

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Golovianko, Mariia; Gryshko, Svitlana; Terziyan, Vagan; Tuunanen, Tuure

Title: Responsible cognitive digital clones as decision-makers : a design science research study

Year: 2023

Version: Published version

Copyright: © 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis

Rights: CC BY-NC-ND 4.0

Rights url: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Please cite the original version:

Golovianko, M., Gryshko, S., Terziyan, V., & Tuunanen, T. (2023). Responsible cognitive digital clones as decision-makers : a design science research study. *European Journal of Information Systems*, 32(5), 879-901. <https://doi.org/10.1080/0960085x.2022.2073278>

Responsible cognitive digital clones as decision-makers: a design science research study

Mariia Golovianko, Svitlana Gryshko, Vagan Terziyan & Tuure Tuunanen

To cite this article: Mariia Golovianko, Svitlana Gryshko, Vagan Terziyan & Tuure Tuunanen (2022): Responsible cognitive digital clones as decision-makers: a design science research study, European Journal of Information Systems, DOI: [10.1080/0960085X.2022.2073278](https://doi.org/10.1080/0960085X.2022.2073278)

To link to this article: <https://doi.org/10.1080/0960085X.2022.2073278>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 16 May 2022.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

Responsible cognitive digital clones as decision-makers: a design science research study

Mariia Golovianko ^a, Svitlana Gryshko ^b, Vagan Terziyan ^c and Tuure Tuunanen ^c

^aDepartment of Artificial Intelligence, Kharkiv National University of Radioelectronics, Ukraine; ^bDepartment of Economic Cybernetics, Kharkiv National University of Radioelectronics, Ukraine; ^cFaculty of Information Technology, University of Jyväskylä, Finland

ABSTRACT

This study uses a design science research methodology to develop and evaluate the Pi-Mind agent, an information technology artefact that acts as a responsible, resilient, ubiquitous cognitive clone – or a digital copy – and an autonomous representative of a human decision-maker. Pi-Mind agents can learn the decision-making capabilities of their “donors” in a specific training environment based on generative adversarial networks. A trained clone can be used by a decision-maker as an additional resource for one’s own cognitive enhancement, as an autonomous representative, or even as a replacement when appropriate. The assumption regarding this approach is as follows: when someone was forced to leave a critical process because of, for example, sickness, or wanted to take care of several simultaneously running processes, then they would be more confident knowing that their autonomous digital representatives were as capable and predictable as their exact personal “copy”. The Pi-Mind agent was evaluated in a Ukrainian higher education environment and a military logistics laboratory. In this paper, in addition to describing the artefact, its expected utility, and its design process within different contexts, we include the corresponding proof of concept, proof of value, and proof of use.

ARTICLE HISTORY

Received 02 June 2020
Accepted 20 April 2022

KEYWORDS

Artificial intelligence;
cognitive clones; decision-
making; design science
research; digital twinning

1. Introduction

The current vision of transformation across various industries and fields embraces the idea of human-centric, resilient, and sustainable processes (Breque et al., 2021). Despite tremendous advances in digitalisation and automation (Tuunanen et al., 2019), human employees play the dominant role, especially in creative, strategic, and emergent decision-making (Jarrahi, 2018). A human-centric perspective, however, leads to more subjective decisions and high dependency on particular employees, therefore creating preconditions for the processes’ bottlenecks, information, and work overload (Matthews et al., 2019; Phillips-Wren & Adya, 2020), thus threatening organisational resilience and sustainability (Raetze et al., 2021). Recent emergencies caused by refugee crises, global hacker attacks, hybrid activities, wars, COVID-19, etc. (Sakurai & Chughtai, 2020) have provided many illustrative examples of the disruption of organisational decision-making due to the human factor. We call this challenge “overload vs. organizational resilience”.

One way to increase organisational resilience and sustainability for human-centricity is digitalisation and cybernization of real-world human-driven decision environments by digital twinning, which translates assets, processes, information systems, and

devices into digital resources (Jones et al., 2020). This study focuses on the digital twinning of human resources (HR) to reduce overload and achieve process resilience. Human digital twins, or cognitive clones in digitising human cognitive capabilities (Somers et al., 2020), act as representatives of human actors in cyberspace. They can either (a) copy (simulate) both correct and incorrect (former and potential) decisions by the “donor” (the human owner), acting according to an “exact”, imperfect, and biased model of a particular “donor”, or (b) be capable of correcting potential decision-making mistakes of the “donor” using an automated machine model of a “perfect”, unbiased decision-maker). Option (b) has long been the goal of AI community studies, whereas Option (a) has not yet been studied enough. We intend to uncover the possibilities of a new class of applications of Option (a), leveraging the twinning of a decision maker’s mental model with all personal preferences, features, and biases. Shifting the focus of attention from the long-living Option (b) to the emerging Option (a) is a relatively new phenomenon. The shift is the recently recognised need to bring humans back into the loop of highly automated processes’ control and the newly emerged opportunities due to the evident success of deep learning models and tools in the cognitive computing domain.

CONTACT Tuure Tuunanen  tuure@tuunanen.fi

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

Our research questions are as follows:

- How can the digital cognitive clone of a donor be designed to inherit the personal preferences, responsibilities, decision features, and biases of a decision-maker?
- How can the cognitive clone be applied for resilient, sustainable, and human-centric organisational decision-making?

To achieve this, we apply design science research (DSR) (Hevner et al., 2004; Peffers et al., 2007) to develop, demonstrate, and evaluate an information technology (IT) artefact (the Pi-Mind agent) capable of making decisions on behalf of its human donor concerning the personal preferences, decision features, and biases of a particular decision-maker.

Previous studies have shown how cognitive clones, which have been created with semantic modelling and deep learning, can digitalise decision systems in Industry 4.0 (Longo et al., 2017; Terziyan et al., 2018b). This study focuses on developing an applicable IT artefact suitable for various organisational and business processes.

The rest of the paper is organised as follows. We present how we applied DSR methodology (DSRM) in the research. Next, we study the context of

digital transformation, digital twinning, and decision-making automation, formulate the design principles, and describe the design of the Pi-Mind agent. We then demonstrate and evaluate the developed artefact. Lastly, we discuss the findings and conclude.

2. Research methodology

To develop cognitive digital clones capable of acting as decision-makers, we apply the DSRM (see, Figure 1). The DSRM (Peffers et al., 2007) comprises the following phases: identifying the problem and motivation, defining the objectives (for a solution), designing and developing artefact(s), demonstrating the solution to the problem, and evaluating the solution (Peffers et al., 2007). We adopt the DSR evaluation approach suggested by Venable et al. (2012) and Nunamaker et al. (2015). Namely, we demonstrate the effectiveness of the artefact (proof of concept), evaluate its efficiency (proof of value) for achieving its stated purpose, and identify the side effects or undesirable consequences of its use for Ukrainian higher education (HE) and society at large (proof of use). This evaluation approach follows Tuunanen and Peffers (2018) and Nguyen et al. (2020).

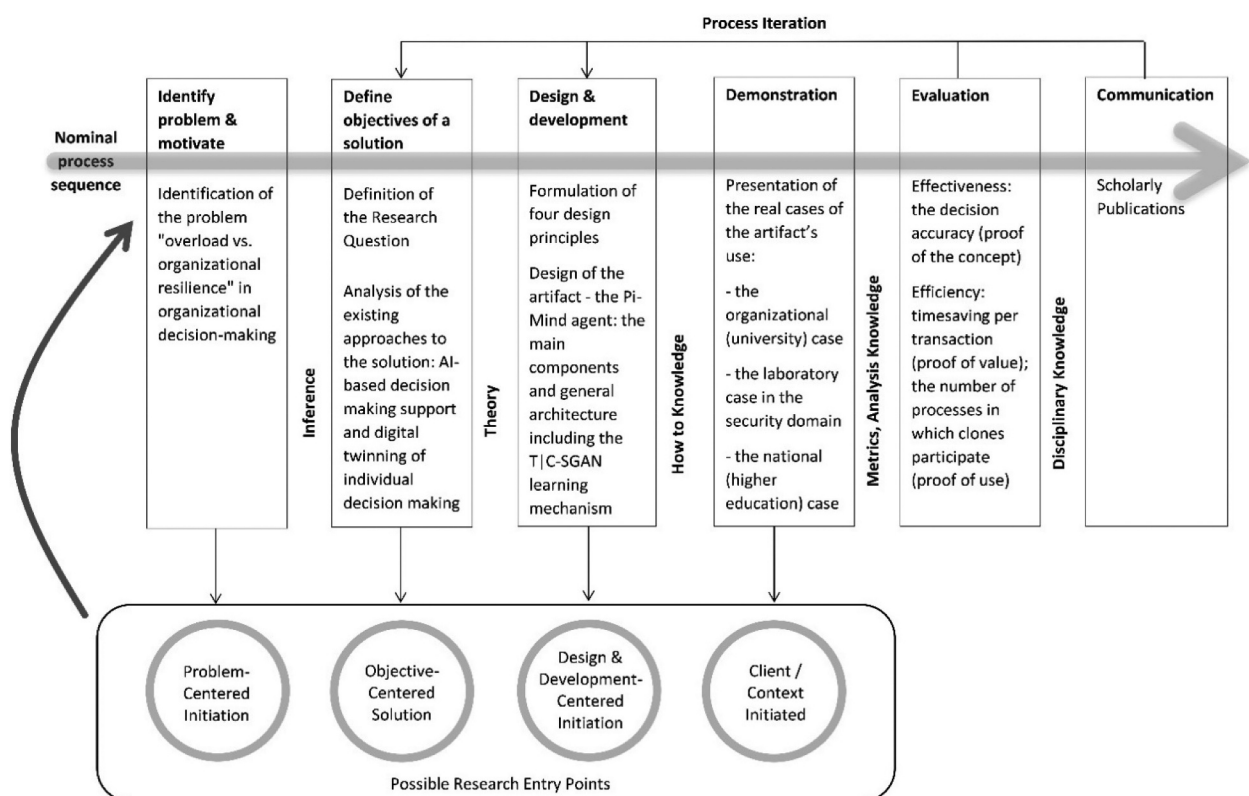


Figure 1. The DSRM applied for the study, adapted from Peffers et al. (2007).

3. Related work

The relationship between digital and physical realities is changing dramatically. It is not just about the emergence of new scenarios in underlying process management (Filip et al., 2017) or digital transformation in organisational structures and management concepts (Kuziemski & Misuraca, 2020). An ontological shift towards a “digital-first” (Ågerfalk, 2020; Baskerville et al., 2019) has led to a new logic of organisational decision-making regarding the underlying models of human–computer interaction.

Traditional organisational decision-making deals with overload in the workplace with simplified models of human–computer interaction. The three basic approaches are straightforward and represent a certain compromise between the two extremes: (1) entirely human decision-making (HR management approach) and (2) completely automated decision-making (see, Table 1).

The new decision-making logic requires clear interaction of the data-driven insights and the behavioural drivers behind human analysts and managers (Li & Tuunanen, 2022; Sharma et al., 2014). The new logic should be “structured to be modifiable rather than rigidly fixed”, light touch processes, infrastructurally flexible to enable the “flexibility and configurability of process data flow”, and mindful actors who act “based on the prevailing circumstances of the context” (Baiyere et al., 2020, p. 3). All can be facilitated by the smart integration of human and machine intelligence. Human workforce decisions are combined with those made by artificial mindful actors (Duan et al., 2019).

There are different views on model joint human and AI decision-making. Ultimately, all human-AI decision-making approaches can be combined into several potential collaborative decision-making scenarios: complete human-to-AI delegation, hybrid (human-to-AI and AI-to-human sequential decision-making), and aggregated human-AI decision-making that aggregates the decisions of a group of individuals (Shrestha et al., 2019). Human judgement and individual preferences are expected to remain vital (Agrawal et al., 2019), especially in agile environments (Drury-Grogan et al., 2017). Therefore, hybrid scenarios of organisational decision-making based on human–AI symbiosis are estimated to prevail (Jarrahi, 2018). Human–AI symbiosis is seen as intelligence augmentation in such hybrid models, demanding that AI extend human cognition when addressing complexity. By contrast, humans offer a holistic, intuitive approach to dealing with uncertainty and equivocality. As machine-prediction technology improves, complex decisions (besides easy-to-automate jobs) will be increasingly delegated to an artificial workforce (Bughin et al., 2018; Kolbjørnsrud et al., 2016).

Acharya et al. (2018) reported a hybrid model in which human operators evaluate automated advice based on their preferences and knowledge. Ruijten et al. (2018), Wang et al. (2019), and Golovianko et al. (2021) demonstrated that AI models are twins of human decision behaviour in various organisational processes.

Even though digital twins have already been implemented, the common vision of “digital representatives” is still under development (Rathore et al., 2021). The theoretical foundation for the creation of digital representatives for the human owner (Pi-Mind agent) was 1) the concept of digital twins (Grieves, 2019) as high-quality simulations or digital replicas of physical objects, and 2) cognitive/digital clones (Al Faruque et al., 2021; Terziyan et al., 2018b) as a “cloning” technology regarding the cognitive skills of humans.

Cognitive cloning is quickly becoming mainstream in today’s applied IT developments (Becue et al., 2020; Booyse et al., 2020; Hou et al., 2021). NTT Secure Platform Laboratories has defined a digital twin as an autonomous, agent-driven entity (Takahashi, 2020). Microsoft has patented the technology for creating digital clones (to the extent of conversational bots) for specific persons (Abramson & Johnson, 2020), and Truby and Brown (2021) discussed the ethical implications of “digital thought clones” for customer experience personalisation.

4. Design of the Pi-Mind agent

4.1. Design principles of the artefact

Attempts to simulate or automate decision-making always encounter some polarising opinions: what should one rely on when facing a decision problem: expertise, intuition, heuristics, biases, calculations, or algorithms (Kahneman & Klein, 2009)? Can human decision-making be improved by either helping a person (partial automation and augmentation) or replacing the decision-maker with advanced AI algorithms (full autonomy)? Where should we get the decision-making procedures? Should they come from the direct transfer of knowledge from humans to decision-making systems, or the automated discovery of knowledge based on observations and machine learning (ML)? What contexts should we foresee for decision-making? Should these be simple or complex, with full information or under uncertainty, unlimited, or limited resources (time, memory, etc.), and business-as-usual situations or crisis management? To have a personal, autonomous, always available, as-smart-as-a-human, reliable, fast, and responsible digital substitute for oneself, it is necessary to abandon attempts to choose just one direction of research. After all, they are all present (one way or another) in the human mental model.

Table 1. Existing solutions for decision-making in organisations.

| Approach | Disadvantages | Advantages |
|--|--|--|
| HR management | The organisation development (Cummings & Worley, 2014), the change management (Hayes, 2018), and the business process management (Jeston & Nelis, 2014) consider this problem in the context of organisations, here being their hierarchical structures with the top-down coordination and procedural actors. This context does not cover the logic of peer-to-peer networking workflows typical of virtual collaboration (Lechner & Hummel, 2002). | Human decision-making is still considered the “gold standard”: people can explain the reasons behind their recommendations and take responsibility for their decisions (Miller, 2019). |
| Decision support systems (DSS) | DSS can reduce the workload by empowering people and helping report, analyse, and interpret data rather than executing business decisions. However, the use of DSS has its limitations. It can even increase the workload because of fragmentation of workflows, excessive/unimportant alerts, dependence on the technological proficiency of users, the need for permanent maintenance, the operational impact of poor data quality and information incorrectness, and interoperability issues (Sutton et al., 2020). | DSS increases resilience in the workplace or the ability to cope with and adapt to new situations (Hartmann et al., 2019) by distributing the complex decision-making process between the information system, taking over routine and complex but automated operations. Thus, leaving the person with duties depends on the cognitive style and approach to problem-solving of the decision-maker (Sprague, 1980). Recently, DSS have been taking on more and more human tasks, for example, web-based group DSS (Carneiro et al., 2019), and cognitive DSS (Lai et al., 2020). |
| Automated decision-making systems (ADMS) | Substituting a human decision-maker with ADMS, which can produce a decision or a recommendation, eliminates the disadvantages of DSS. It can also increase workloads and stress in the workplace because of additional problems related to expert knowledge extraction and formalisation, limited opportunities to become an expert, legal implications, and so forth (Harris & Davenport, 2005). | The use of ADMS creates ubiquity “as an ongoing relationship between the individual and the technological possibilities offered to this individual” (Sorensen, 2010, p. 6), everywhere, anywhere, at any time, in our location and context (Greenfield, 2010) – the employee gets the opportunity to stay in the loop without being overwhelmed. |

The main idea comprises a partial shift from human-driven to AI-driven decision-making supported by an IT artefact, the Pi-Mind agent, which enables the replacement (when needed and appropriate) of a human actor (donor) at decision points with a personal cognitive clone (aka proactive digital twin). This ensures ubiquity or the involvement of a particular person in many processes simultaneously without losing the characteristics of decision-making (responsibility) and cognitive tension (resilience).

Thus, we formulate design principles (DPs), prescriptive statements to constitute the basis of the design actions (Chatterjee et al., 2017), and approaches for the evaluation of the DPs’ implementation (see, Table 2). The artefact should be designed as a responsible (DP1), resilient (DP2), and ubiquitous (DP3) cognitive copy (DP0) of its donor. The primary and fundamental design principle is the Turing principle (DP0), which is inspired by the historical principle of AI – the so-called Turing test that examines the ability of AI to exhibit intelligent behaviour equivalent to or indistinguishable from human behaviour (Turing, 1950).

To assess the implementation of all the DPs, we measure the effectiveness of the artefact in terms of the correspondence between a clone and donor; we also measure efficiency in terms of the benefits gained from replacing the human decision-maker with AI.

Several metrics of the correspondence between a clone and the donor allow evaluation of the quality of the implementation of DP0-DP2. Considering the decision evaluation as a binary classification problem (with two possible output classes: “correct decision” and “incorrect decision”), we apply the F1-score as a metric of decision accuracy. The F1 score is

a widely used metric for model performance in both AI and management, making it universal and applicable to all our cases.

Efficiency indicates user utilities; thus, it can be interpreted and measured using several metrics. We consider the challenge of overload versus organisational resilience to be the main benefit of utilising the Pi-Mind agent. The clone enables the participation of a human in various processes simultaneously (DP3). The donor saves time by delegating part of the work to the clone. Therefore, we use the following efficiency metrics: timesaving per transaction and the number of processes in which a clone participates simultaneously.

4.2. Twinning decision behaviour

The Pi-Mind agent is an intelligent model for ubiquitous decision-making based on the digital twinning of human decision behaviour. A specific instance of the Pi-Mind agent is a digitally shared proactive copy of a particular person’s decision system in terms of decision schemes and preferences that depend on specific tasks, domains, and contexts. Digital decision-specific knowledge is stored as interrelated semantic resources in a set of ontologies. Personal preferences are defined as a unique set of decision criteria formalised with various information models, such as mathematical models (e.g., neural networks and decision trees), detailed specifications (algorithms), or explicit scoring (judgement) systems based on customised sets of values.

To provide a new instance of the Pi-Mind agent with knowledge, we extract decision-specific knowledge from the human donor, annotate it in terms of

Table 2. Pi-Mind agent design principles.

| Label | Name | Description | Evaluation approach |
|-------|------------------|---|---|
| DP0 | Turing principle | There is a minimal deviation of a clone's cognitive behaviour from the behaviour of its donor in similar or identical situations. It ensures that a human is at the centre of decision-making (Shneiderman, 2020). | The precision of the imitation: The inverse of a measured difference between the decisions made by the clone and the donor over the same inputs (minimal difference – maximal precision). |
| DP1 | Responsibility | A clone acts as a representative of a human donor, inheriting her personal behavioural characteristics to take over specific responsibilities. This identical behaviour ensures that a clone acts as responsibly as its donor. Thus, a donor remains morally and legally responsible for decisions. | The percentage of responsibilities (decision-making duties) delegated to a clone by its donor. |
| DP2 | Resiliency | A clone can address the capabilities (with personal specifics) of a donor to handle challenging situations (emergencies, attacks, etc.) reactively and proactively. | The precision of the imitation: The inverse of a measured difference between the decisions made by the clone and the donor over the same inputs in an emergency context (new, challenging, confusing, stressful, etc. situations) (minimal difference – maximal precision). |
| DP3 | Ubiquity | A clone enables the digital involvement of a particular person in many processes simultaneously. | The number of persons/hours saved because of synchronous and efficient use of the clone within several processes (parallelisation). This value will grow with the number of synchronously running processes. |

decision ontologies, and store it in the corresponding knowledge base. Suppose that the decision behaviour can be explicitly explained. In that case, this task is reduced to configuring the most accurate mathematical decision-making model (identifying the correct type of model, the appropriate parameters, and the relationships between them). The general laws, principles, and rules are rigidly fixed in this case.

The basic model is shared among all decision-makers and functions as the basis for making decisions. However, the final decisions are unique because each decision-maker customises and personalises the model by setting up the preferences (consciously assigned values for the parameters) within it. With this approach, the digitised human is a carrier of a vector of values for the model parameters. Decisions made according to this method are explainable because the essence of each parameter and its comparative value can be understood. This approach is well established in DSS, formalising human (or group) multi-criteria decision behaviour in business-as-usual situations.

However, this approach encounters problems if the situation ceases to be ordinary. When decision-makers leave their comfort zones because of fuzziness, uncertainty, or a lack of information, the previously fixed model becomes irrelevant. In this situation, humans are still capable of making decisions; however, they unconsciously use hidden personal biases, heuristics, and intuition. It is difficult for a person to explain their decisions in this situation. Therefore, a different technology for extracting knowledge is needed.

This approach assumes the model (the architecture and parameters) to be self-configurable when observing the human donor's decision behaviour. The outcomes of this model cannot be explained by analysing its parameters, because the values for these parameters are not provided by donors but by artificial

(computational) intelligence when simulating human behaviour. In this approach, the information system can be trained to make personalised decisions in various situations, simulating concrete human decision-maker behaviour.

We use both fundamental AI approaches to extract expert knowledge (see, Figure 2).

In the first facet shown in Figure 2, the top-down (“symbolic”) approach, which is based on semantic annotation, experts specify and explicitly conceptualise how they evaluate alternatives and choose a solution. In the second facet, the bottom-up approach, computational intelligence enables learning decision-specific knowledge from observation for more intuitive decision practices and weakly formalised problems. The third facet enables the autonomous behaviour of the Pi-Mind, exhibiting proactivity and the abilities of situation- and self-awareness and self-management. These are considered important components of future strong AI, artificial general intelligence, and super-intelligence and are mainly associated with the development of reinforcement learning techniques (Silver et al., 2021).

4.3. Creating cognitive clones as AI agents

Agent-based models drive the autonomous (the third AI facet in Figure 2) and proactive behaviour of the clones as ubiquitous activity within decision environments (DP3). During the technology's life cycle, the Pi-Mind agent “lives” in two types of modes almost synchronously. In the “operational” mode, the clone addresses different decision-making tasks (jobs). In the “university” mode, the agent develops its cognitive capabilities for better performance, aiming to adapt its behaviour continuously to new challenges, contexts, and tasks. Within such an environment, the Pi-Mind agent learns the following:

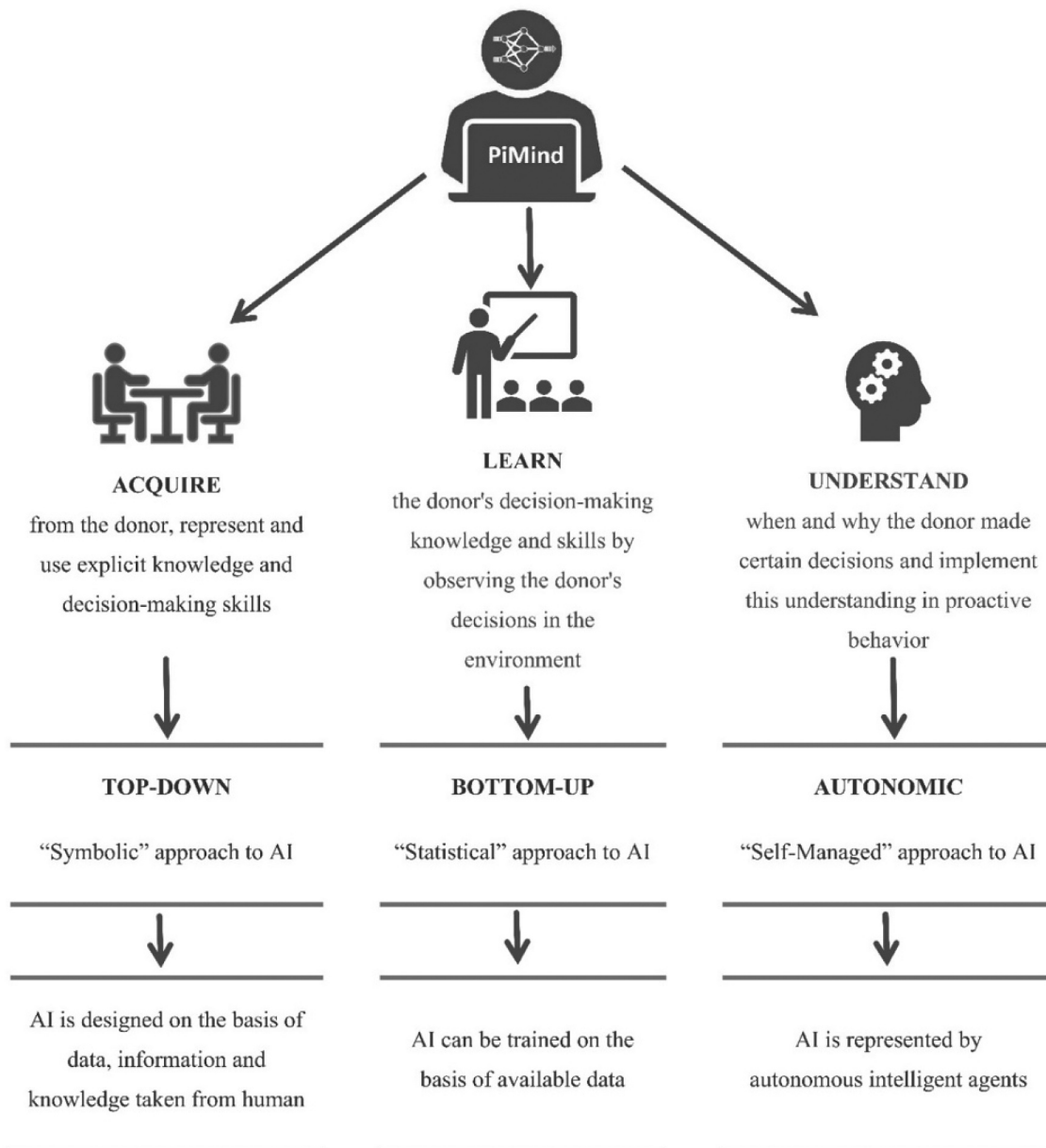


Figure 2. Three facets of the Pi-Mind agent.

- (1) Understanding and using the personal decision preferences and values of its human donors and choosing the most appropriate solution among alternatives in business-as-usual processes.
- (2) Applying known decision models to different but related problems with transfer learning.
- (3) Obtaining critical decision-making experience and preparing for new, difficult tasks, complex contexts, and disruptions from continuous retraining in adversarial settings.
- (4) Training the AI immunity (Castro et al., 2002) against cognitive attacks (Biggio & Roli, 2018). Such attacks threaten an agent's intelligence capabilities by breaking ML models with specifically crafted adversarial inputs.

Unlike simulators with embedded typical behaviour patterns, the Pi-Mind agent acquires the ability to learn and proactively exhibits the unique decision-making behaviour of a particular person (donor). A simple formula connecting all three ideas with the concepts in Figure 2 is as follows:

$$\begin{aligned}
 \text{Pi - MindAgent}(\text{cognitive clone of a human - as - a - decision maker}) = & \\
 = \text{DigitalTwin} & \left(\begin{array}{l} \text{SYMBOLIC_+_STATISTICAL_AI_--} \\ \text{DRIVEN personalized, learnable model of} \\ \text{human cognitive skills} \end{array} \right) + \\
 + \text{SmartResource} & \left(\begin{array}{l} \text{SELF - MANAGED_AI_--} \\ \text{DRIVEN generic model of digital} \\ \text{consciousness : autonomy, proactivity,} \\ \text{self - awareness, and self - management} \end{array} \right).
 \end{aligned}$$

Here, we use a pragmatic notion of “consciousness”, which supports the concept of the Pi-Mind agent. According to our definition, consciousness (for a human or for an AI agent) includes (a) self-awareness – understanding the boundary between everything within the accessible “me” and the accessible “the rest of the world” and (b) self-management, which is an autonomic, proactive activity for keeping the balance between these two. The “balance” means the following: (a) for a human, sustainable opportunity to complete the personal mission statement, and (b) for an AI agent, sustainable opportunity to complete its design objectives.

This extension of the digital twin concept also influences the distribution of responsibilities among the relevant players. The players are as follows: the donor (the object for twinning), the designer (the subject performing twinning), the clone (the outcome of twinning), and the user (the user of the designed clone). Consider the case when the clone is a simple digital twin of some donor. In this case, the user takes remote control of the clone. If some severe fault happens in use, the responsibility will be divided between the user (for possible incorrectness in control) and in the designer (when the clone is not functioning according to the agreed-upon specification). Consider the case when Pi-Mind drives the clone of a donor. In this case, the designer and donor are the same person; the clone works (behaves and decides) proactively, following the donor’s decision and self-management logic. Therefore, here, the user will not be responsible for the potential problems, and the responsibility will be entirely on the side of the clone – that is, the donor and/or designer. The difference between these two scenarios is another reason for using the term “responsible cognitive clones” when discussing Pi-Mind agents.

This property allows us to talk about “soft substitution” that preserves jobs. The clone does not push the donor out of the workplace but instead strengthens the donor’s capabilities, allowing simultaneous virtual participation in several processes (DP3). This creates an understudy for emergencies and a sparring partner for coevolution.

4.4. Clone’s cognitive development by ML

The training environment for Pi-Mind agents is built to ensure the implementation of the declared design principles (Table 2), namely DP0 (Turing principle or the minimal donor-clone deviation), DP1 (transfer of specific responsibilities to the agent), and DP2 (resiliency or ability to act both reactively and proactively):

- (1) A Pi-Mind agent learns to act as a smart cognitive system capable of creativity, cognition, and computing in real-world settings, where it is impossible to predict future tasks and

problems. The agent should observe new realities and be capable of generating new alternatives and parameters. Therefore, an agent’s cognitive capabilities are developed primarily through adversarial learning using variations and enhancements of the generative adversarial network (GAN) architecture (Goodfellow et al., 2014).

- (2) A Pi-Mind agent must react similarly to the target (human) donor. We suggest updating the basic GAN architecture by finding a place for the donor (human) component. We call this mix (agent + donor) the “Turing discriminator” (TD), and all the GAN architectures that are included in this mix will have the letter “T” in their acronym. A direct extension of the basic GAN is called the T-GAN.
- (3) A Pi-Mind agent is required to operate not only as a binary classifier (“real” or “fake”), as the basic GAN does, but can also address the complete decision options spectrum (class labels and fake detection). To achieve this, we use some architectural features from SGAN (Odena, 2016), which is an extension of a generic GAN architecture with classification capability. Classification is a kind of decision-making problem that involves choosing a particular class label from those available. Behaviour is also a decision-making problem of choosing a specific action from the available ones.
- (4) A Pi-Mind agent is trained in different contexts to make decisions similar to those of its donor. This means that the agent will learn not only one but several decision models personalised in different contexts, including critical decision-making. We add the letter “C” to the acronym of such context-aware GAN architectures; for example, the direct extension of the basic GAN will be named C-GAN (not to be confused with CGAN, which is the conditional GAN).

The aggregation of these requirements forms a target adversarial architecture for training Pi-Mind agents (clones): T|C-SGAN (see, Figure 3).

T|C-SGAN includes a TD with different semantics than a traditional GAN discriminator. The TD is a small collective intelligence team consisting of a “human” (*H*) and a trainable digital “clone” (*C*). The TD gets the input samples from the reality samples set and is expected to guess the correct decision (class or action label). Additional inputs come in the form of a specific context label, which means that the correct decision derived by the TD must also be appropriate for the given context. *H* and *C* (independent of each other) suggest their decisions (correct labels) for the input. The decisions are compared, and the

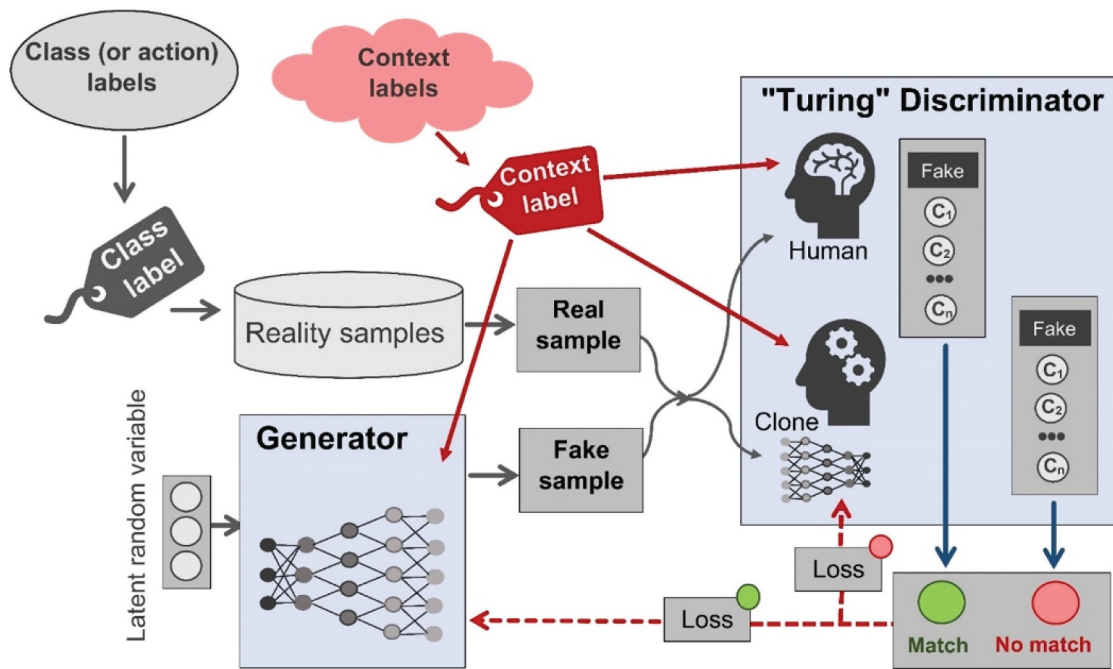


Figure 3. Pi-Mind technology-based T[C]-SGAN adversarial architecture for transferring decision-making skills from a human to a digital clone.

numeric evaluation of the mismatch is computed. This mismatch is used as feedback (loss function value) for the neural network model of C , and the network parameters are updated, aiming for better performance next time. Thus, this procedure comprises supervised context-aware (backpropagation-driven) learning by C , where the clone is trained to guess the decisions made by its donor in similar situations and contexts as precisely as possible.

4.5. Proactivity training with adversarial ML

A more challenging case occurs when the information system is trained in adversarial or conflicting interactions to copy human decision-making behaviour in complex emergencies. Observing the donor, the information system captures and configures the donor's hidden critical decision-making model. We consider the concrete (but hidden) heuristics and biases people use to make personal judgements or choices in cases of ignorance or uncertainty (Tversky & Kahneman, 1974). Cognitive clones must capture personal specifics while learning from donors. This type of learning is based on adversarial ML (Bai et al., 2021; Kumar et al., 2020; Kurakin et al., 2016), in which situations with maximum ignorance and uncertainty are generated and addressed synchronously by humans and clones. Using this principle of (inexplicable) knowledge extraction, even in complex adversarial conditions, the information system copies an individual model of the donor's behaviour without understanding and explaining its parameters.

The GAN philosophy requires that the training process be executed in adversarial conditions, assuming that a challenging training environment facilitates attaining the intended learning outcomes. Therefore, the "generator" (G) adds an adversary component to the architecture. G constantly challenges the TD ($H + C$ team) by aiming to generate input samples similar to those from the reality samples set to confuse the TD . The goal of G is to maximise the mismatch or difference between the H and C reactions. The H and C teams are expected to synchronously address the inputs (classify) and uncover fake inputs. G is also a trainable and neural network-driven component. If G 's content cannot confuse the TD , then G receives feedback (as the loss function value), and the parameters of G will be updated accordingly. With time, G improves its adversarial performance. This "game" (TD vs. G) drives the process of the coevolution of TD and G towards perfection in their competing objectives. Improvement of the TD implies that C learns how to make the same decisions as H (possibly even incorrect ones) in the same situation and the same context and while under pressure (if there are any). This training ensures the Turing principle of minimal donor-clone deviation for business-as-usual and critical decisions.

The quality of the resulting artefacts regarding their fit with the declared DPs could be assessed according to the metrics used in ML, as shown in Table 2. However, one crucial advantage of ML is the possibility of applying quality assurance along with quality assessment. Design quality assurance requires correspondence between the artefact itself and the declared

principles. The additional requirement concerns designing the artefact to enable the production of the artefact with the declared principles when any noticed deviation of the process from the intended one is self-corrected according to the feedback from real-time process monitoring.

More specifically, the feedback provided by neural networks during the training process is a value of the so-called loss function. The process of self-correction based on feedback is called backpropagation (Amari, 1993). Neural network training environments and algorithms ensure self-design and self-evaluation with these instruments. It is important to note that the *TD* provides two feedback loops: one to the clone being trained and one to the *G*, which plays the role of a challenger in the training process (see, Figure 3). Regarding the first feedback loop, the *TD* in the architecture plays the role of a digital measuring device to monitor the fitness of the current design iteration to DP0 (minimal donor-clone deviation). At each iteration of the cloning process, the *TD* outputs the value of a loss function (quality assurance measure regarding DP0), which is a precise assessment of the deviation and is used as feedback for both *TD* and *G* via backpropagation. The cloning (particularly clone training) process stops when the loss function tends to zero, meaning that DP0 is satisfied.

Regarding the second feedback loop, the *TD* in the architecture plays the role of a digital measuring device to check the fitness of the current design iteration to the DP2 design principle (resiliency or the capability of a clone to adapt or recover being under pressure of complex, challenging situations or adversarial attacks). The organisation of the corresponding GAN architecture guarantees that *G* improves its performance in generating challenging (adversarial) inputs to the *TD* at each iteration of the cloning process. Therefore, the feedback (loss function value), which *G* gets, actually evaluates the quality of the challenge for the *TD*, and (because the clone learns synchronously to address the challenge), this feedback can be used to ensure the quality of DP2. This is because the *TD* is learning to adapt to potentially the most challenging inputs and – by doing this – performs as a resilient decision-maker.

Finally, the donor (a human being cloned) in the architecture also receives some feedback (a kind of reward or punishment) from the environment as a response to the decisions she has made, and this feedback reinforces the human's ability to change (adapt or upgrade) from some hidden model that drives her choices. This evolution indicates how seriously the person takes personal responsibility for their own decisions. Personal responsibility is challenging to measure directly from a human perspective. However, suppose we undertake lifelong retraining of the clone (Crowder et al., 2020) according to the

architecture shown in Figure 3. In that case, the clone will coevolve with its donor and perform with the same level of responsibility. Therefore, the suggested GAN architecture will also guarantee the fitness of the design process to the principle of DP1 (responsibility). Finally, the clone (as a trained neural network) can be easily copied and used as an autonomic decision-maker within several processes where the same decision task is required. Consequently, the average number of synchronously running clones from the same donor could be used to measure ubiquity (DP3).

4.6. An example of cognitive cloning

Let us assume that we want to train the clone to make the same decisions as the donor when facing the same decision problems within the same decision context in the future. Figure 4(a) illustrates a simple explicit donor-clone knowledge transfer case. Here, we have a two-dimensional decision space, meaning that each decision task is defined by two parameters (x, y). We also have two different decision options (“YES” or “NO”).

In Figure 4(a), the donor is supposed to know the rules for addressing particular types of decision cases. Each rule is a kind of formal definition of the bounded subspaces within the decision space, corresponding to each decision option (the area of “YES” decisions and the area of “NO” decisions in the figure). Given the parameters (coordinates) of the decision task (a point within the multidimensional decision space), the donor applies the rules to locate the point within one of the decision subspaces (decision options) to make the corresponding decision. Even for popular decision problems, each donor may have specifics in their decision boundaries, meaning that different donors may make different decisions in some cases. Therefore, knowing the explicit definitions of the decision boundaries (decision rules) allows the donor to design the clone top-down via explicit personal decision skill transfer.

Humans may not always know exactly how and why they make certain decisions. To capture hidden decision skills, one needs to “interview” the target donor on many decision cases and collect the chosen options for each case. The collected samples can then be used as training data for various computational intelligence techniques, drawing the decision boundaries and capturing the rules for designing the clone in a bottom-up (ML-driven) way. This option is illustrated in Figure 4(b). Here, based on several cases of “YES” and “NO” decisions, some ML algorithms draw the decision boundary (the separation curve between the “YES” and “NO” decision subspaces). After this, new decision-making cases can be addressed accordingly. This method of cloning entails a kind of supervised learning, wherein the donor labels the set of

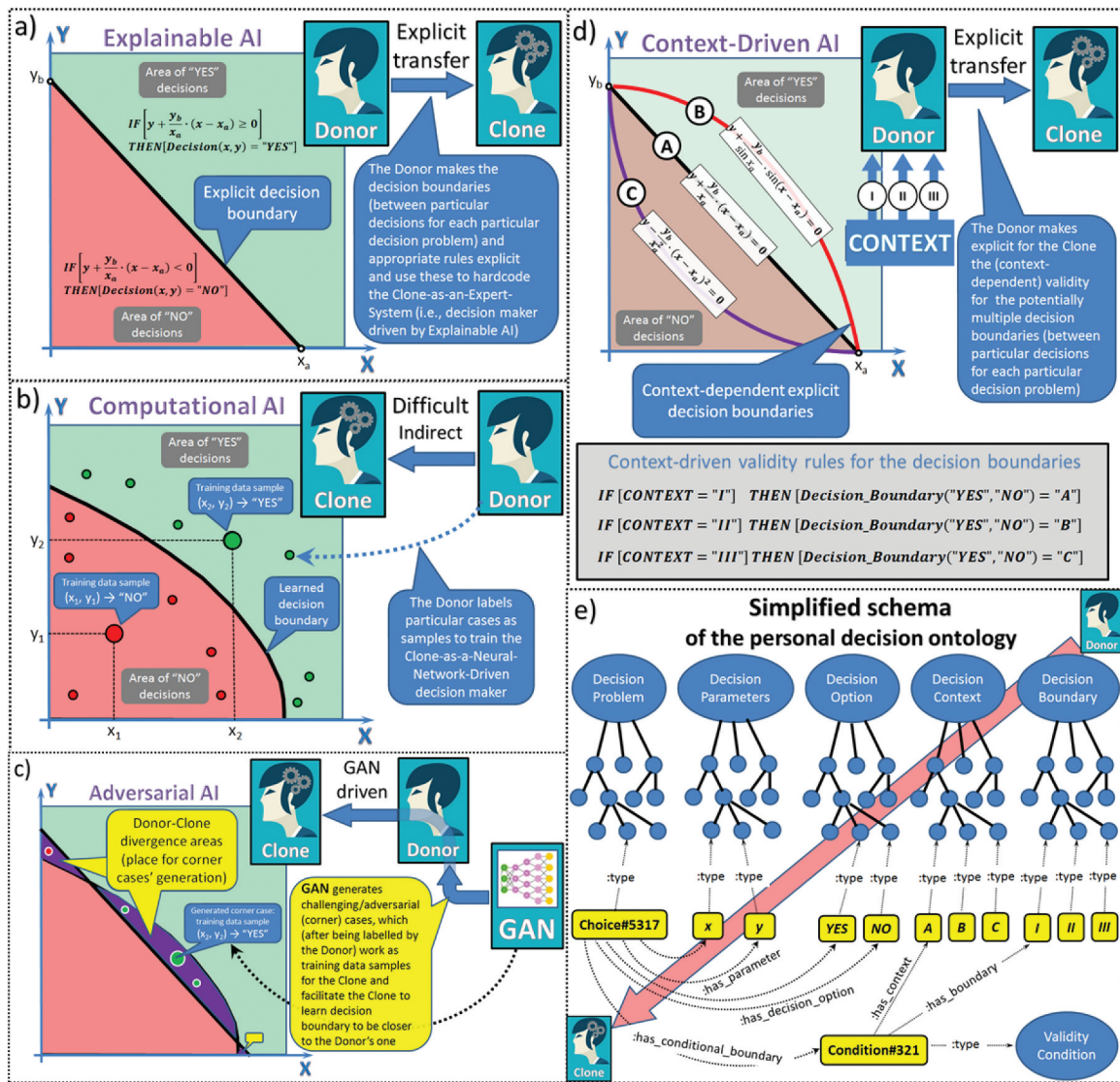


Figure 4. Challenges regarding cognitive cloning: a) explicit (donor to clone) knowledge transfer (e.g., as a set of decision rules); b) machine learning-driven training of the clone (donor labels the particular decision contexts, and the clone learns the boundaries between different decision options by discovering the hidden decision rules of the donor); c) adversarial learning (driven by GANs) helps facilitate the training process in b) by discovering the corner cases for challenging decisions, hence making the clone's decision boundaries and rules closer to the donor's; d) discovering and making explicit the contexts that influence the donor's decision boundaries and rules and training the clone specifically for all such contexts; e) integrating both explicit and learned decision knowledge into different decision tasks and decision contexts under the umbrella of personal decision ontology, which will be used by the Pi-Mind agent when acting as a clone of a particular human donor.

decision cases with the chosen decision option, and the clone learns based on this set (by guessing the hidden decision boundary that the donor uses when making decisions).

Such supervised ML depends heavily on the training data. Let us assume that the actual (but hidden) decision boundary for a particular donor is the line shown in Figure 4(a). However, some ML algorithms (e.g., neural network backpropagation learning) draw the curve as the decision boundary based on the labelled data, as shown in Figure 4(b). The donor will continue to make further decisions with different boundaries from the clone. As shown in Figure 4(c), the difference between actual and guessed decision boundaries creates some

divergence areas within the decision space; if they happen to belong to these areas, all the decision tasks will be addressed differently by the donor and clone. To minimise discrepancies between the donor's and clone's opinions, we have to provide the clone with better training data. We used GANs for this purpose because they are capable of discovering (within the real-time training process) divergence areas and generating new (corner) cases (to be labelled by the donor) from these areas. This type of (adversarial) training, as illustrated in Figure 4(c), facilitates the learning of precise decision boundaries and makes the clone capable of making almost the same decisions as the donor would.

Humans often make decisions differently in different contexts. Consequently, if the clone learns the donor's decision logic (appropriate decision boundaries and corresponding decision rules) in one context, this will not mean that the same logic would work in another context for the same set of decision tasks. Figure 4(d) illustrates the context-dependent decision boundaries and the appropriate (meta-) rules. When a particular decision context is known, the particular decision boundary (and hence the corresponding decision rules) becomes valid (according to explicitly defined meta-rules) and will be used for further decisions. In the most complicated cases, the cloning challenge would mean that both the hidden decision rules and hidden context meta-rules of the donor must be learned bottom-up using adversarial, context-aware, GAN-driven techniques, as we described earlier and illustrated in Figure 3.

All these decision-making skills, which the intended clone either gets explicitly from the donor or learns by observing the donor's decisions for different decision problems and contexts, must be integrated as an interconnected set of capabilities controlled by the Pi-Mind agent. This content (the taxonomies and semantic graphs of acquired or learned decision problems, parameters, options, contexts, boundaries, rules, and meta-rules that characterise the decision-making specifics of the donor) is constructed automatically under the umbrella of the personal decision ontology, as shown in Figure 4(e). Semantic (machine-processable) representation on the top of decision models allows for automated processing (by the Pi-Mind agent-driven clone), seamless integration of available decision-making knowledge and skills, openness to lifelong learning of new knowledge and skills (via continuous observation of the donor), and communication and coordination between intelligent agents.

5. Demonstration and evaluation of the Pi-Mind agent

5.1. The specifics of the Pi-Mind demonstration and evaluation

Appropriate conditions must be met to use a Pi-Mind agent (a virtual ecosystem with support for all Pi-Mind options). Therefore, launching a fully functional version without preliminary testing of the individual components and tools is too complicated, expensive, and risky. We approached the solution to this problem as follows: 1) through diversification with a pilot launch of the Pi-Mind technology in different procedures (we provide examples called the NATO case, NURE case, and HE case); 2) by leveraging the ecosystems of their previous projects; and 3) by using the minimum

viable product (MVP) (Nguyen-Duc, 2020) approach, not only for development but also to demonstrate the value of the agents (different Pi-Mind options were selected as MVPs). We also used the opportunity to “play” in different contexts:

- Complex contexts (e.g., in security systems) are of special interest because of their difficult and ambiguous tasks, but they also carry significant risks and require enormous resources. We used field testing at a local scale to achieve validation through repeatability (using the test-retest reliability as an indicator of the same-tests – same-results similarity).
- Contexts with simple tasks without excessive risks are also interesting because they make it possible to validate the scalability of a concept.

The NATO case is in the security systems domain (a real laboratory with a real system, real users, and simulated problems). Complex implicit knowledge transfer based on continuous adversarial learning cannot be checked in a social environment. The requirement of large, well-formed, validated datasets for comprehensive training and badly formalised decision processes forced us to turn to the ongoing project funded by NATO Science for Peace and Security (NATO SPS; <http://recode.bg/natog5511>; Terziyan et al., 2018a). We aimed to prove that Pi-Mind agents could be trained to enhance civil and military security infrastructures and to take over control of certain operations on behalf of security officers. The agents “observed” the interroll cassette conveyor, which is an analogue of those used in airports for distributing and inspecting luggage. Their task was to prevent any potential danger caused by cassette loads on the conveyor by applying expert donors' judgement and expertise. Two types of decision processing contexts based on image recognition were tested: in the business-as-usual environment and in adversarial conditions (DP2), which would cause disruptions to the critical infrastructure. A detailed description of the previous and current experiments is available in Golovianko et al.'s (2021) study.

Although the human decision-makers showed better performance in threat recognition, the experiments' results are promising because of the high accuracy of the artificial predictions and (even better) high human-clone correlation regarding the decisions. Experiments with different adversarial conditions showed that artificial and human workers tended to misclassify threatening objects in adversarial settings. However, the agents can be trained to develop a new capability – an artificial cognitive immunity – which can help improve the accuracy of human donors' evaluations by giving artificial advice.

The NURE case is used in the real business processes of organisations, with real users, problems, and information systems. The education domain provides an opportunity to demonstrate the Pi-Mind agent in a social environment with many users in simple contexts. Some of the quality assurance (QA) processes at Ukrainian universities were executed through the TRUST Portal (<http://portal.dovira.eu>), an academic social media and process management platform (see, Figure 5) with ontology-based information storage and inference mechanisms on top (Terziyan et al., 2015).

To introduce Pi-Mind agents into the portal, we transferred explicit knowledge from the donors to the Pi-Mind agents to implement a basic decision-making procedure when choices were made based on a comparative numeric evaluation (ranking) of the registered or available options. The core of this procedure is a personal system of values (PSV) consisting of parameters for the evaluation function in personalised decision-making. Combined with agent technologies and implemented in the semantic environment, a PSV can become a proactive entity capable of accessing digitised information, operating automatically as a personal clone, and participating in collective decision-making.

Over the past few years, the Pi-Mind agent has been used as a decision-making tool on the portal. As the most active user, Kharkiv National University of Radio

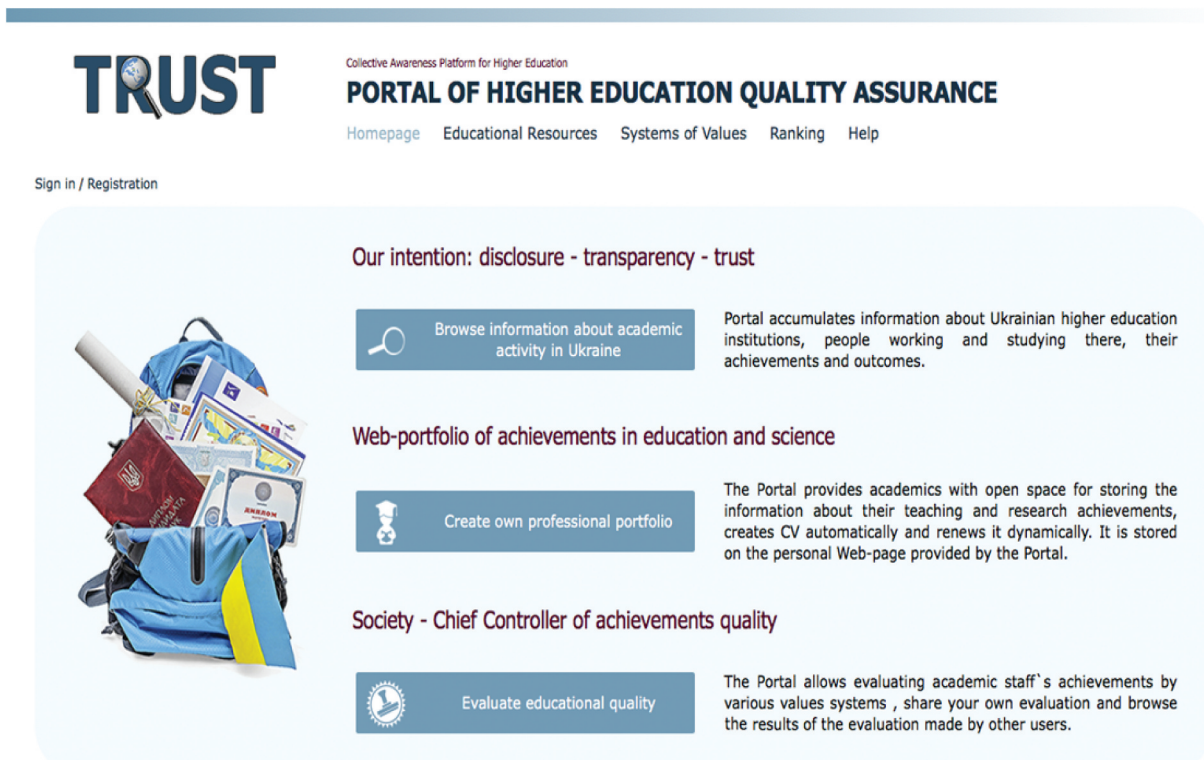
Electronics (NURE) – the leading Ukrainian IT university – was the first to implement this technology in its business processes (see, Table 3).

Next, we present the evaluation of the Pi-Mind agent, showing that the use of the agent meets the following goals: 1) effectiveness, given its improved decision accuracy (proof of concept), and 2) efficiency, given the decreased time for decision-making (proof of value) and the digital transformations in decision-making in HE (proof of use).

5.2. Proof of concept: increasing the accuracy of decisions

In the “employee motivation” in the NURE case (administrative decision-making, see, Table 3), to distribute bonuses among academic staff, the rector – an authorised decision-maker – annually analyzes the achievements of each employee for a certain period. Based on this analysis, the rector either increases the employee’s monetary bonus (for good results regarding certain objectives) or reduces the bonus (for decreased efficiency). The application of Pi-Mind agents significantly changed the parameters of this procedure (see, Tables 4 and 5).

The co-focusing of the bonus objectives (multi-dimensional vector) and actual employee development vector determines the utility of the decision. The effectiveness of the decision (or the accuracy) is assessed following the assessment practices used




TRUST
Collective Awareness Platform for Higher Education

PORTAL OF HIGHER EDUCATION QUALITY ASSURANCE

Homepage Educational Resources Systems of Values Ranking Help


Sign in / Registration

Our intention: disclosure - transparency - trust

 Browse information about academic activity in Ukraine


Portal accumulates information about Ukrainian higher education institutions, people working and studying there, their achievements and outcomes.

Web-portfolio of achievements in education and science

 Create own professional portfolio

The Portal provides academics with open space for storing the information about their teaching and research achievements, creates CV automatically and renews it dynamically. It is stored on the personal Web-page provided by the Portal.

Society - Chief Controller of achievements quality

 Evaluate educational quality

The Portal allows evaluating academic staff’s achievements by various values systems, share your own evaluation and browse the results of the evaluation made by other users.

Figure 5. TRUST Portal, <http://portal.dovira.eu>.

Table 3. Evaluation of the Pi-Mind agent at NURE.

| Type of proof | Proof of concept (Business as usual) | Proof of concept (Crisis management) | Proof of value (Business as usual) |
|---|---|---|---|
| Decision problem | Administrative decision-making | Provoked by COVID | Participative decision-making |
| Type of problem | Employee motivation | Extreme resource reconfiguration | Recruitment |
| Procedure | <i>Distribution of bonuses among academic staff.</i> Practices of various manipulations with the lists of applicants for the award are replaced by generation and publicising the applicants' rankings based on actual achievements proactively assessed by the rector's PSV (launched in 2017) | <i>Redistribution of resources without the direct involvement of the decision-maker.</i> Personal participation of the decision-maker is replaced by the clone, which is capable of suggesting a resource allocation structure among the real and distant processes based on the PSV (launched in 2020) | <i>Job candidate selection:</i> Meetings of a hiring selection committee are replaced by an automated "vote" as an aggregation of ranking lists (proactive assessments of the applicants' achievements by the PSVs of each board member) (launched in 2019) |
| Number of use cycles | 3 completed cycles 1 in the process | 1 completed cycle | 1 completed cycle |
| Donor(s) | University rector | University rector in restricted conditions for personal decision-making | Members of the academic council (at the faculty level) |
| The number of clones involved | 3 Rector's Pi-Mind-2017 Rector's Pi-Mind-2018 Rector's Pi-Mind-2019 | 1 Rector's Pi-Mind-2019 (for business as usual), which was available "at hand" and had suitable criteria | 23 23 Pi-Minds of academic council members |
| People involved in clone-driven processes | 367 (1st cycle) 428 (2nd cycle) 501 (3rd cycle) | 2,118 (all university personnel) | 18 (job candidates) |
| Impact Benefit | Higher responsibility (i) Improving the accuracy of decisions (ii) Activation of personnel interest | Higher resilience Improving the accuracy of decisions made | Higher ubiquity Decreasing the decision time |
| Evaluation (main metrics) | (i) F1-score: 0.49 (before) 0.96 (after) (ii) Site traffic growth: 3–5 times | F1-score 0.95 | More than 1,500 human hours per year (for the whole university) |

Table 4. Results of the decisions regarding the distribution of bonuses among academic staff before using the Pi-Mind agents.

| Parameters | Award Committee data | Portal data (rankings) | 2016 |
|------------------------------|-----------------------------|-------------------------------------|------|
| Total population | Awarded staff members | | 467 |
| True positive | Award increased | Ranking score increased | 105 |
| False positive | Award increased | Ranking score decreased | 195 |
| False negative | Award decreased | Ranking score increased | 27 |
| True negative | Award decreased | Ranking score decreased | 140 |
| Predicted condition positive | All with an increased award | x | 300 |
| Predicted condition negative | All with decreased award | x | 167 |
| Condition positive | x | All with an increased ranking score | 132 |
| Condition negative | x | All with a decreased ranking score | 335 |

in a binary classification (Tharwat, 2021), and it is evaluated based on a confusion matrix and measured through the F1-score, a metric for model performance that combines precision and recall (see, Table 6).

Thus, using the Pi-Mind agent also increased resilience in the rector's workplace: accidental and deliberate errors arising from information overload, lack of time, or the malicious influence of outside forces on the decision-making process were excluded. Neither advocates nor opponents of the new procedure supported by the Pi-Mind agent could reproach the rector with the claim of biased or non-transparent distribution of the bonuses.

In the other example in the NURE case, "extreme resource reconfiguration" (see, Table 3), the announcement of the COVID-19 quarantine required the university to make an urgent transition to crisis management mode and reconfigure resources for the on/off-line processes accordingly. Although the university rector was not available at that moment due to a business trip, the university administration quickly managed to get the rector's opinion on the reallocation of resources from his digital clone (proactive PSV). Although this clone was created for other business processes, it reflected his administrative preferences. With the help of the Pi-Mind agent, the university managed to restructure and adapt its processes completely

Table 5. Results of decisions for the distribution of bonuses among academic staff after using Pi-Mind agents.

| Parameters | Ranking-Based Awards | Data of Appeal Committee | 2017 | 2018 |
|------------------------------|--|--|------|------|
| Total population | Awarded staff members | | 367 | 428 |
| True positive | Rank increased. Award increased | x | 171 | 228 |
| False-positive | Award increased. The increase was later cancelled because of appeals | Rank decreased because of incorrect data (social verification) | 12 | 15 |
| False-negative | Award decreased. The decrease was later cancelled because of appeals | Rank increased because of additional data (previously unseen) | 9 | 4 |
| True negative | Rank decreased Award decreased | x | 175 | 181 |
| Predicted condition positive | All with increased awards (based on ranking) | x | 183 | 243 |
| Predicted condition negative | All with decreased awards (based on ranking) | x | 184 | 185 |
| Condition positive | x | All whose recorded achievements were evidently increased | 180 | 232 |
| Condition negative | x | All whose recorded achievements were evidently decreased | 187 | 196 |

Table 6. Assessment of the decisions for the distribution of bonuses among academic staff.

| Parameters | Parameter value | 2016 | 2017 | 2018 |
|----------------------------|---|-------|--------|-------|
| | | After | Before | |
| Sensitivity | True positive/Condition positive | 0.8 | 0.95 | 0.98 |
| Miss rate | False negative/Condition positive | 0.2 | 0.05 | 0.02 |
| Probability of false alarm | False-positive/Condition negative | 0.58 | 0.06 | 0.08 |
| Selectivity | True negative/Condition negative | 0.42 | 0.94 | 0.92 |
| Prevalence | Condition positive/Total population | 0.28 | 0.49 | 0.54 |
| Precision | True positive/Predicted condition positive | 0.35 | 0.93 | 0.94 |
| False omission rate | False negative/Predicted condition negative | 0.16 | 0.05 | 0.02 |
| Accuracy | (True positive + True negative)/Total population | 0.52 | 0.94 | 0.96 |
| False discovery rate | False-positive/Predicted condition positive | 0.65 | 0.07 | 0.06 |
| Negative predictive value | True negative/Predicted condition negative | 0.84 | 0.95 | 0.98 |
| Positive likelihood ratio | Sensitivity/Probability of false alarm | 1.37 | 14.8 | 12.84 |
| Negative likelihood ratio | Miss rate/Selectivity | 0.49 | 0.05 | 0.02 |
| Diagnostic odds ratio | Positive likelihood ratio/ Negative likelihood ratio | 2.79 | 277.08 | 687.8 |
| F1-score | $2 \times (\text{Precision} \times \text{Sensitivity}) / (\text{Precision} + \text{Sensitivity})$ | 0.49 | 0.94 | 0.96 |

to a remote mode in all the activity areas and was able to do so within three working days. This case demonstrates that the ubiquitous Pi-Mind agent can involve HR, even when the employee is temporarily unavailable. This reduces the risk of interruptions in the workflow (especially in critical circumstances). The Pi-Mind agent protects employees from losing their jobs in the case of temporary unavailability.

The accuracy of the decisions supported by the Pi-Mind agent during the COVID-19 crisis was assessed by comparing the regulatory documents adopted (cancelled or changed) during the transition process (Table 7). Between March 12 and 16, 2020 (three working days), 20 orders were approved, and an additional 31 official instructions were released regarding the transfer of the university processes to a remote work mode. One order was later cancelled, and two orders were changed. Supplements were issued regarding one order and one instruction. Seven drafts remained unapproved.

5.3. Proof of value: decreasing the time needed for decision-making

Our third example in the NURE case (“recruitment”; see, Table 3) shows that using a Pi-Mind agent improves the system’s efficiency with the organisation’s resources. The decisive (voting) stage of the “job candidates’ selection” procedure includes a meeting of the hiring selection committee, wherein members of the academic council are familiarised with summaries of the processed documents for each candidate. A private vote then takes place at the academic council meeting. A Pi-Mind agent launched at this decision-making point duplicates the HR (i.e., the voter is only a professional expert) and is not subject to being influenced by the situation (like a vulnerable human voter). Such a decision becomes explainable and reviewable, and, accordingly, responsible. Another benefit is that such decision-making requires significantly less time. The former procedure required at least 10 minutes to review each candidate. Involvement of the Pi-Mind agents minimises the time spent searching for an

Table 7. Assessment of the decisions for the redistribution of resources without the direct involvement of the decision-maker.

| Parameters | Parameter value | 2020 |
|------------------------------|---|------|
| Total population | Quarantine-related orders | 60 |
| True positive | Timely high-quality orders | 48 |
| False-positive | Orders activated on time but changed later | 3 |
| False-negative | Orders activated later | 2 |
| True negative | Orders prepared but not activated | 7 |
| Predicted condition positive | Orders activated in critical situation (Pi-Mind clone result) | 51 |
| Predicted condition negative | Cancelled orders (Pi-Mind clone result) | 9 |
| Condition positive | Really necessary orders | 50 |
| Condition negative | Orders appeared to be unnecessary | 10 |
| Sensitivity | True positive/Condition positive | 0.96 |
| Miss rate | False-negative/Condition positive | 0.04 |
| Probability of false alarm | False-positive/Condition negative | 0.3 |
| Selectivity | True negative/Condition negative | 0.7 |
| Prevalence | Condition positive/Total population | 0.83 |
| Precision | True positive/Predicted condition positive | 0.94 |
| False Omission rate | False-negative/Predicted condition negative | 0.22 |
| Accuracy | True (positive + negative)/Total population | 0.92 |
| False discovery rate | False-positive/Predicted condition positive | 0.06 |
| Negative predictive value | True negative/Predicted condition negative | 0.78 |
| F1-score | $2 \times (\text{Precision} \times \text{Sensitivity}) / (\text{Precision} + \text{Sensitivity})$ | 0.95 |

acquaintance with the documents, allowing for the creation of individual ranking lists for candidates without a separate meeting. As a result, the time for approving each candidate was reduced to one minute. Nine minutes saved per transaction resulted in tangible savings throughout the university (Table 8).

5.4. Proof of use: The digitally accelerated transformation of Ukrainian HE decision-making and society at large

Proof of use of the Pi-Mind agent was performed in numerous case studies to justify the current artefact's applicability in various domains, as promoted by Nunamaker et al. (2015), to identify its use's side effects or undesirable consequences (Venable et al., 2012). The transformation from specific, one-time decisions to more complex, fuzzy decision models allowed us to reveal the real scope of the artefact's

implicit and explicit capabilities and, therefore, its potential to influence change management processes and collective decision-making.

The HE case impacts the national level with real users, problems, and information systems. The Pi-Mind agent was implemented as part of the digital infrastructure, contributing to accelerating national reforms and managing change. Social change happens as a set of transitional processes that qualitatively shape different systems and communities. Even developed countries often experience difficulties when reforming because of the complex, nonlinear, dynamic, and difficult-to-predict nature of the underlying transition processes. In post-Soviet developing countries, such as Ukraine, unfavourable "starting positions" make this challenge even greater. Our goal was to use digital cloning (both decision-makers and processes) to enhance the efficiency of transitional processes, decrease the level of corruption, and increase societal trust in the process participants. We studied the effect of the Pi-Mind agent on the educational domain in Ukraine in several phases.

First, the web-based platform TRUST (www.portal.dovira.eu) was launched in four universities (NURE, Yuriy Fedkovych Chernivtsi National University, Ukrainian Catholic University, and the National Academy of Managing Personnel of Culture and Art) with the support of the Ministry of Education and Science of Ukraine. A meta-procedure for QA procedures was developed and tested on five procedures at each of these universities: management of academic recruitment, academic staff assessment, and motivation; innovation management at the level of HE institutions; student feedback management; internationalisation;

Table 8. Savings in human workload per year for the whole university.

| Parameters | Academic councils at the faculty level | Academic councils at the university level |
|--|--|---|
| Number of academic councils | 8 | 1 |
| Number of academic council members | 23 | 48 |
| The average number of candidates considered by the academic council per year | 20 | 140 |
| Savings of human hours per transaction | 0.153 | 0.153 |
| Savings of human hours per year | 563 | 1,028 |
| Savings of human hours per year for the university | | 1,591 |

academic networking management; and management of academic staff's lifelong learning. All the developed procedures were properly tested and documented (Semenets et al., 2021). The first phase of implementing the Pi-Mind agent showed that even a reasonably constructed QA system required the support and commitment of its players. This conclusion was confirmed by the Organization for Economic Cooperation and Development (Melchor, 2008). Transparent, and the systematic use of the Pi-Mind agent has become a support for the agents of change in universities, addressing the lack of academic consolidation, awareness, and acceptance of the reforms.

The second phase of evaluating the proof of use was shaped by deploying an entire ecosystem for accelerated cognitive development. This allowed cloning processes that could simulate new scenarios to predict the system's behaviour, processes, and individual players within the new "rules of the game", and then publish and compare the results. The Pi-Mind agent was first officially implemented at the national level during the election of private HE institutions' representatives to the first national independent QA agency in Ukraine in 2015. The congress of private HE institutions used the tools and services of TRUST as digital support for the election process. The Pi-Mind agent (as a carrier of the collective vision) evaluated the candidates' applications, choosing the most qualified members and publishing the final scores so that everybody could verify them and check the validity of the achievements. An attempt by the Ukrainian Congress to ignore the results and push previously chosen candidates behind closed doors was described in detail by the independent analytical platform VoxUkraine. The report on its analysis of the election results was later published with the

title "Reform of HE: One Step Ahead and Two Steps Back". The article illustrated (step by step with screenshots) that the best candidates (the Pi-Mind agent's decisions) and the nominated candidates (decisions of the officials) were different. The Ministry of Education and Science (responsible for the elections), under pressure from the solid and transparent facts provided by the Pi-Mind agent, subsequently annulled these nominations. For more than seven years, the Pi-Mind agent has continued to be used in various processes, helping to adapt people's mindsets, culture, attitudes, and practices to new environments.

The results of the functioning of the ecosystem (the TRUST Portal) with the support of Pi-Mind technology are as follows:

- The ontological knowledge base stores more than 400,000 resources registered by users, and 23 million knowledge triples. Knowledge (semantic) triples (subject-predicate-object statement) are elementary units that define one connection between two entities (often called "resources") within a shared ontological knowledge graph (often called a "semantic graph").
- More than 5,000 individual and corporate users (with advanced usage powers) were registered on the portal.
- More than 8,000 academic achievements were registered on the portal.
- More than 500 Pi-Mind agents, with their value systems, were created on the portal.
- More than 1,700 procedures were launched by Pi-Mind agents and stored on the portal.

The artefact was rigorously evaluated for its effectiveness and efficiency in business-as-usual and crisis management processes for the design principles (DP0–3) used for its design (Table 9).

Table 9. Evaluation of the effectiveness and efficiency of the Pi-Mind agent and the design principles DP0–DP3.

| Essence | Design Principle | Metric | Value |
|---------------|-------------------------|---|---|
| Effectiveness | DP0 Turing principle | F1-score in business as usual processes | Increase from 0.49 to 0.96 – the correlation between artificial and human academic staff in the NURE case (Table 6). From 0.92 to 0.995 – the correlation between artificial and human security inspectors in the NATO case (Golovianko et al., 2021). |
| | DP1 Responsibility | The percentage of responsibilities delegated to a clone | 100% – because of the stored PSVs in the information system (in the NURE case). > 85% in business-as-usual processes in the NATO case (Golovianko et al., 2021). > 65% in adversarial environments in the NATO case (Golovianko et al., 2021). |
| | DP2 Resiliency | F1-score in critical processes | 0.95 – in benign environments (NURE case), Table 7. Increase to 0.7 in adversarial environments, here in the case of evasion attacks in the NATO case (Golovianko et al., 2021). |
| Efficiency | DP3 Ubiquity | Human hours savings | > 1,500 human hours for one procedure in an organisation per year (see, Table 8). |
| | | Number of processes run with Pi-Mind agents | > 1,700 processes in a real environment. |

The performed test runs convinced us that not only personal decision expertise but also entire processes can be cloned. In cloning processes, by placing different Pi-Mind agents at the appropriate decision points, the organisation can ensure the continuity and quality of the processes based on collective intelligence and dynamic assessments. These findings led us to work on twinning and simulating transitional processes to test various expert approaches, which are extremely important in situations with limited human and time resources. More specifically, the Pi-Mind agents can be considered key providers of the processes' sustainability in the context of hybrid threats in the environment created in the "Academic Response to Hybrid Threats" (WARN) Project (<https://warn-erasmus.eu>). We are currently simulating complex situations in which a cognitive clone (or an "artificial clone + human donor" team) learns to make decisions in real processes when challenged by complex, non-standard adversarial conditions.

6. Discussion

6.1. Use prospects for the artefact

Our study makes a novel artifactual contribution to the research (Ågerfalk & Karlsson, 2020) on improving organisational resilience in points of decision-making supported by information systems (Duchek, 2020; Heeks & Ospina, 2019; Hillmann & Guenther, 2021; Linnenluecke, 2017; Riolli & Savicki, 2003). Global crises provoked by "unknown unknowns" have made this task especially important because situations such as the COVID-19 pandemic have revealed the considerable vulnerability of decision-makers in critical situations. The search for appropriate solutions led to AI technologies, enabling a smooth transformation from human-driven to automated decision-making.

The IT artefact (Pi-Mind agent) is an AI solution designed and implemented to create digital clones of professional experts. With its basic concepts (cognitive clone and cloning environment), models, methods, techniques, interfaces, and tools, the agent will ensure responsible, resilient, and ubiquitous decision-making in business-as-usual and crisis management conditions.

The applicability and utility of the IT artefact were demonstrated and evaluated via several successful implementations and deployments. First, we presented the design and implementation of the top-down facet of the Pi-Mind agent within organisations in the university environment. The administrative and participative decision-making processes were reformed based on this implementation, but the critical decision-making processes also

changed completely. The implementation assessments confirmed an increase in the accuracy of HR decisions and savings, with an increase in personnel's general interest in the new technology. Second, the bottom-up facet of the Pi-Mind agent was implemented based on new GAN-oriented architectures. Third, the implementation of the Pi-Mind ecosystem in Ukrainian HE institutions has led to several societal and legislative changes:

- Several initiatives for improving regulatory and legislative acts were successfully adopted at the level of the Cabinet of Ministries,¹ the Parliament Committee of Education and Science,² and the Ministry of Education and Science.³
- Members of the project team participated in activities for the development of the new law on HE (September 2014).
- Expert recommendations for the draft of the concept of reforms in the system of accreditation and licencing of HE institutions and a roadmap for establishing the National Independent QA Agency were presented.⁴

Therefore, we claim that our "last-mile DSR" has directly impacted Ukrainian society and how decisions are made in Ukrainian HE institutions. Consequently, our study has implications for practice and provides an example of how proof of use can be evaluated in a DSR project.

Furthermore, the developed DPs make a novel contribution to the development of both responsible AI (Arrieta et al., 2020; Gupta et al., 2021) and ethical AI (see, e.g., the Montreal Declaration, 2017: montreal-declaration-responsibleai.com/the-declaration; the Asilomar AI principles: futureoflife.org/ai-principles; Berente et al., 2021; Floridi & Cowls, 2019; Henz, 2021). Regarding the "responsibility" concept, we consider three dimensions: "taking over a responsibility", "being responsible", and "deciding responsibly". "Being responsible" involves the ethical dilemma of who will be responsible for incorrect decisions made by the clone (Royackers et al., 2018). The DPs create the technical prerequisites for transparent distribution of responsibilities between the donor, developer, and user and enable a good balance between AI autonomy and high control of the donor or of those the Pi-Mind collaborates with. Putting humans at the centre of systems design thinking, the Pi-Mind agent validates Shneiderman's human-centred AI framework (Schneiderman, 2020).

Our study also promotes the concept of "deciding responsibly" as a computational aspect of measurable responsibility for mistakes. We capture and measure personal bias and consider it an indication and a measure of responsibility as the extent to which the

decision-maker is concerned regarding the potential consequences of one's decisions. The study also contributes to solving several ethical issues in the workplace: 1) the Pi-Mind agent provides a guarantee of the virtual presence of a needed specialist without overloading him; 2) the opportunity to participate simultaneously in several critical processes increases an individual's confidence regarding employment; 3) in cases where an employee has exceptionally unique, strong, and potentially reusable expertise, they can "patent" the clone and sell copies elsewhere; therefore, as a technology acronym, "Pi" (or "π" as another option) refers to "patented intelligence".

6.2. Further development of the artefact

The IT artefact (Pi-Mind agent) is a testing ground for extensive research on "human-like" decision-making. For example, donors have different attitudes regarding how they behave in new situations, either choosing a rational (statistically more rewarding) option or bravely trying something new out of curiosity. Clones must also capture these attitudes during training. In Terziyan and Nikulin's (2021) study, the "gray zones" within the data (collected as experiences for future training) were defined as the voids within the decision space. Such zones indicate the boundaries for potential situations for which no decisions have been made. The grey zones can be handled and used for "curiosity-driven learning" when training and testing samples are intentionally generated deep inside the grey zones. This would force the intelligent algorithm (e.g., a potential cognitive clone) to learn faster about how to decide in cases of uncertainty and ignorance, such as the current COVID-19 crisis.

Our study also contributes to the literature on human-centric AI and the change from computational thinking (Wing, 2006) to AI thinking (Zeng, 2013), as well as to human-centric AI thinking (How et al., 2020), human-centric AI (Bryson & Theodorou, 2019), and trustworthy AI applications (García-Magariño et al., 2019; Kaur et al., 2020). More specifically, our study addresses how AI helps respond to emergencies and how human-aware AI (Kambhampati, 2019) should be designed to address these challenges by modelling the mental states of a human in the loop, recognising their desires and intentions, proactively addressing them, behaving safely and clearly by giving detailed explanations of demand, and so forth. We argue that this development will result in the form of collaborative AI, which starts with two former extremes: AI-agent-driven swarm intelligence (Schranz et al., 2020), which is based on nature-inspired collective behaviour models for self-organised multi-agent systems, and human-team-driven collective intelligence (Suran et al., 2020),

which focuses on the search for a compromise among individual (human) decision-makers. Collaborative AI is a compromise of both, thus including human + AI-agent-driven decision-making (Paschen et al., 2020). This benefits from AI but keeps humans in the loop, as in the case of "human swarming" (Rosenberg, 2016), which combines the benefits of an efficient computational infrastructure with the unique values that humans bring to the decision process.

The foreseen development of collaborative AI has implications for many areas of society. One such example is the manufacturing industry's digital transformation towards smart factories ("Industry 4.0") and the use of cyber-physical systems to cybernize manufacturing processes (Tuunanen et al., 2019). Given the increasing role of AI in the digital transformation of industrial processes, one can assume that the supervisory role of humans in future industries will decrease (Rajnai & Kocsis, 2017). However, one way to keep humans in the loop would be to create cognitive clones of humans, as suggested here. The use of cognitive clones of real workers, operators, and decision-makers will preserve the human-centric nature of cybernized manufacturing processes, enabling highly personalised product manufacturing (Lu et al., 2020).

A more everyday area of society related to the current study is the use of self-driving vehicles, which have been seriously impacted by various ethical dilemmas and responsibility distribution issues (Bennett et al., 2020; Lobschat et al., 2021; Myers, 2021). However, if the digital driver (as a clone of the customer) will make ethically similar choices (among the legal ones) as the customer would make in similar critical situations (having enough time to think properly), then the customer-to-vehicle trust would be much higher. Another important case for cognitive clones as "digital customers" would include improving the (digital) customer experience because of customer involvement in the design and manufacturing processes via corresponding digital clones. Such cyber-physical systems would allow them to interconnect the processes related to a cybernized customer experience (Rekettye & Rekettye, 2020; Tuunanen et al., 2019) and supply chain innovation (Hahn, 2020).

Finally, the obvious intensification of various global and local crises of various natures (manufactured disasters, terrorist attacks, refugee crises, global hacker attacks, hybrid and real wars, COVID-19, other pandemics, etc.) play a role as catalysers for the emergence of all these AI-related trends. Therefore, we believe that society today does not need faceless and ruthless AI analytics. On the contrary, people need AI with a "human face" or a sustainable and trusted partner capable of helping humans overcome problems and feel safe in this challenging world. This was the main driver of this study.

7. Conclusions

The synergy of vulnerabilities of different natures has become a real-world crash test for decision-making mechanisms. Most processes supported by information systems have decision branch points driven mainly by humans. The key dispatcher role a human (as a thinker and decision-maker) has in process management makes many believe that this role is the reason for our existence. Descartes's famous "Cogito, ergo sum" ("I think, therefore I am") confirms the vital human need to be involved in decision-making. Any interruption of our active involvement in a cognitive activity is interpreted as a threat to our existence. However, this phrase takes on a new meaning with cognitive cloning technology. We offer an IT artefact (Pi-Mind agent) for the digital duplication of human decision-makers with their unique competencies. The Pi-Mind agent, which can imitate a human donor's decision-making, has an invaluable advantage: the technology is not exposed to the intentional and unintentional biological threats related to, for example, infections, pollution, toxic substances, and so forth. Therefore, the Pi-Mind agent will preserve sustainable cognitive involvement ("existence") for itself and its human donor. Consequently, through the resilience of the Pi-Mind technology, we can modernise the quote with a new interpretation: "My clone thinks when I cannot, therefore, I (still) am".

Pi-Mind as a technology covers the top-down, bottom-up, and autonomic approaches to AI, along with the technology of decision behaviour twinning and the technology of creating decision clones. This allows us to determine how to make a digital, proactive copy of a person's decision system:

- The developed DP1–3 brings flexibility to the solution because of the ability to choose one of the options as the MVP (Nguyen-Duc, 2020) and join the remaining options when appropriate.
- To teach an AI agent to capture personal decision preferences, values, and skills from humans (DP0), we propose a hybrid approach to knowledge extraction: AI can be developed based on data, information, and knowledge taken from humans, or AI can be trained based on the available data; both options can be used alone or in combination with each other.
- To teach an AI agent to (consciously) choose the most appropriate solution among the alternatives at critical decision points, we propose supplementing agent-based technologies with training agents in adversarial environments (within the GAN paradigm).

The developed IT artefact is not simply a digital twin, an intelligent agent, or a decision-making system; it is a more complex IT artefact with heterogeneous features.

The research questions in this study focused on how to ensure organisational resilience using the Pi-Mind agent. Our answer is empowering (and complementing, when appropriate) humans as decision-makers and, at the same time, decreasing the vulnerabilities of human-dependent decision-making processes due to special techniques used for IT artefact design and further adversarial training.

The current study has some limitations. First, the Pi-Mind agent requires special digital (agent-enabled) environments that must be integrated with the target information system before being used. In addition, we did not address the fact that using Pi-Mind agents as digital decision-making clones also has social and, in particular, ethical challenges. The use of Pi-Mind agents in organisations may reveal many economic and legal challenges (copyrights, liability, rewards for using such an asset, etc.), which should be studied separately. Despite these limitations, we believe that the results present a clear benefit for ensuring the efficiency of critical decision-making and organisational resilience.

For future research, we also recognise the importance of understanding (during cloning) the donor's emotional state, which often influences their decisions. The extent to which emotions influence choices during decision-making is very personal. Each clone must capture these specifics. Humans often make decisions in groups, and social interactions affect each individual's preferences. Everyone must balance individual and group biases while making decisions; therefore, personal clones or Pi-Mind agents must capture the specific limits of compromises for every individual. Finally, cloning the groups could be a potentially important topic for generalising the Pi-Mind concept from individual to collective intelligence. Such clones will learn to compromise between individual and collective choices. However, there is much to be done in this area, and we welcome other researchers to join us.

Notes

1. August 2013, http://dovira.eu/Law_proposal.pdf.
2. November 2013, http://dovira.eu/Round_table_%208_11.pdf.
3. February 2014, http://dovira.eu/UCU_proposals.pdf.
4. March 2014, http://dovira.eu/Accreditation_proposals.pdf.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Mariia Golovianko  <http://orcid.org/0000-0003-0734-4028>

Svitlana Gryshko  <http://orcid.org/0000-0001-7286-413X>

Vagan Terziyan  <http://orcid.org/0000-0001-7732-2962>

Tuure Tuunanen  <http://orcid.org/0000-0001-7119-1412>

References

- Abramson, D. I., & Johnson, J. (2020, December 1). *Creating a conversational chatbot of a specific person* (U. S. Patent No. 10,853,717). U.S. Patent and Trademark Office. <https://pdfpiw.uspto.gov/piw?PageNum=0&docid=10853717&IDKey=&HomeUrl=%2F>
- Acharya, A., Howes, A., Baber, C., & Marshall, T. (2018). Automation reliability and decision strategy: A sequential decision-making model for automation interaction. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 62(1), 144–148. <https://doi.org/10.1177/1541931218621033>
- Ågerfalk, P. J. (2020). Artificial intelligence as a digital agency. *European Journal of Information Systems*, 29(1), 1–8. <https://doi.org/10.1080/0960085X.2020.1721947>
- Ågerfalk, P. J., & Karlsson, F. (2020). Artefactual and empirical contributions in information systems research. *European Journal of Information Systems*, 29(2), 109–113. <https://doi.org/10.1080/0960085X.2020.1743051>
- Agrawal, A., Gans, J. S., & Goldfarb, A. (2019). Exploring the impact of artificial intelligence: Prediction versus judgment. *Information Economics and Policy*, 47, 1–6. <https://doi.org/10.1016/j.infoecopol.2019.05.001>
- Al Faruque, M. A., Muthirayan, D., Yu, S. Y., & Khargonekar, P. P. (2021). Cognitive digital twin for manufacturing systems. In *Proceedings of the 2021 design, automation & test in Europe conference & exhibition* (pp. 440–445). IEEE. <https://doi.org/10.23919/DATE51398.2021.9474166>
- Amari, S. I. (1993). Backpropagation and stochastic gradient descent method. *Neurocomputing*, 5(4–5), 185–196. [https://doi.org/10.1016/0925-2312\(93\)90006-O](https://doi.org/10.1016/0925-2312(93)90006-O)
- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>
- Bai, T., Luo, J., Zhao, J., & Wen, B. (2021). Recent advances in adversarial training for adversarial robustness. *arXiv preprint*. arXiv:2102.01356.
- Baiyere, A., Salmela, H., & Tapanainen, T. (2020). Digital transformation and the new logics of business process management. *European Journal of Information Systems*, 29(3), 238–259. <https://doi.org/10.1080/0960085X.2020.1718007>
- Baskerville, R. L., Myers, M. D., & Yoo, Y. (2019). *Digital first: The ontological reversal and new challenges for IS research*. MIS Quarterly. Advance online publication. https://scholarworks.gsu.edu/cgi/viewcontent.cgi?article=1009&context=ebsc_articles
- Becue, A., Maia, E., Feeken, L., Borchers, P., & Praça, I. (2020). A new concept of digital twin supporting optimization and resilience of factories of the future. *Applied Sciences*, 10(13), 4482. <https://doi.org/10.3390/app10134482>
- Bennett, J. M., Challinor, K. L., Modesto, O., & Prabhakaran, P. (2020). Attribution of blame of crash causation across varying levels of vehicle automation. *Safety Science*, 132, 104968. <https://doi.org/10.1016/j.ssci.2020.104968>
- Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, 45(3), 1433–1450. <https://doi.org/10.25300/MISQ/2021/16274>
- Biggio, B., & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. *Pattern Recognition*, 84, 317–331. <https://doi.org/10.1016/j.patcog.2018.07.023>
- Booyes, W., Wilke, D. N., & Heyns, S. (2020). Deep digital twins for detection, diagnostics and prognostics. *Mechanical Systems and Signal Processing*, 140, 106612. <https://doi.org/10.1016/j.ymssp.2019.106612>
- Breque, M., Nul, L. D., & Petridis, A. (2021). *Industry 5.0: Towards a sustainable, human-centric and resilient European industry*. European Commission Directorate-General for Research and Innovation. https://ec.europa.eu/info/news/industry-50-towards-more-sustainable-resilient-and-human-centric-industry-2021-jan-07_en
- Bryson, J. J., & Theodorou, A. (2019). How society can maintain human-centric artificial intelligence. In M. Toivonen & E. Saari (Eds.), *Human-centered digitalization and services* (Vol. 19, pp. 305–323). Springer. https://doi.org/10.1007/978-981-13-7725-9_16
- Bughin, J., Hazan, E., Lund, S., Dahlström, P., Wiesinger, A., & Subramaniam, A. (2018). Skill shift: Automation and the future of the workforce. *McKinsey Global Institute*, 1, 3–84. <https://www.mckinsey.com/featured-insights/future-of-work/skill-shift-automation-and-the-future-of-the-workforce>
- Carneiro, J., Martinho, D., Marreiros, G., & Novais, P. (2019). Arguing with behavior influence: A model for web-based group decision support systems. *International Journal of Information Technology & Decision Making*, 18(2), 517–553. <https://doi.org/10.1142/S0219622018500542>
- Castro, L. N., De Castro, L. N., & Timmis, J. (2002). *Artificial immune systems: A new computational intelligence approach*. Springer-Verlag.
- Chatterjee, S., Xiao, X., Elbanna, A., & Saker, S. (2017). The information systems artifact: A conceptualization based on general systems theory. In *Proceedings of the 50th Hawaii International Conference on System Sciences* (pp. 5717–5726). <https://doi.org/10.24251/HICSS.2017.689>
- Crowder, J. A., Carbone, J., & Friess, S. (2020). Methodologies for continuous, life-long machine learning for AI systems. In *Artificial psychology* (pp. 129–138). Springer. https://doi.org/10.1007/978-3-030-17081-3_11
- Cummings, T. G., & Worley, C. G. (2014). *Organization development and change*. Cengage Learning.

- Drury-Grogan, M. L., Conboy, K., & Acton, T. (2017). Examining decision characteristics & challenges for agile software development. *Journal of Systems and Software*, 131, 248–265. <https://doi.org/10.1016/j.jss.2017.06.003>
- Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of big data—Evolution, challenges and research agenda. *International Journal of Information Management*, 48, 63–71. <https://doi.org/10.1016/j.ijinfomgt.2019.01.021>
- Duchek, S. (2020). Organizational resilience: A capability-based conceptualization. *Business Research*, 13(1), 215–246. <https://doi.org/10.1007/s40685-019-0085-7>
- Filip, F. G., Zamfirescu, C. B., & Ciurea, C. (2017). *Computer-supported collaborative decision-making*. Springer. <https://doi.org/10.1007/978-3-319-47221-8>
- Floridi, L., & Cows, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 1(1), 1–15. <https://doi.org/10.1162/99608f92.8cd550d1>
- García-Magariño, I., Muttukrishnan, R., & Lloret, J. (2019). Human-centric AI for trustworthy IoT systems with explainable multilayer perceptrons. *IEEE Access*, 7, 125562–125574. <https://doi.org/10.1109/ACCESS.2019.2937521>
- Golovianko, M., Gryshko, S., Terziyan, V., & Tuunanen, T. (2021). Towards digital cognitive clones for the decision-makers: Adversarial training experiments. *Procedia Computer Science*, 180, 180–189. Elsevier. <https://doi.org/10.1016/j.procs.2021.01.155>
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (Vol. 2, pp. 2672–2680). MIT Press. <https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf>
- Greenfield, A. (2010). *Everyware: The dawning age of ubiquitous computing*. New Riders.
- Grieves, M. (2019). Virtually intelligent product systems: Digital and physical twins. In S. Flumerfelt, K. G. Schwartz, D. Mavris, & S. Briceno (Eds.), *Complex systems engineering: Theory and practice* (pp. 175–200). American Institute of Aeronautics and Astronautics. <https://doi.org/10.2514/4.105654>
- Gupta, S., Kamboj, S., & Bag, S. (2021). Role of risks in the development of responsible artificial intelligence in the digital healthcare domain. *Information Systems Frontiers*, 1–18. <https://doi.org/10.1007/s10796-021-10174-0>
- Hahn, G. J. (2020). Industry 4.0: A supply chain innovation perspective. *International Journal of Production Research*, 58(5), 1425–1441. <https://doi.org/10.1080/00207543.2019.1641642>
- Harris, J. G., & Davenport, T. H. (2005). Automated decision making comes of age. *MIT Sloan Management Review*, 46(4), 83–89.
- Hartmann, S., Weiss, M., Newman, A., & Hoegl, M. (2019). Resilience in the workplace: A multilevel review and synthesis. *Applied Psychology*, 69(3), 913–959. <https://doi.org/10.1111/apps.12191>
- Hayes, J. (2018). *The theory and practice of change management*. Red Globe Press.
- Heeks, R., & Ospina, A. V. (2019). Conceptualising the link between information systems and resilience: A developing country field study. *Information Systems Journal*, 29(1), 70–96. <https://doi.org/10.1111/isj.12177>
- Henz, P. (2021). Ethical and legal responsibility for artificial intelligence. *Discover Artificial Intelligence*, 1(1), 1–5. <https://doi.org/10.1007/s44163-021-00002-4>
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105. <https://doi.org/10.2307/25148625>
- Hillmann, J., & Guenther, E. (2021). Organizational resilience: A valuable construct for management research? *International Journal of Management Reviews*, 23(1), 7–44. <https://doi.org/10.1111/ijmr.12239>
- Hou, L., Wu, S., Zhang, G. K., Tan, Y., & Wang, X. (2021). Literature review of digital twins applications in construction workforce safety. *Applied Sciences*, 11(1), 339. <https://doi.org/10.3390/app11010339>
- How, M. L., Cheah, S. M., Chan, Y. J., Khor, A. C., & Say, E. M. P. (2020). Artificial intelligence-enhanced decision support for informing global sustainable development: A human-centric AI-thinking approach. *Information*, 11(1), 39. <https://doi.org/10.3390/info11010039>
- Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making. *Business Horizons*, 61(4), 577–586. <https://doi.org/10.1016/j.bushor.2018.03.007>
- Jeston, J., & Nelis, J. (2014). *Business process management*. Routledge.
- Jones, D., Snider, C., Nassehi, A., Yon, J., & Hicks, B. (2020). Characterising the digital twin: A systematic literature review. *CIRP Journal of Manufacturing Science and Technology*, 29(Part A), 36–52. <https://doi.org/10.1016/j.cirpj.2020.02.002>
- Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, 64(6), 515–526. <https://doi.org/10.1037/a0016755>
- Kambhampati, S. (2019). Challenges of human-aware AI systems. *arXiv preprint*. arXiv:1910.07089
- Kaur, D., Uslu, S., & Durrezi, A. (2020). Requirements for trustworthy artificial intelligence—A review. In L. Barolli, K. Li, T. Enokido, & M. Takizawa (Eds.), *Advances in networked-based information systems. Advances in intelligent systems and computing* (Vol. 1264, pp. 105–115). Springer. https://doi.org/10.1007/978-3-030-57811-4_11
- Kolbjørnsrud, V., Amico, R., & Thomas, R. J. (2016). How artificial intelligence will redefine management. *Harvard Business Review*, 2, 1–6. <https://www.pegacom/system/files/resources/2018-05/hbr-how-ai-will-redefine-management.pdf>
- Kumar, R. S. S., Nyström, M., Lambert, J., Marshall, A., Goertzel, M., Comissoneru, A., & Xia, S. (2020). Adversarial machine learning—Industry perspectives. *arXiv preprint*. arXiv:2002.05646
- Kurakin, A., Goodfellow, I., & Bengio, S. (2016). Adversarial machine learning at scale. *arXiv preprint* arXiv:1611.01236.
- Kuziemski, M., & Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. *Telecommunications Policy*, 44(6), 101976. <https://doi.org/10.1016/j.telpol.2020.101976>
- Lai, K., Oliveira, H. C., Hou, M., Yanushkevich, S. N., & Shmerko, V. (2020). Assessing risks of biases in cognitive decision support systems. In *Proceedings of the 28th European Signal Processing Conference* (pp. 840–844). IEEE. <https://doi.org/10.23919/Eusipco47968.2020.9287384>

- Lechner, U., & Hummel, J. (2002). Business models and system architectures of virtual communities: From a sociological phenomenon to peer-to-peer architectures. *International Journal of Electronic Commerce*, 6(3), 41–53. <https://doi.org/10.1080/10864415.2002.11044242>
- Li, M., & Tuunanen, T. (2022). Information technology-supported value co-creation and co-destruction via social interaction and resource integration in service systems. *Journal of Strategic Information Systems*, 31(2).
- Linnenluecke, M. K. (2017). Resilience in business and management research: A review of influential publications and a research agenda. *International Journal of Management Reviews*, 19(1), 4–30. <https://doi.org/10.1111/ijmr.12076>
- Lobschat, L., Mueller, B., Eggers, F., Brandimarte, L., Diefenbach, S., Kroschke, M., & Wirtz, J. (2021). Corporate digital responsibility. *Journal of Business Research*, 122, 875–888. <https://doi.org/10.1016/j.jbusres.2019.10.006>
- Longo, F., Nicoletti, L., & Padovano, A. (2017). Smart operators in industry 4.0: A human-centered approach to enhance operators' capabilities and competencies within the new smart factory context. *Computers & Industrial Engineering*, 113, 144–159. <https://doi.org/10.1016/j.cie.2017.09.016>
- Lu, Y., Xu, X., & Wang, L. (2020). Smart manufacturing process and system automation—A critical review of the standards and envisioned scenarios. *Journal of Manufacturing Systems*, 56, 312–325. <https://doi.org/10.1016/j.jmsy.2020.06.010>
- Matthews, G., Wohleber, R., & Lin, J. (2019). Stress, skilled performance, and expertise: Overload and beyond. In P. Ward, J. M. Schraagen, J. Gore, & E. M. Roth (Eds.), *Oxford handbook of expertise: Research and application* (pp. 490–524). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198795872.013.22>
- Melchor, O. H. (2008). Managing change in OECD governments: An introductory framework. *OECD Working Papers on Public Governance*, 12. OECD Publishing. <https://doi.org/10.1787/227141782188>
- Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1–38. <https://doi.org/10.1016/j.artint.2018.07.007>
- Myers, M. D. (2021). Big data analytics: Ethical dilemmas, power imbalances, and design science research. *Communications of the Association for Information Systems*, 49(1), 19. <https://doi.org/10.17705/1CAIS.04919>
- Nguyen-Duc, A. (2020). An analytical framework for planning minimum viable products. In A. Nguyen-Duc, J. Münch, R. Prikladnicki, X. Wang, & P. Abrahamsson (Eds.), *Fundamentals of software startups* (pp. 81–95). Springer. https://doi.org/10.1007/978-3-030-35983-6_5
- Nguyen, K., Tuunanen, T., Gardner, L., & Sheridan, D. (2020). Design principles for learning analytics information systems in higher education. *European Journal of Information Systems*, 30(5), 541–569. <https://doi.org/10.1080/0960085X.2020.1816144>
- Nunamaker, J. F., Jr., Briggs, R. O., Derrick, D. C., & Schwabe, G. (2015). The last research mile: Achieving both rigor and relevance in information systems research. *Journal of Management Information Systems*, 32(3), 10–47. <https://doi.org/10.1080/07421222.2015.1094961>
- Odena, A. (2016). Semi-supervised learning with generative adversarial networks. *arXiv preprint*. arXiv:1606.01583v2.
- Paschen, J., Wilson, M., & Ferreira, J. J. (2020). Collaborative intelligence: How human and artificial intelligence create value along the B2B sales funnel. *Business Horizons*, 63(3), 403–414. <https://doi.org/10.1016/j.bushor.2020.01.003>
- Peppers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3), 45–77. <https://doi.org/10.2753/MIS0742-1222240302>
- Phillips-Wren, G., & Adya, M. (2020). Decision making under stress: The role of information overload, time pressure, complexity, and uncertainty. *Journal of Decision Systems*, 29(sup1), 213–225. <https://doi.org/10.1080/12460125.2020.1768680>
- Raetz, S., Duchek, S., Maynard, M. T., & Kirkman, B. L. (2021). Resilience in organizations: An integrative multi-level review and editorial introduction. *Group & Organization Management*, 46(4), 607–656. <https://doi.org/10.1177/10596011211032129>
- Rajnai, Z., & Kocsis, I. (2017). Labor market risks of Industry 4.0, digitization, robots and AI. In *Proceedings of the 15th IEEE International Symposium on Intelligent Systems and Informatics* (pp. 343–346). IEEE. <https://doi.org/10.1109/SISY.2017.8080580>
- Rathore, M. M., Shah, S. A., Shukla, D., Bentafat, E., & Bakiras, S. (2021). The role of AI, machine learning, and big data in digital twinning: A systematic literature review, challenges, and opportunities. *IEEE Access*, 9, 32030–32052. <https://doi.org/10.1109/ACCESS.2021.3060863>
- Reketye, G., & Reketye, G., Jr. (2020). The changing role of customer experience in the age of Industry 4.0. *Marketing & Menedzsmnt*, 54(1), 17–27. <https://doi.org/10.15170/MM.2020.54.01.02>
- Riulli, L., & Savicki, V. (2003). Information system organizational resilience. *Omega*, 31(3), 227–233. [https://doi.org/10.1016/S0305-0483\(03\)00023-9](https://doi.org/10.1016/S0305-0483(03)00023-9)
- Rosenberg, L. (2016). Artificial swarm intelligence, a human-in-the-loop approach to A.I. *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, 30(1), 4381–4382. <https://ojs.aaai.org/index.php/AAAI/article/view/9833>
- Royackers, L., Timmer, J., Kool, L., & van Est, R. (2018). Societal and ethical issues of digitization. *Ethics and Information Technology*, 20(2), 127–142. <https://doi.org/10.1007/s10676-018-9452-x>
- Ruijten, P. A., Terken, J., & Chandramouli, S. N. (2018). Enhancing trust in autonomous vehicles through intelligent user interfaces that mimic human behavior. *Multimodal Technologies and Interaction*, 2(4), 62. <https://doi.org/10.3390/mti2040062>
- Sakurai, M., & Chughtai, H. (2020). Resilience against crises: COVID-19 and lessons from natural disasters. *European Journal of Information Systems*, 29(5), 585–594. <https://doi.org/10.1080/0960085X.2020.1814171>
- Schranz, M., Di Caro, G. A., Schmickl, T., Elmenreich, W., Arvin, F., Şekercioglu, A., & Sende, M. (2020). Swarm intelligence and cyber-physical systems: Concepts, challenges and future trends. *Swarm and Evolutionary Computation*, 60, 100762. <https://doi.org/10.1016/j.swevo.2020.100762>
- Semenets, V., Terziyan, V., Gryshko, S., & Golovianko, M. (2021). Assessment and decision-making in universities: Analytics of the administration-staff compromises. *arXiv preprint arXiv:2105.10560*.

- Sharma, R., Mithas, S., & Kankanhalli, A. (2014). Transforming decision-making processes: A research agenda for understanding the impact of business analytics on organisations. *European Journal of Information Systems*, 23(4), 433–441. <https://doi.org/10.1057/ejis.2014.17>
- Shneiderman, B. (2020). Human-Centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human-computer Interaction*, 36(6), 495–504. <https://doi.org/10.1080/10447318.2020.1741118>
- Shneiderman, B. (2020). Human-centered artificial intelligence: Three fresh ideas. *AIS Transactions on Human-Computer Interaction*, 12(3), 109–124. <https://doi.org/10.17705/1thci.00131>
- Shrestha, Y. R., Ben-Menahem, S. M., & Von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. *California Management Review*, 61(4), 66–83. <https://doi.org/10.1177/0008125619862257>
- Silver, D., Singh, S., Precup, D., & Sutton, R. S. (2021). Reward is enough. *Artificial Intelligence*, 299, 103535. <https://doi.org/10.1016/j.artint.2021.103535>
- Somers, S., Oltramari, A., & Lebiere, C. (2020). Cognitive twin: A cognitive approach to personalized assistants. In *Proceedings of the AAAI Spring Symposium: Combining Machine Learning with Knowledge Engineering* (vol. 1). <http://ceur-ws.org/Vol-2600/paper13.pdf>
- Sprague, R. H., Jr. (1980). A framework for the development of decision support systems. *MIS Quarterly*, 4(4), 1–26. <https://doi.org/10.2307/248957>
- Suran, S., Pattanaik, V., & Draheim, D. (2020). Frameworks for collective intelligence: A systematic literature review. *ACM Computing Surveys (CSUR)*, 53(1), 1–36. <https://doi.org/10.1145/3368986>
- Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeker, K. I. (2020). An overview of clinical decision support systems: Benefits, risks, and strategies for success. *NPJ Digital Medicine*, 3(1), 1–10. <https://doi.org/10.1038/s41746-020-0221-y>
- Sorensen, C. (2010). Cultivating interaction ubiquity at work. *The Information Society*, 26(4), 276–287. <https://doi.org/10.1080/01972243.2010.489856>
- Takahashi, K. (2020). Social issues with digital twin computing. *NTT Technical Review*, 18(9), 36–39. <https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202009fa5.html>
- Terziyan, V., Golovianko, M., & Gryshko, S. (2018a). Industry 4.0 intelligence under attack: From cognitive hack to data poisoning. In K. Dimitrov (Ed.), *Cyber defence in Industry 4.0 systems and related logistics and IT infrastructure (NATO Science for Peace and Security Series D: Information and Communication Security)* (Vol. 51, pp. 110–125). IOS Press. <https://doi.org/10.3233/978-1-61499-888-4-110>
- Terziyan, V., Golovianko, M., & Shevchenko, O. (2015). Semantic portal as a tool for structural reform of the Ukrainian educational system. *Information Technology for Development*, 21(3), 381–402. Taylor & Francis. <https://doi.org/10.1080/02681102.2014.899955>
- Terziyan, V., Gryshko, S., & Golovianko, M. (2018b). Patented intelligence: Cloning human decision models for Industry 4.0. *Journal of Manufacturing Systems*, 48 (Part C), 204–217. Elsevier. <https://doi.org/10.1016/j.jmsy.2018.04.019>
- Terziyan, V., & Nikulin, A. (2021). Semantics of voids within data: Ignorance-aware machine learning. *ISPRS International Journal of Geo-Information*, 10(4), 246. <https://doi.org/10.3390/ijgi10040246>
- Tharwat, A. (2021). Classification assessment methods. *Applied Computing and Informatics*, 17(1), 169–192. <https://doi.org/10.1016/j.aci.2018.08.003>
- Truby, J., & Brown, R. (2021). Human digital thought clones: The Holy Grail of artificial intelligence for big data. *Information & Communications Technology Law*, 30(2), 140–168. <https://doi.org/10.1080/13600834.2020.1850174>
- Turing, A. (1950). Computing machinery and intelligence. *Mind*, LIX(236), 433–460. <https://doi.org/10.1093/mind/LIX.236.433>
- Tuunanen, T., Kazan, E., Salo, M., Leskelä, R. L., & Gupta, S. (2019). From digitalization to cybernization: Delivering value with cybernized services. *Scandinavian Journal of Information Systems*, 31(2), 83–96. <https://aisel.aisnet.org/sjis/vol31/iss2/3>
- Tuunanen, T., & Peffers, K. (2018). Population targeted requirements acquisition. *European Journal of Information Systems*, 27(6), 686–711. <https://doi.org/10.1080/0960085X.2018.1476015>
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>
- Venable, J., Pries-Heje, J., & Baskerville, R. (2012). A comprehensive framework for evaluation in design science research. In *International Conference on Design Science Research in Information Systems* (pp. 423–438). Springer. http://link.springer.com/10.1007/978-3-642-29863-9_31
- Wang, D., Yang, Q., Abdul, A., & Lim, B. Y. (2019). Designing theory-driven user-centric explainable AI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–15). <https://doi.org/10.1145/3290605.3300831>
- Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33–35. <https://doi.org/10.1145/1118178.1118215>
- Zeng, D. (2013). From computational thinking to AI thinking. *IEEE Intelligent Systems*, 28(6), 2–4. <https://doi.org/10.1109/MIS.2013.141>