., Reynolds, R. E.: What is learning anyway? a topographical perspective considered. Educational Psychologist **44**(3), 176–192 (2009)
2. Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P.N., Inkpen, K., Teevan, J., Kikin-Gil, R. & Horvitz, E: Guidelines for Human-AI Interaction. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19), pp. 1-13. Association for Computing Machinery, New York (2019) https://dl.acm.org/doi/10.1145/3290605.3300233
3. Bainbridge, L.: Ironies of Automation. In: J. Rasmussen, K. Duncan & J. Leplat (eds.). New Technology and Human Error, pp. 271–283. Chichester: Wiley (1987)
4. Bansal, G., Nushi, B., Kamar, E., Lasecki, W. S., Weld, D. S., Horvitz, E.: Beyond Accuracy: The Role of Mental Models in Human-AI Team Performance. Proceedings of the AAAI Conference on Human Computation and Crowdsourcing, vol. 7, 2–11 (2019)
5. Bessa, R. J., Yagoubi, M., Leyli-Abadi, M., Meddeb, M., Boos, D., Borst, C., Castagna, A., Egli, A., Eisenegger, A., Fuxjäger, A., Hamouche, S., Hassouna, M., Lemetayer, B., Marot, A., Liessner, R., Schneider, M., Sturm, I., Usher, J., Van Hoof, H,. Viebahn, J., Kop, S., Waefler, T., Geraldes, J., Felix, C., Sales, H., Leto, G., Ellerbroek, J., Chavarriaga, R., Schiaffonati, V., Zanotti, G., Lundberg, J., Fedorova, A.: Ai4realnet framework and use cases. Technical Report D1.1, European Project AI4REALNET, https://ai4realnet.eu/wp-content/uploads/2024/12/D1.1-AI4REALNET-framework-and-use-cases_v1.0-3.pdf (2024) last accessed 2025/01/22
6. Buçinca, Z., Malaya, M. B., Gajos, K. Z.: To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making. Proc. ACM Hum.-Comput. Interact. 5, CSCW1, Article 188 21 pages (2021) https://doi.org/10.1145/3449287
7. Buçinca, Z., Swaroop, S., Paluch, A. E., Doshi-Velez, F., Gajos, K. Z.: Contrastive Explanations That Anticipate Human Misconceptions Can Improve Human Decision-Making Skills. In Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2024, Woodstock, NY. ACM, New York, NY, USA, 35 pages (2024)
8. Cahour, B., Forzy, J.-F.: Does projection into use improve trust and exploration? An example with a cruise control system. Safety Science, **47**(9), 1260–1270 (2009) https://doi.org/10.1016/j.ssci.2009.03.015
9. Clegg, C.W: Sociotechnical Principles for System Design. Applied Ergonomics 31, pp. 463-477 (2000)
10. Dell'Acqua, F., McFowland III, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., Lakhani, K. R.: Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality. Harvard Business School Technology & Operations Mgt. Unit Working Paper No. 24-013, The Wharton School Research Paper (2023) https://dx.doi.org/10.2139/ssrn.4573321

11. Dubey, A., Abhinav, K., Jain, S., Arora, V., Puttaveerana, A.: HACO: A Framework for Developing Human-AI Teaming. In 13th Innovations in Software Engineering Conference (formerly known as India Software Engineering Conference) (ISEC 2020), February 27–29, 2020, Jabalpur, India. ACM, New York, NY, USA, 9 pages (2020) https://doi.org/10.1145/3385032.3385044
12. Emery, F.E.: Characteristics of Sicio-Technical Systems. Tavistok Institute of Human Relations. Document Nr. 527 (1959)
13. Endsley, M.R.: Towards a theory of situation awareness in dynamic systems. Human Factors. **37**(1), pp. 32-64 (1995)
14. Endsley, M. R. (2023). Ironies of artificial intelligence. Ergonomics, **66**(11), 1656–1668 (2023) https://doi.org/10.1080/00140139.2023.2243404
15. Faust, B.: Implementation of tacit knowledge preservation and transfer methods (2007) available at https://www.academia.edu/36648357/Implementation_of_Tacit_Knowledge_Preservation_and_Transfer_Methods
16. Grote, G., Weik, S., Waefler, T.: KOMPASS: Complementary allocation of production tasks in sociotechnical systems. In S. A. Robertson (Ed.), Contemporary Ergonomics 1996 (pp. 306-311). Taylor & Francis, London (1996)
17. Hackman, J. R., Oldham, G. R.: Motivation through the design of work: Test of a theory. Organizational Behavior and Human Performance, **16**(2), 250–279 (1976)
18. Hoffman, R.: A taxonomy of emergent trusting in the human–machine relationship. In Cognitive Systems Engineering: The Future for a Changing World, pp. 137–163. CRC Press (2017)
19. Hoffman, R., Mueller, S. T., Klein, G., Litman, J.: Measuring Trust in the XAI Context (Explainable AI Program) [Technical Report]. DARPA (2018) https://doi.org/10.31234/osf.io/e3kv9
20. Hollnagel, E.: Handbook of Cognitive Task Design. Lawrence Erlbaum, Mahwah, NJ (2003)
21. Hollnagel, E.: The ETTO Principle. Ashgate, Farnham (2009)
22. Hutchins. E.: Cognition in the Wild. The MIT Press, Cambridge MA (1995)
23. Jelodari, M., Amirhosseini, M. H., Giraldez-Hayes, A.: An AI powered system to enhance self-reflection practice in coaching. Cognitive Computation and Systems, **4**(5), 243–254 (2023) https://doi.org/10.1049/ccs2.12087
24. Jordan, N.: Allocation of functions between man and machines in automated systems. Journal of Applied Psychology, **47**(3), 161-165 (1963)
25. Kahneman, D.: Thinking, Fast and Slow. Farrar, Straus and Giroux, New York (2011)
26. Kaplan, D., Kessler, T. T., Brill, J. C., Hancock, P. A.: Trust in artificial intelligence: Meta-analytic findings. Human Factors: The Journal of the Human Factors and Ergonomics Society, **65**(2), 337–359 (2023)
27. Kirwan, B.: Human Factors Requirements for Human-AI Teaming in Aviation. Preprints (2025) https://doi.org/10.20944/preprints202501.0974.v1
28. Klein, G.: Sources of Power. How People make Decisions. The MIT Press, Cambridge MA: (1998)
29. Klein. G.: Naturalistic decision making. Human Factors: The Journal of the Human Factors and Ergonomics Society. **50**(3), 456–460 (2008)
30. Klein, G.: Macrocognitive Measures for Evaluating Cognitive Work. In E. S. Patterson & J. E. Miller (Eds.), Macrocognition Metrics and Scenarios: Design and Evaluation for Real-World Teams. pp. 47–64. Ashgate, Farnham (2018)
31. Klein, G., Ross, K. G., Moon, B. M., Klein, D. E., Hoffman R. R., and Hollnagel E. Macrocognition. IEEE Computer Society. pp. 81–84 (2003)

32. Kluwe, R.H.: Informationsaufnahme und Informationsverarbeitung. In: B. Zimolong &. U. Konradt (Hrsg.). Ingenieurpsychologie, Enzyklopädie der Psychologie (Themenbereich D, Serie III, Band 2, S. 35-70). Hogrefe, Göttingen (2006)
33. Kolb, D. A.: Experimental learning: Experience as the source of learning and development. Prentice-Hall: Englewood Cliffs, NJ (1984)
34. Lam, A.: Tacit Knowledge, Organizational Learning and Societal Institutions: An Integrated Framework (2000) available at: http://oss.sagepub.com/content/21/3/487
35. Miller, T.: Explainable AI is Dead, Long Live Explainable AI! Hypothesis-driven decision support (2023) https://arxiv.org/abs/2302.12389
36. Parasuraman, R., Manzey, D.H.: Complacency and bias in human use of automation: An attentional integration. Human Factors. **52**(3), pp. 381–410 (2010)
37. Parker, S. K., Grote, G.: Automation, algorithms, and beyond: Why work design matters more than ever in a digital world. Applied Psychology **71**(4), 1171–1204 (2022)
38. Polanyi, M.: Personal Knowledge: Towards a Post-Critical Philosophy. Taylor & Francis, London (1962)
39. Pronin, E.: Perception and misperception of bias in human judgment. Trends in Cognitive Sciences **11**(1), 37–43 (2007) https://doi.org/10.1016/j.tics.2006.11.001
40. Schaap, G., Bosse, T. & Hendriks Vettehen, P. The ABC of algorithmic aversion: not agent, but benefits and control determine the acceptance of automated decision-making. AI & Soc 39, 1947–1960 (2024). https://doi.org/10.1007/s00146-023-01649-6
41. Schmid, U.: Trustworthy Artifcial Intelligence: Comprehensible, Transparent and Correctable. In: A. Werthner, C. Ghezzi, J. Kramer, J. Nida-Rümelin, B. Nuseibeh, E. Penn & A. Stranger (eds.). Introduction to digital humanism. Springer, Cham (2024) https://doi.org/10.1007/978-3-031-45304-5_10
42. Sheridan, T.B.: Supervisory Control. In: G. Salvendy (eds.), Handbook of Human Factors, pp. 1243-1268. Wiley, New York (1987)
43. Stanton, N. A., Stewart, R., Harris, D., Houghton, R. J., Baber, C., McMaster, R., Salmon, P., Hoyle, G., Walker, G., Young, M.S., Linsell, M., Dymott, R. Green, D.: Distributed situation awareness in dynamic systems: theoretical development and application of an ergonomics methodology. Ergonomics, 49(12–13), 1288–1311 https://doi.org/10.1080/00140130600612762 (2006)
44. Ulich, E.: Arbeitspsychologie (11. Auflage). Schäffer-Poeschl Verlag, Stuttgart (2011)
45. Waefler T.: Progressive Intensity of Human-Technology Teaming. Proceedings of the 5th International Virtual Conference on Human Interaction and Emerging Technologies, IHIET 2021, August 27–29, 2021, France, pp. 28–36 (2021)
46. Waefler, T., Grote, G., Windischer, A., Ryser, C.: KOMPASS: A Method for Complementary System Design. In: Handbook of Cognitive Task Design. E. Hollnagel (Ed.), pp. 477–502. Lawrence Erlbaum, Mahwah, NJ (2003)
47. Waefler, T., Rack, O.: Kooperation und künstliche Intelligenz. In O. Geramanis, S. Hutmacher, & L. Walser (Eds.), Kooperation in der digitalen Arbeitswelt (pp. 77–88). Springer Fachmedien, Wiesbaden (2021) https://doi.org/10.1007/978-3-658-34497-9_5
48. Wahde, M. & and Virgolin, M.: The five Is: Key principles for interpretable and safe conversational AI. In: Proceedings of the 2021 4th International Conference on Computational Intelligence and Intelligent Systems (CIIS '21), pp. 50–54. Association for Computing Machinery, New York, NY, (2022) https://doi.org/10.1145/3507623.3507632