Cognitive Systems Research 82 (2023) 101166

Contents lists available at ScienceDirect

Cognitive Systems Research

journal homepage: www.elsevier.com/locate/cogsys

Fundamental concepts of cognitive mimetics

Antero Karvonen [a,*], Tuomo Kujala [b], Tommi Kärkkäinen [b], Pertti Saariluoma [b]

[a] VTT Technical Research Centre of Finland Ltd, P.O. Box 1000, FI-02044 VTT, Finland
[b] University of Jyväskylä, Faculty of Information Technology, P.O. Box 35, FI-40014 Jyväskylä, Finland

ARTICLE INFO

Action editor: Alexei V. Samsonovich

Keywords:
AI design
Cognitive mimetics
Design
Artificial intelligence

ABSTRACT

The rapid development and widespread adoption of Artificial Intelligence (AI) technologies have made the development of AI-specific design methods an important topic to advance. In recent decades, the centre of gravity in AI has shifted away from cognitive science and related fields like psychology. However, there is a clear need and potential for added value in returning to stronger interaction. One potential challenge for this interaction may be the lack of common conceptual grounds and design languages.

In this article, we aim to contribute to the development of conceptual interfaces for human-based AI-specific design methods through the idea of cognitive mimetics. We begin by introducing basic concepts from mimetic design and interpret them in the context of this thematic area. These provide some of the basic building blocks for a design language and bring to the surface key questions. These in turn provide a ground for explicating cognitive mimetics. In the second part of this paper, we focus on specifying a key aspect in cognitive mimetics: the contents of information processes.

Others engaged in this field can derive value from using or developing the basic conceptual machinery to specify their own approaches in this interdisciplinary field that is still shaping itself. Furthermore, those who resonate with the idea of cognitive mimetics, as specified here, can join in taking this particular approach further.

1. Introduction

The emergence of Artificial Intelligence (AI) on a massive scale means that it is important to advance AI design methods. A key aspect of any design process is to ask how to get new ideas and how to frame and solve design problems (Cross, 2006, 2011; Schön, 2017; Simon, 2019). One family of approaches for AI design are those characterised by the fact that they take their inspiration from some facet of human mental processes (Haugeland, 1997; Lake et al., 2017; Lieto, 2021; Muggleton & Chater, 2021; Saariluoma et al., 2018). Cognitive mimetics is one approach, in which the goal is to imitate, though not slavishly, human information processing in intelligent technology. Cognitive mimetics builds on the core ideas of mimetic design by interpreting them in the context of the design of intelligent technology.

Despite a rich historical pedigree, the centre of gravity in AI and research into the human mind have become more distant in recent years and decades. One possible reason for this is a lack of common design-oriented frameworks and core concepts. Given the sophistication of human thought, action, and perception, there is potential in this thematic area, both in terms of novel and better functioning of technology, as well as aligning it with human action and values. Harnessing this potential requires both making sense of the complexities of human information processing and making use of it in a design context.

Given that both psychology and cognitive science contain various perspectives, models, and theories to mind and the explanation of behaviour and action, it is important to create conceptual tools to clarify differences and similarities within the space of possible approaches. This may help establish common discourse and take thinking and design forward. To fulfil this need, design-oriented conceptual work is needed. This is the essential research gap and problem this article addresses itself. Our method to doing so is by introducing core concepts from mimetic design and using them as a vehicle for specifying cognitive mimetics in a general sense and then focus on one aspect it, which we call the content-based approach.

2. Background

Historically, the relationship between cognitive science and AI was strong. Although its connotation has shifted, the early idea of cognitive science was conceptualised as referring to those facets of human thought

* Corresponding author.
E-mail addresses: antero.karvonen@vtt.fi (A. Karvonen), tuomo.kujala@jyu.fi (T. Kujala), tommi.karkkainen@jyu.fi (T. Kärkkäinen), ps@jyu.fi (P. Saariluoma).

https://doi.org/10.1016/j.cogsys.2023.101166
Received 30 April 2023; Received in revised form 8 August 2023; Accepted 10 August 2023
Available online 11 August 2023

and perception which could benefit from using artificial intelligence as a method and experimental system (Longuet-Higgins, 1973). Important pioneers in cognitive science Newell and Simon did exactly this, they modelled human thought on computers with the idea that a model based on empirical human research that replicated human behaviour was proof of the validity of the theory and model it embodied, and of course simultaneously was an instance of AI itself (Newell & Simon, 1972, 1976; Simon, 1995). Since the early days, AI and cognitive science have diverged and branched out in various directions. While influence has continued, calls for a stronger return of this interaction have also been raised (Anderson, 2015). There are clear signals that it may be on the rise again (Friston et al., 2022; Lake et al., 2017; Lieto, 2021; Littman et al., 2021; Muggleton & Chater, 2021), perhaps never went away, e.g., Goel (2013), and may find a place even in modern machine learning as well (Eppe et al., 2022; Littman et al., 2021). Some have called for taking inspiration from the neurocognitive basis of human information processing (e.g. LeCun, 2022; Zador et al., 2023). Work on biologically-inspired cognitive architectures (BICAs) has been ongoing for more than a decade (Samsonovich, 2012). Traditional cognitive architectures remain an important field as well in this context (Lieto, 2021).

Nevertheless, the main lines of modern AI has distanced itself from attempting to imitate the human mind. Russell and Norvig (2010), for example, explicitly formulated the fundamental stance of AI in accordance with the principle of (bounded) rational action (Nilsson, 2009), explicitly distancing themselves from human thought and action as a source of imitation. We will only mention here the partial irony that this concept originates in the work of Herbert Simon (Newell & Simon, 1972; Simon, 1995, 2019) as a general concept covering *both* natural and artificial intelligence. The basic ground for this position is apparently in the idea that human thought and action is imperfect and idiosyncratic, comparatively slow and based on quite different strengths than computers, and thus does not provide a good basis for AI. In their formulation (Russell & Norvig, 2010), attempting to model human mental processes on computers is the domain of cognitive modelling, i.e., models *of* cognition. “The bitter lesson”, according to a critical remark by Sutton (2019) has been that leveraging computational power using simple techniques over large sets of data has yielded orders of magnitude better results in practical applications than attempting to encode human knowledge and information processing in artificial systems. However, such a claim builds upon narrow conceptions of what taking inspiration from the human mind entails and are apt to generate controversies where from the perspective of AI development there should be active dialogue. The core idea of cognitive mimetic design is conceptually useful for mitigating this, because from a mimetic perspective it is uncontroversial that artificial systems are different and should play on different strengths than the source of inspiration does. Mimetic design solutions are never identical with their sources, rather they exist on a gradient of similarity when viewed at a particular level of abstraction (Floridi, 2009). Furthermore, mimetic design is not only about comparative similarity between natural and designed systems, but about the design process itself in which the source is made to stand in relation to a design target (and vice versa). In fact, it can be asked whether the methods highlighted by Sutton (2019), *search* and *learning*, are wholly *un*inspired by concepts developed from investigations into the human mind. Given the rich descriptions of the origins of reinforcement learning (RL) in Sutton and Barto (2018) we can infer that he would not disagree with the assertion that various forms of mimetic-like thinking have played an important, though complex and probably rather serendipitous role in the development of RL. Finally, as Sutton points out, leveraging human knowledge (even if that would be the extent of cognitive mimetics) and leveraging computation need not run counter to each other – which is congruent with cognitive mimetic thinking, and in accordance with common sense.

A broader issue is that it is not always clear at which level(s) of abstraction the discourse is proceeding on either in terms of the source of inspiration or the technological target, nor whether there are sufficient common conceptual grounds to achieve sufficient understanding from a mimetic perspective. Thus, the questions of whether, to what extent, how, and on which levels AI should mimic human thinking and action is impossible to answer in simple blanket terms. It is important to point out, that in practice such mimetic processes have occurred starting with Turing (1936):

- ”Turing defined his machine not to model physical digital computers, but to **model the thinking process of a human**.” (Fortnow, 2012), emphasis added
- ”Gandy also emphasised that Turing’s original argument was drawn from **modelling the action of a human being working to a rule** and was not based on modelling machines …” (Hodges, 2013), emphasis added
- The computer is a particular type of **automaton** that as an artefact was “designed to simulate and imitate certain kinds of human *thought* processes” (Dasgupta, 2016), emphasis added

Certainly, if thinking in purely functional terms, it makes little practical sense to mimic human thinking in domains where it has already been surpassed by AI, like chess, Go, and many other games. However, the world contains countless domains, many of which are far less formal than games, where AI is barely making headway. There, it remains to be discovered what counts as intelligent action, and very probably it is currently captured in human thought and action and thus provides a reasonable place to start and allows building intelligent technology on solid foundations. Pure performance, often narrowly construed in competitive terms, is only one aspect of intelligent technology. Framing human and machine intelligence in terms of competition and exclusion in tasks and activities is neither desirable or realistic, as far more likely use cases are those where the two form a “hybrid intelligence” system (Dellermann et al., 2019). Even autonomous intelligent technology interacts with and is embedded in human life. For instance, autonomous cars of the future should both behave in human-predictable ways as well as be able to predict human driving, and such requirements imply being structured in accordance with folk psychological categories as well as being able to apply similar heuristics to human drivers (Schumann et al., 2023; see also Dennett, 1987) – a clear case of design problems suitable for cognitive mimetics. As another example, an integration of human–machine intelligence is present at the design level in the generative pre-trained transformers (GPT), which are fine-tuned using so-called reinforcement learning with human feedback (RLHF) (Liu, 2023).

At base, even many modern machine learning-based systems are mimetic at the level of the data they learn from, which is already rich with human-generated structure. Furthermore, even “pure” machine learning methods that leverage computational power require human thought in setting up, for example, reward functions, network architectures or descriptions of states and actions in Markov’s decision processes (Littman et al., 2021; Silver et al., 2021; Sutton & Barto, 2018). From the present point of view, in novel domains with uncertainty and complexity, setting up goals, rewards, state descriptions, and actions may benefit from contextual human research, because programmers are only occasionally experts in the domains for which intelligent technology is being developed (Saariluoma, Karvonen, et al., 2022). Most domains of human action do not come with an explicit rulebook written on the back and thus require in-depth investigation. Moreover, expertise is largely tacit, in the sense that while experts can perform well, the reasons for successful action are subtle and even under the threshold of their conscious thinking, requiring methods for successful elucidation (Ericsson et al., 2018). Furthermore, it may be interesting to speculate on the origins of the concepts of attention or next-token prediction in transformer architectures and the fine conceptual structures in the human mind that emerge as algorithmic or architectural ideas in software (Freed, 2013; Jackson, 2013) – already Boden (1990) noted that programmers apply implicit theories of human action and thought into programming.

Broadly speaking, the human mind is *in general* a very capable and adaptive system and many general principles and abilities, some of which may be realisable in technology, remain to be discovered and mapped into technology. The extensive work done with cognitive architectures for decades (Lieto, 2021) are a key field from the overall cognitive mimetic perspective in this respect. Indeed fields like cognitive psychology that seek to discover invariants or universals (Anderson, 2015; Simon, 1990, 2019) of human cognitive processes are apt to provide novel insights for technology design.

Furthermore, there may be normative reasons for wanting to have some similarity with humans in the use cases typical for AI. Intelligent machines offer new possibilities for designers, such as the ability to shift tasks from humans to machines. Designing intelligent information processes often involves transforming human processes into ones carried out by machines. However, this transformation requires an understanding of what humans do. Without a clear idea of human work, it is not easy or risk-free to transfer it to machines. Finally, there seems little doubt that future AI systems should be aligned with human values (Gabriel, 2020). So-called 'AI ethics' is a burgeoning field of research (Coeckelbergh, 2020). These questions can be formulated from a mimetic perspective (Karvonen, 2020) as well. The point is that if AI systems should be aligned with human values, they are mimetic to human thought on at least a normative and ideal level. For example, in the discourse on AI ethics, it is becoming rather clear that the enumeration of abstract principles does not automatically yield an ethical AI system, and the problem is importantly one of design languages that facilitate such design and bridge the principles-practices gap (Saariluoma, Myllylä, et al., 2022).

These examples suffice to note that AI design approaches in the broad net cast by a very loosely defined cognitive mimetics remain important even (or perhaps particularly) in light of the recent advances in AI.

2.1. Basic concepts in mimetic design

The basic idea of cognitive mimetics is straightforward. It is an approach for the design of intelligent technology using human information processing and contents as a source of solutions – with an emphasis on so-called higher cognitive processes like thinking and reasoning rather than sensory-motor processes (Kujala & Saariluoma, 2018). The premise is that human intelligence remains the most prominent example of intelligent information processing. From a mimetic perspective this provides a rationale for using it as a source of inspiration. However, this also calls for a conceptual clarification, since, among other things, (human) information processing presents a multisided phenomenon as does the concept of information itself (Deacon, 2007; Floridi, 2009). Furthermore, "the mind" is open to multiple (Allen, 2017), sometimes mutually exclusive, interpretations and foundational assumptions (Haugeland, 1997). The basic concepts of mimetic design such as *source* and *target* provide useful tools for making things more explicit and for structuring intelligent technology design discourse from a mimetic perspective.

The examples in the previous section, interpreted as mimetic in a general sense, imply quite different types or aspects of mimetic design. In the language of this article, they imply in different perspectives (Lee & Dewhurst, 2021) on the source, as well as different design targets and problems. For example, if the design question relates to the structuring of the overall actions of an intelligent agent, this implies a particular level of abstraction (Floridi, 2008) or perspective on the design target which in turn constrains the relevant level of abstraction on the source used as inspiration. The same logic applies across all possible design targets. This image of the thematic field is complex, since there are many perspectives possible on both the source and the target. Which perspectives are relevant for which aspects of design is an important topic in the future discourse on intelligent technology design.

Furthermore, the nature of the relation between source and target has many possibilities. For example one relation is *search* (Goel & Hancock, 2021; Helms et al., 2009). Search can start from either the *design problem (the target)* and proceed to search for *solutions (the source)* in a pluralistic fashion. It may also start from a *solution* and proceed to search for suitable *problems.* Both types of relation are predicated on the perspective taken on the source. Another relation can be called *mapping.* Here, assuming that a tentative solution-problem pair has been identified, the question is how to transfer the idea from the source to the target, which can be called the problem of alternative realisation. Note the importance of conceptual languages: in fact each level of abstraction is manifest in a conceptual language. In terms of the source, this may be a scientific theory language in cognitive science or psychology. In terms of the target, this may be a design language relating to an aspect of intelligent technology, including those based on theory languages in computer science.

In addition to the elements of source and target, a further general property of mimetic design is the goal of multiple or alternative realisation. In classical material-structural biomimetic examples, it is assumed that non-living artificial systems, structures, or materials can replicate similar functional characteristics that are useful for human purposes. Similarly, in cognitive mimetics, it is assumed that similar if not identical information processes can be carried out in different material systems, for example, human brains and computers. The goal of mimetic design is in implementing this alternative realisation based on the way the source is understood to operate.

To reiterate, conceptually all mimetic types of design are seemingly variations of a simple structure consisting of a *source* and a *target*, and various iterative relations between them. One of the most important properties these concepts imply is what can be called a *perspective* (Lee & Dewhurst, 2021), a *stance* (Dennett, 1987), or a level of abstraction (LoA) (Floridi, 2009). In terms of the source these are related to the concepts of explanation and understanding in science. On the target side, they are related to the various design aspects of (intelligent) technology and problems of technical realisation. Mimetic design itself can be characterised as a metastance, which seeks to bring together via a mapping relation the different perspectives in a coherent and fruitful way. These basic concepts can be transposed into a simple model, which can be called the SMT model (source, mapping, target). Here we will not seek to formalise the model or the relations, or explicate all possible perspectives on the source or the relations, rather, its' purpose is to provide some conceptual scaffolding for the discussion. The point rather is that when such reflective explication and practical work is done, a mimetic approach starts to turn into an explicit method or approach.

One overall theoretical problem in cognitive mimetics on a general level is to offer a framework in which different perspectives can be applied and to provide metadesign concepts and languages to provide some structure to such efforts. One way to proceed is by way of example, which means at the conceptual level providing explicit answers to the methodological questions implied by the concepts sketched above. This is the move from methodology to concrete method and specified approach. On an empirical level, it means providing cognitive mimetic case examples. Here we seek make some progress towards these goals on the conceptual level. To begin discussion, it is useful to build on fundamentals, and for this purpose we can briefly turn to two key antecedent approaches, biomimetics and design by analogy, and then turn to design issues.

2.2. Analogical, biomimetic, and cognitive mimetic design

Two major strands of research and practice in mimetic design are biomimetics and design-by-analogy. Cognitive mimetics is conceptually and in practice closely related to both, sharing in and adopting many of the basic ideas, but also differing in important respects, in particular by its general target as intelligent technology. Both biomimetics and design-by-analogy are quite naturally less explicitly defined in this respect. In short, cognitive mimetics differs from biomimetics by having an explicit explanatory and implementational framework in information processing, and an explicit specification of both the *source* and the *target* on a conceptual level as *human information processing* and *intelligent technology,* respectively. Design-by-analogy on the other hand is a design science focused on the general cognitive, psychological, and methodological features of any analogical design and consequently does not specify an explicit type of mimetic design. It is a design science, and not a particular and specified approach to design.

Design-by-analogy is a design science field which studies the use of analogy in design from many perspectives, cognitive-psychological, methodological and otherwise. Mimetic design can be understood conceptually as closely related – perhaps subset – to design-by-analogy (Fu et al., 2014; Jiang et al., 2021; Moreno et al., 2014; Qian & Gero, 1996; Song & Fu, 2022). In design-by-analogy, the source can be anything – natural or artificial, a simple idea or a complex model, whereas biomimetics focuses on concrete and functional biological solutions. Where biomimetics has focused more on the method and biology itself, design-by-analogy has paid attention to the psychological processes relevant for analogical design. Both have attempted to systematise the design processes, generate ontologies and databases of solutions, and develop computer tools to aid the process (Chakrabarti et al., 2005; Goel & Bhatta, 2004; Goel, 1997; Jiang et al., 2021; Song & Fu, 2022; Vattam et al., 2010).

Biomimetics and bio-inspired design is a well-known, but relatively young and not particularly well-understood mimetic design field (Bhushan, 2009; Goel & Hancock, 2021; Vincent et al., 2006). The example of Velcro is known by many, but there are countless examples (Lenau et al., 2018; Lepora et al., 2013). Biomimetics is defined in its ISO standard as "the interdisciplinary cooperation of biology and technology or other fields of innovation with the goal of solving practical problems through the function analysis of biological systems, their abstraction into models, and the transfer into and application of these models to the solution" (Graeff et al., 2020).

While these criteria can be interpreted in many ways, in practice, biomimetics has focused on materials, structures and (physical) mechanisms for realising functions (Lenau et al., 2018; Lepora et al., 2013), despite the insight by Vincent et al. (2006) that nature in fact often uses *information* to solve problems. Thus, the idea of information processing in nature has been raised in the biomimetic context (Gruber et al., 2011), but in practice this viewpoint has been overshadowed by the mechanical-material engineering design perspective and therefore have operated on a theory language largely predicated on the concept of causal, structural, behavioural and mechanical determination (Bunge, 2017). Nevertheless, biomimetic thinking has played an important role in AI as the examples in bio-inspired computing show, such as the swarm intelligence algorithms (artificial bee colony, ant colony, firefly algorithm, and particle swarm) (Del Ser et al., 2019; Kar, 2016; Linja et al., 2023). However, these examples still have a relatively behaviouristic and structural perspective on the source. When the information processing analysis is refined by including, for example, internal states such as perceptual processes, information contents, representations, cognitive models, etc., of the source of inspiration a new form of mimetic design emerges which we have felt appropriate to call *cognitive.*

Cognitive mimetic design is characterised by its use of information as a general framework for explanation, understanding and implementation. From this general starting point, many further specifications can be made to yield a specific aspect of cognitive mimetics, identified by its primary stance towards the source, which is here the *content-oriented or content-based* approach to information (Myllylä & Saariluoma, 2022; Saariluoma, 1995, 1997). Content-oriented is a wider concept here, but content-based refers to explaining intelligent action on the grounds of involved mental contents. From the viewpoint of this ontology, approaches in this thematic field that view the source of inspiration from a *neuro*cognitive perspective, such as BICAs (Samsonovich, 2012), fall on an overlapping area between bio- and cognitive mimetics, and thus include the biological basis of information processing as an additional source of ideas. However, we specify cognitive mimetics purely in terms of information processing.

The rest of the paper will be devoted to specifying one aspect of cognitive mimetics, in which the relevance of *mental contents* within the *correct and appropriate processing of mental information in complex tasks in context* is highlighted and its relevance to intelligent technology is sketched out. First, to provide some additional context to the discussion, it is useful to visit some basic concepts of design and sketch out a phased and iterative design process for cognitive mimetics that is holistic, and in which the content-based perspective is a key component.

3. Design and research cycles in cognitive mimetics

To focus our discussion, we assume that a key class of problems of intelligent technology for which cognitive mimetics can provide value, are (complex) tasks for which *correct or appropriate processing of mental information is critical* (Carroll, 1993). This means that if done by intelligent technology, the task is completed successfully, i.e., reaches an appropriate terminal state, and that it is done right, meaning that the way in which it is done is both acceptable (for example, morally) and efficient. We can further assume that there is some present situation A, where human mental processes carry critical parts of the completion of the task. The overall goal for intelligent technology is to move to situation B, where the technology can in part or whole carry out this task, possibly completely autonomously. There are important nuances to this, such as possible hybrid intelligence (Dellermann et al., 2019) models, but for the sake of argument, we will bracket them for now. Overall, cognitive mimetics is neutral with respect to this aspect of the design goal (the target), i.e., what is the intended model of use and operations for the intelligent technology. It is good to point out however, that the basic methodological cycles of cognitive mimetics can support both approaches, as exemplified in our concept of human digital twins (Karvonen & Saariluoma, 2023; Saariluoma et al., 2020; Saariluoma, Karvonen, et al., 2022).

The central question from this starting point, from a cognitive mimetic perspective, is: what are the mental information processes and information contents that are necessary and sufficient grounds for successful task completion? The hypothesis is that elucidating these information processes in depth yields a possible source of ideas for solving the same (and possibly related or even unrelated) tasks through intelligent technology. Such elucidation can proceed in a number of complementary directions, each of which has a role in the methodology of cognitive mimetics. It is crucial to begin from a holistic standpoint. Many if not all complex tasks require the coordination of multiple universal aspects of human information processing: perception, motor control, attention, representation, memory, problem-solving, reasoning, decision-making, expertise, language and communication. Each of these human information processes have both invariant and variant aspects: for example, the goal of cognitive psychology writ large is to discover the *invariants* or *universals* of these processes that in some important respects explain the variants in conjunction with the environment (Anderson, 2015; Oulasvirta et al., 2018; Simon, 1990, 2019).

However, as in perception, it is *something* that is seen, in memory *something* that is remembered, in representation *something* that is represented, or in attention *something* which is attended, there is endless variance in the contents of information processes. Contents, in turn, have their own explanatory grounds, and provide explanatory grounds for others, arguably in fact, intelligent action (Myllylä & Saariluoma, 2022; Saariluoma, 1995, 1997). In addition, as the highest human performance in a task is the best starting point for cognitive mimetics, expertise and skill become key components in the approach, which in turn highlight domain-specificity (Feltovich et al., 2006). While mental contents vary both between and within subjects, they also can be analysed according to their general properties at various levels of abstraction, whether personal, sub-personal or tacit (Saariluoma & Karvonen

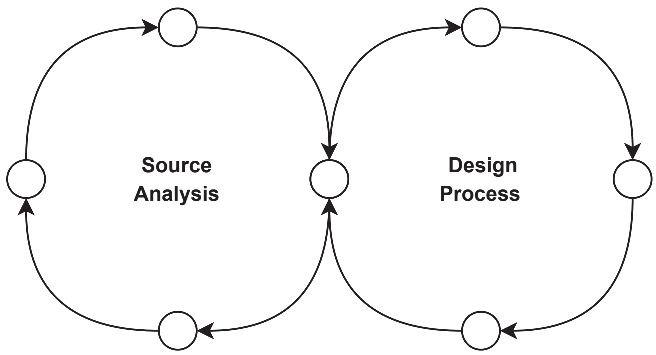

Fig. 1. The source analysis and target design process cycles are partially independent but lock into each other and exchange information at (at least) four key points.

et al., 2019; Von Eckardt, 2012).

A holistic approach to cognitive mimetics would include all the relevant aspects of mind in the completion of a task. It would also include a holistic description of the flow and contents of the information process. From this perspective, it is possible to describe the information processes relevant for particular tasks in an integrated fashion which is both compositional and specific, and also semantically evaluable, all while giving them explanatory grounds, and moreover, connecting information processes with actions and the environment. This aspect of cognitive mimetics can be called the *analysis of the source.* In practice, it is iterative and progressive work which is in cyclical dialogue with the design of the target technology. Both cycles mutually specify each other. The four phases of cognitive mimetic research will be described below after establishing a relation with four nominal phases in design processes (see Fig. 1).

Above, in describing the core question of cognitive mimetics, the task was assumed. But how do we arrive at this stage? This is essentially where the necessity of aligning with design problems and needs becomes paramount and the need for embedding cognitive mimetics in a more general and iterative design-research approach becomes apparent (see e.g. Peffers et al. (2007). The target design process side can be described along the four main general phases of design identified by Kannengiesser and Gero (2015). In the first phase, the goal is to understand and define the design problem. In the second, the goal is to generate a concept structure. In the third, the goal is to generate a solution structure. In the fourth, the goal is to finalise and deliver the design solution. This can be understood as a process of elaboration from conceptual design to specification and implementation (Gibbons et al., 2008). These correspond to distinct research cycles on the source analysis side.

For context, next we can briefly describe the envisioned overall process for cognitive mimetics. The idea for cognitive mimetics is to "lock in" to these design phases and offer inputs for ideas at each stage. On the source analysis side, this corresponds to different levels of abstraction for investigation. Roughly, these can be grouped into four research phases corresponding to the idea of progressive deepening.

The first can be called the context, task, and joint cognitive system (Hollnagel & Woods, 2005) research phase. Here, the main goal is to understand the system of regular actions (Saariluoma et al., 2016) in which the ultimate target of design is one future element. We completely agree with the view that intelligent technology should be designed as systems (Brown, 2022; Saariluoma et al., 2016), and not only at the algorithm and implementation level. This phase offers inputs particularly to understanding and defining the design goal and problem, as well as conceptual design.

The second phase begins with the selection of a task, understood as goal-seeking action. This aspect locks in with the conceptual design phase. Here the goal is to describe the information process *and* its necessary grounds both from a technological and human information processing perspective. Here, in particular, based on the information process in question, the stance towards the source becomes more specific and the analysis more demanding, including the necessity of explicating tacit knowledge (Ericsson et al., 2018; Saariluoma & Karvonen et al., 2019; Saariluoma, 1995).

The third phase is where the information process and the necessary grounds selected are modelled: this can be called the *human digital twinning (HDT)* phase (Karvonen & Saariluoma, 2023; Saariluoma et al., 2020; Saariluoma, Karvonen, et al., 2022). The input to design at this stage is quite concrete, since it is a computational model of the human information processes and contents in the completion of the task. Thus, it locks in naturally with the *solution structure* phase of design and may even become identical to it.

The final phases of implementation and finalisation in the design process correspond to iterating the HDT model. Generalising across cases, the omega phase of the process would include some ways of both communicating and sharing the findings for future reference and advancement within the thematic field. These could be called cognitive mimetic design patterns (Alexander, 1977; Gamma et al., 2009). Within a particular case-context, the final phase may simply mean a return to a lighter version of the first phase, to discover new tasks for intelligent technology, while the HDT model itself can be spun off into other application areas, such as operations design or iterated further. At this stage, new possibilities both for modelling hybrid intelligence scenarios and new task role distributions becomes possible using HDTs. At the same time, elements of the HDT can be used as a basis of autonomous AI systems, but crucially in such a manner that the entire system has been subject to analysis (Brown, 2022; Hollnagel & Woods, 2005).

Next, we will focus in more closely on the idea of information processes, and their contents, as key specifications of cognitive mimetics. This corresponds most strongly with the second phase described above.

4. Cognitive mimetics: Basic specification

The core perspectival commitment in cognitive mimetics is information. This means that information and its processing is the general interpretative (and implementation) framework for cognitive mimetics. Information here, to a first approximation, follows Wiener (1985) idea that it is not matter or energy, despite requiring them as a substrate. Information organises work and thus matter and energy (Deacon, 2011).

Associating cognitive mimetics with information however begs the question of what is meant by information. Information is a multisided phenomenon which has been approached from numerous perspectives both in terms of its quantitative and qualitative aspects (Adriaans, 2020; Floridi, 2009; Sequoiah-Grayson & Floridi, 2022). Aspray (1985) sketched out the main figures and intellectual background which formulated a scientific conceptualisation of information, and subsequently gave rise to the computer and information age in which we live. The influence of Shannon, Wiener, McCulloch and Pitts, von Neumann and Turing cannot be understated, but within the very conceptualisation of information, particularly in Shannon (1948) was also a significant conceptual problem from the present point of view (Deacon, 2003, 2007, 2011). Shannon (1948) achieved his general results by abstracting information from its natural conceptualisation, namely, as meaningful and with contents. Turing (1936) work was a significant step in that he managed to conceptualise an abstract, automatic machine that took care of information processing tasks, in particular mathematical operations. However, the work was otherwise unspecified in terms of the contents of information processing, because the symbols in the machine and related operation could be defined in any way, respecting certain constraints (Piccinini, 2004). The Turing machine was both limited and enabled by the level of abstraction at which Turing operated, namely computational processes (Saariluoma & Rauterberg, 2016). This has resulted in the basic problem of symbol grounding (Harnad, 1990) and related problems such as the Chinese Room (Searle, 1980). Conceptually similar grounding problems apply to human mental representations as well (Von Eckardt, 2012). However, in this context, we will not seek to open the problems of how mental representations are grounded, but rather focus on how they provide grounds for intelligent action, and therefore open the path to methodological issues of mapping mental contents onto computational processes.

4.1. Computational information processing and contents

Computation and information processing are sometimes used as interchangeable terms (Piccinini & Scarantino, 2010, 2011). Here we define computation as the *mechanisation of abstraction* (Aho & Ullman, 1992). It involves a *mapping* of an information process, such as addition, to some mechanical artificial system to achieve an isomorphic correspondence between these two levels (Chalmers, 2011; Crane, 2015; Deacon, 2003; Egan, 2014). From the present perspective, computation is a type of information processing, and information processing is a large class (Floridi, 2009) not subsumed under computation as defined here. From a cognitive mimetic perspective, the essential question is *what gets mapped* and *interpreted* as computation – and *how*.

From a levels perspective, we may thus distinguish three:

1. the level of information processes
2. the level of computational information processes
3. the level of realising mechanisms.

The cognitive mimetic process can be characterised as a mapping from 1 to 2, pragmatically constrained by 3 (but not in the focus of cognitive mimetics).

Computation is thus approached pragmatically, as the ground of computers and thus the target "as we find it". There is no pressing need to take a stand on the ontological status of computation with respect to mental processes, i.e., whether they are constituted as such (Clark, 1996; Dreyfus, 2007; Lucas, 1996; Newell & Simon, 1976; Penrose, 1990; Piccinini, 2004; Rapaport, 2012; Simon, 1995), other than to note that if the target is understood as computational information processing, then over the mapping relationship, any representations of the source need to land eventually in the computational idiom. This makes, for example, *computational* cognitive science a prima facie suitable theory language for cognitive mimetics, where the empirical question is not whether cognition *is* computation, but whether it is *computable* (Rapaport, 2012). For present purposes to advance our specification, we may take computation in terms of the target as given. Insofar as this is the case, the mapping relation logically must have the computational element.

However, the fact that we can implement all sorts of information processes on computers, and not just "calculations", is because they are defined as symbol processing machines (Newell & Simon, 1976) and the contents of computations can be defined any way we like. For present purposes, it makes no difference whether the computations are based on digital symbol strings, patterns of activation in networks, or analog systems, as these are for present purposes more like representation-bearers which do not intrinsically have contents, only operational principles (Clark, 1996; Von Eckardt, 2012). As Adriaans (2020) points out, since Turing (1936) machines are defined as purely operational mappings between computations, they are ontologically neutral. A problem in cognitive science and AI has been that the contents of the symbols do not enter into the computational processes as such (Harnad, 1990; Sayre, 1987). In the philosophy of mind and cognition, the problem of computation and contents has arrived at a similar conclusion, at least according to Piccinini (2004), that questions of computation are independent of questions of content. In the simplistic addition example, the computational process runs the same way even if we assign it to count apples, oranges or money, and this happens not to be a problem since the operation of addition is abstract and applies to all cases of numerical addition. This is why Egan (2014), for example, called the contents of the symbols "glosses" over the actual operations in, for example, cognitive modelling.

While the above distinction has some heuristic value, it makes little sense to say that information processes, computational or otherwise, are ever content-free. Rather, it is more accurate to say that some information contents are more *abstract* than others, which gives them wider range but less specificity. In addition, some information processes (as well as contents) are more fixed than others. What the mechanisation of abstraction actually entails, is that fixed processes are combined with abstract and fixed contents, which makes them susceptible to implementation in mechanical systems (see Saariluoma & Karvonen, 2023 for more).

Human information processing is far more complex as the contents of information processes *do matter;* they define the correctness and accuracy of thinking, among other things. This is obvious, (though controversial; Dennett (1987)), but the fact that we can assign all kinds of informational contents and relation to computational processes in practice simply means that we must obtain a clear and distinct understanding of the information process in question. This is precisely where explicating mental contents becomes a plausible methodological approach for seeking clear and distinct descriptions of human information processes. It is precisely due to the arbitrariness and "glossiness" (Egan, 2014) of contents of symbols in machine information processing that we need to be precise in the process of analysis and mapping. The computational idiom forces the issue. It is a related but different question how this is implemented computationally, herein a facet of the *mapping relation* understood as alternative realisation.

4.2. Information contents and expertise

We can define intelligent technology to correspond to domain and context-specific ability of the technical system for completing complex tasks, based on information processing. Within the logic of cognitive mimetics, the question is what is the corresponding source and aspects therein that explain intelligent action. Given variance in human proficiency in complex tasks, logically, the source should be the highest level of human ability in the same task and domain, i.e. expertise. One line of investigation into the basis of expertise has been mental contents (Saariluoma, 1995, 1997). The core idea, according to Myllylä and Saariluoma (2022), is to analyse what a particular piece of mental content can explain about some aspect of human action (or goal-directed behaviour). Thus, expertise and mental contents provide a coherent basis for content-based cognitive mimetics.

An essential component of intelligent information processing are mental representations, their contents, and transformations (Newell & Simon, 1972). Mental representations are mental objects with semantic properties (Pitt, 2022; Von Eckardt, 2012). A mental representation can be called a state or structure within a cognitive system which stands for something, i.e., is intentional, and thereby has contents (Frankish and Ramsey, 2012). Thus, in the present discussion, it makes reference to the semantic conceptions of information (Sequoiah-Grayson & Floridi, 2022) and semiotic processes (Deacon, 2003, 2007).

When taking the content-based view, an explanation of intelligent mental processes and thus actions is based on representational contents, rather than, for example, schemas, production systems or associative networks (Myllylä & Saariluoma, 2022; Saariluoma, Myllylä, et al., 2022), which are at base relatively content-neutral systems that provide various theoretical (and practical) scaffoldings for information contents and in particular, their relations and transformations. Importantly, the "contents as contents" starting point entails that the goal is not the explanation (naturalistic or otherwise) of contents by attempting a reduction to more primitive notion, conceptual or material. Rather, they provide an explanatory ground for intelligent action.

There are many ways to seek to frame mental contents, such as conceptual spaces (Gärdenfors, 2000) cognitive architectures (Anderson et al., 2004; Sun, 2001) or languages of thought (Fodor, 1975), which are relevant from the overall cognitive mimetic perspective. However, the most direct attack on the black box of intelligent action is to open it up *as such*, i.e. to *explicate mental contents* with respect to *action.* Mental contents, in this schema, explain intelligent action, the latter which is conceptually what intelligent technology should achieve. If mental contents provide explanatory grounds for intelligent action, they make sense as the source within the logic of mimetic design, under the assumption that intelligent technology in fact means the ability for intelligent action. This level of abstraction presents an autonomous level of investigation which cannot be reduced to any description, however valuable, of underlying "mechanisms" or other necessary realising conditions – brains in particular.

This way of framing the importance of mental contents finds corroboration in expert studies (Ericsson & Lehmann, 1996; Feltovich et al., 2006) and in AI is recognised most strongly in expert systems (Buchanan et al., 2006). In fact, it is possible to take the concepts of expertise and skills into the context of cognitive mimetics, given the framing that what intelligent technology actually means is the ability of the technology to complete complex tasks at a level exceeding human performance. From a mimetic perspective, expertise thus provides the supporting level of specification from the questions of contents. Thus, the fact that mental representations have different contents which accounts for success in tasks, points the way towards a specification of the source. Namely, given that experts have more non-perceptual contents in experience which accounts for their expertise (Myllylä & Saariluoma, 2022; Saariluoma, 1995), cognitive mimetics can be based on expert mental contents (Ericsson et al., 2018).

The concepts and methods in studies of expertise are thus highly relevant for content-based cognitive mimetics. Indeed, expertise studies (Feltovich et al., 2006) have found that it was untenable to maintain a distinction between domain-specific skills and general cognitive abilities. Domain knowledge is now seen as the dominant source expertise, and correspondingly, expertise is *highly* domain-specific, even in terms of basic cognitive abilities (Ericsson & Lehmann, 1996; Feltovich et al., 2006). It is interesting to note that the often remarked point that AI programs that achieve expert-level play in one game can't even get started in another is analogous to human expertise. This is both evidence for the specificity of domains and the idea that one can't get very far in domain-specific intelligent information processing using *general* mechanisms and high abstractions alone (Buchanan et al., 2006). One way or another domain-specific complexity needs to be mapped into intelligent technology and one way to think about the evolution of AI since the 1950s is by considering how the mapping process has changed emphasis from intelligence in the program code to intelligence in learning from massive amounts data (Jaakkola et al., 2022). One major phase were expert systems (Jaakkola et al., 2022), which provide an important reference point for content-based cognitive mimetics (Buchanan et al., 2006). While there are strong common ground methods and issues, from a cognitive mimetic perspective, expert systems (a vast field in itself (Liao, 2005) are best understood as an extant approach in AI to be interfaced, rather than identified with.

Above, we highlighted the conceptual importance of information contents for intelligent technology, which is a key to understanding correct and appropriate processing of mental information. This point of view does not exclude other aspects in the ecosystem of cognition, but simply highlights its importance from the point of view of explaining intelligent information processes – from the expert mental content point of view, the source is simply investigated from a scientific stance and the process is thus not intuitive or based on non-expert assumptions.

Cognitive mimetics can be used in a conceptually wide sense to capture a huge set of past and present ideas and concepts in AI, starting with Turing (1936) rigorous if introspective insight into the structure of action and thought of a human executing an algorithm. In fact, it is in a basic sense impossible to remove the human mental component from the design of AI programs, for they are always constructed by human minds and thus reflect human thinking under a broad definition (Boden, 1990; Freed, 2013; Jackson, 2013) This entanglement of human thought and artificial information processes provides crude but important evidence for the presence of mimetic thinking in computer science, and therefore AI as well. This provides some further background reasoning into why in our previous work (e.g. Saariluoma, Karvonen, et al., 2022), we have sought to develop the methodology of cognitive mimetics towards empirical research into human thought and action in a context. There, expert studies methods, including protocol analysis (Ericsson et al., 2018) were used to capture a simple but extendable cognitive control model based on empirical paper mill operator action and thought, called the IEC (ideal-exception-correction) model. This provides an example of the ideas that empirical cognitive mimetics can generate, but there is endless room for iteration and improvement. The reader is referred to these works for further details as in this article our focus is on more fundamental conceptual issues.

As the final part, we will discuss theory and design languages and iterate the primary components of a consciously developed language for cognitive mimetic design.

5. Conclusion: Fundamentals of a design language for cognitive mimetics

A key conceptual issue for stronger interaction between cognitive science and AI is developing frameworks for common discourse. This can be interpreted as a process of developing design languages. Professionals of all kinds live within their professional discourses (Goodwin, 2015; Saariluoma, 1997). Designers and design teams use design languages to communicate and realise their ideas. Scientists act and think based on their theory languages, as well as develop them further. Communication and therefore language emerges from common goal-seeking within constraints (Wilden, 1987). Large parts of common design languages have evolved through such processes. However, design languages can also be designed, which adds to their effectiveness (Gibbons et al., 2008; Rheinfrank & Evenson, 1996). A common conceptual ground is useful for getting the evolution going, which means that while the initial design language looks different years in the future, it is important to start somewhere.

This need can be conceptualised on two levels. From the perspective of the thematic field as a whole, there is a need for a metalanguage, a design science language that enables discourse across different specific perspectives with respect to the source of inspiration and the technological target. Such a language makes sense of some features of the overall endeavour. However, a metalanguage is not sufficient by itself, because a design language which embodies a perspective on the source and the target is needed to operationalise the work. Importantly, such a design language needs to interface with scientific theory languages as well as extant design languages for intelligent technology, corresponding to different approaches and layers of the design problems.

In the above, we have sought to specify certain elements of both the meta and design languages. The main concepts for the metalanguage are embodied in the SMT model (*source, target, mapping)*. Important concepts that specify the model are what we have called *levels*. These can be understood in various ways as determined through stances (Dennett, 1987), perspectives (Lee & Dewhurst, 2021), or levels of abstraction (Floridi, 2008). These, in turn, refer to different theory languages, which carve out and specify the source in different ways. These can be understood as posing relevant questions for all in this thematic field. The relation between source and target itself opens many conceptual questions, such as relevance of a particular stance on the source in terms of a design problem, as well as realisability in the technical context. As actual design processes and design thinking are complex and recursive processes, it is exceedingly difficult to capture them in a single abstract description (Parnas & Clements, 1986; Petre et al., 2019; Ralph, 2018; Vermaas & Dorst, 2007). This means that mimetic design remains an open world of possibilities which benefits from structure but should not be thought as reducable to formal operations. Actual design processes are "messy" (Parnas & Clements, 1986; Ralph, 2018): problem and

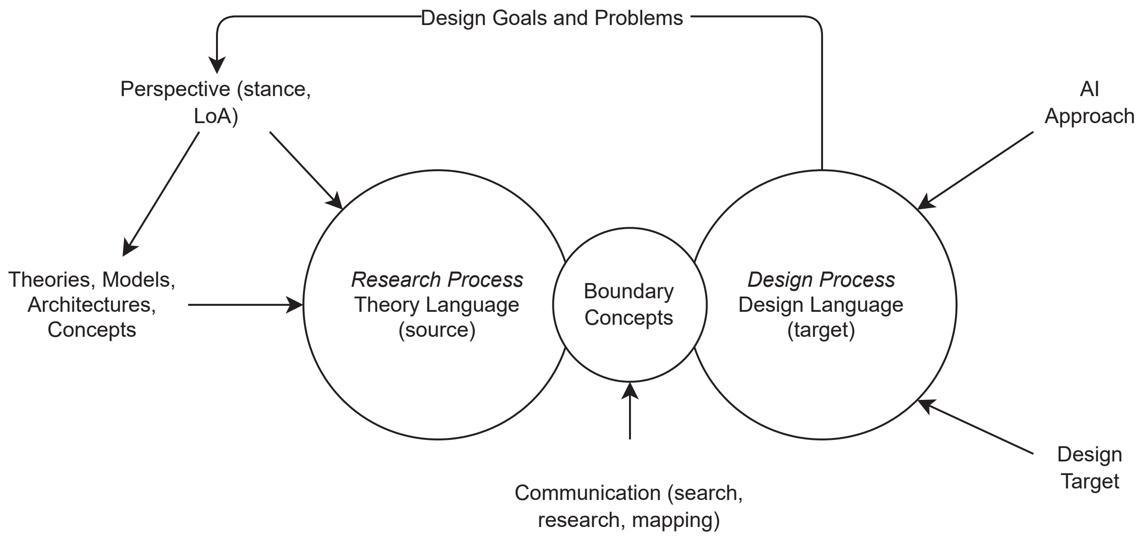

Fig. 2. A schema of some key properties and relations of cognitive mimetic design in terms of theory and design languages. Boundary concepts facilitate communication between scientific theory languages in the source analysis and design languages in the target design. This forms the basis for effective idea generation and mapping in design.

solution spaces coevolve; projects present unique events and sequences of events; improvisation is crucial and thus, the process is discursive, dialectical, heuristic, and exploratory (Schön, 2017). Nevertheless, the rational part of thinking cognitive mimetics is crucially concerned with making sense of the approach and clarifying the position for further refinement and development.

On the design language level, we have sought to illustrate what making commitments with respect to these elements can mean. First, we took information processing (in general) to be a reasonable interpretive and implementational framework for cognitive mimetics. Then, we turned attention to the target "as we find it", namely computers and computational information processing as the ground of the target. Third, noting the multifaceted nature of information (and cognition), we sought to highlight a specific aspect of information, its contents and representation. Finally, we discussed the relevance of mental contents for intelligent action in the source, which lead us to the importance of expertise and the fundamental problem of the mapping relation: How can mental contents be mapped to computational information processes? We can, without any problems, begin by using the concept of contents, and asking *how* could this be implemented computationally (Denning & Tedre, 2019), and readily many answers present themselves, such as ontologies (Chandrasekaran et al., 1999), concepts and methods in expert systems (Buchanan et al., 2006; Liao, 2005) and the learning processes in modern machine learning such as the generative pre-trained transformers (GPT), which are fine-tuned using so-called reinforcement learning with human feedback (RLHF) (Liu, 2023). From this perspective, how problems and solutions are framed is in part determined by the design language that is used.

Design languages enable communication, understanding, and reflective action within a design process, not only between people but within a single person in the way they are able to grasp and get a "handle" (Sheridan, 2017) on situations. Design languages evolve and are used both naturally and relatively unconsciously, but they can also be developed and designed for a purpose. The advantage of conscious efforts to develop design languages is that they can dramatically improve the designed objects from all perspectives and their position in life (Rheinfrank & Evenson, 1996; Saariluoma et al., 2016). An important aspect of a design language is that they are not necessarily formal (Rheinfrank & Evenson, 1996), which is essential in the context of cognitive mimetics based on the idea that computational systems design requires an informal language to balance the formal core of computation (Colburn & Shute, 2007; Rosen, 1999), that is, to assign meanings and information contents to the abstract machinery that computation is. Theoretically, this presents interesting problems for taking the content-based grounding of symbols (or patterns of activation in networks, etc.) forward (Coelho Mollo & Millière, 2023; Harnad, 1990; Liu, 2023). Importantly however, these are no longer purely theoretical problems, but *design* problems, once embedded within holistic design processes.

Rheinfrank and Evenson (1996) conceptualised the development of design languages as consisting of five steps: characterisation, re-registration, development and demonstration, evaluation, and evolution. In characterisation, existing underlying assumptions and preceding design languages are described. In re-registration, a new assumption set and design framework is created. In development and demonstration, the new design language is concretised in real design settings and their use demonstrated. In evaluation, the design language is evaluated in terms of its resonance with users and the results of the design processes. Evolution simply means that languages change when in use as additions are made and refinements are recognised during use.

The present article can be situated within the process of developing a design language for cognitive mimetics, particularly in the first two steps of the process. From the present perspective, the SMT model and the various conceptual discussions that follow should be understood as part of developing the language and its conceptual grounds. From this perspective, given different sources of inspiration, psychological or cognitive, new aspects of design languages should be developed, but it is believed that the common ground discussions and language elements will be useful for all work in this thematic area. These, furthermore, must interface with concepts in the domain of AI. For example, certain concepts in ML like *attention, context, reward,* and *prediction* provide interesting boundary concepts for interfacing between theory languages in cognitive science and AI (Silver et al., 2021; Sutton & Barto, 2018). On the other hand, concepts in content-based cognitive mimetics, such as *mental content, apperception, or restructuring* (Myllylä & Saariluoma, 2022; Saariluoma, 1995, 1997) can be used as research driven boundary concepts. The essential idea is that within the mimetic process, these concepts can be interpreted in the constraints of, for example, computational information processing or in terms of other aspects of the design of intelligent technology.

Key issues in developing cognitive mimetic on a conceptual level revolve around different conceptual languages in the source and the target, and the importance of developing boundary concepts to mediate and establish communication, and therefore a basis for the mapping relation (see Fig. 2). Two directions for future research present themselves for cognitive mimetics. One the one hand, theoretical work can be done for specifying explicit boundary concepts with respect to various design targets and source theory languages. This work is akin to conceptual engineering in philosophy (Chalmers, 2020). On the other hand, it is not assumed that such mediating design languages can be fully developed in the abstract, rather concepts gain meaning and evolve in use (Rheinfrank & Evenson, 1996; Wittgenstein, 1953). Thus, it is important to simply get to work, as we have proceeded by using the concrete boundary objects of human digital twins (HDTs) (Karvonen & Saariluoma, 2023; Saariluoma et al., 2020; Saariluoma, Karvonen, et al., 2022). These will feed new contents and concepts to the interface between the psychological and cognitive sources and intelligent technology targets. These, in turn, could in the future be organised into cognitive mimetic design patterns (Alexander, 1977; Gamma et al., 2009) that illustrate, communicate, and provide a basis for reuse. The structure and contents of these patterns is future work, but it can be based on the basic concepts and design languages illustrated in this article.

In summary, developing the work both theoretically and through practice will be important for cognitive mimetics, as it is the only way to truly develop a shared design language, and to demonstrate that in-depth human research can offer significant value to the design of intelligent technology in all its' aspects. In the present circumstances of rapid AI development, such work even carries a moral imperative.

Taking a broader view, emerging trends like the Society 5.0 program (Deguchi et al., 2020; Fukuyama, 2018) will shift the focus of technology design thinking. While Society 4.0 was cyber-physically oriented and based on the idea of intelligent technical artifacts, Society 5.0 includes human research in design discourses. Therefore, it is crucial to develop new and holistic design theoretical concepts to gain sufficiently broad perspectives on technology design and development. When designers must think about technical, cyber-physical, individual, and social intelligence, the design thinking must be holistic. In developing future intelligent technology design processes, cognitive mimetics and human digital twins offer effective tools.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Antero Karvonen reports financial support was provided by Business Finland.

Data availability

No data was used for the research described in the article.

Acknowledgements

This work was supported by Business Finland through the COACH project (3289/31/2022) as part of the SEED Forest ecosystem.

References

Adriaans, P. (2020). Information. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy (Fall 2020 ed.)*. Metaphysics Research Lab, Stanford University.

Aho, A. V., & Ullman, J. D. (1992). *Foundations of computer science*. Computer Science Press Inc.

Alexander, C. (1977). *A pattern language: Towns, buildings, construction*. Oxford University Press.

Allen, C. (2017). On (not) defining cognition. *Synthese, 194*(11), 4233–4249. https://doi.org/10.1007/s11229-017-1454-4

Anderson, J. R. (2015). *Cognitive psychology and its implications* (8th ed.). Worth Publishers.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological review, 111*(4), 1036.

Aspray, W. F. (1985). The scientific conceptualization of information: A survey. *Annals of the History of Computing, 7*(2), 117–140.

Bhushan, B. (2009). Biomimetics: Lessons from nature–an overview. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 367*(1893), 1445–1486. 10.1098/rsta.2009.0011.

Boden, M. (1990). *The philosophy of artificial intelligence*. New York: Oxford University Press.

Brown, B. R. (2022). *Engineering intelligent systems: Systems engineering and design with artificial intelligence, visual modeling, and systems thinking*. John Wiley & Sons, Incorporated.

Buchanan, B. G., Davis, R., & Feigenbaum, E. A. (2006). Expert systems: A perspective from computer science. In *The Cambridge handbook of expertise and expert performance* (pp. 87–103).

Bunge, M. (2017). *Causality and modern science*. Routledge.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.

Chakrabarti, A., Sarkar, P., Leelavathamma, B., & Nataraju, B. S. (2005). A functional representation for aiding biomimetic and artificial inspiration of new ideas. *AI EDAM, 19*(2), 113–132. https://doi.org/10.1017/S0890060405050109

Chalmers, D. J. (2011). A computational foundation for the study of cognition. *Journal of Cognitive Science, 12*(4), 325–359.

Chalmers, D. J. (2020). What is conceptual engineering and what should it be? *Inquiry*, 1–18.

Chandrasekaran, B., Josephson, J. R., & Benjamins, V. R. (1999). What are ontologies, and why do we need them? *IEEE Intelligent Systems and Their Applications, 14*(1), 20–26.

Clark, A. (1996). *Being there: Putting brain, body, and world together again*. The MIT Press. doi: 10.7551/mitpress/1552.001.0001.

Coeckelbergh, M. (2020). *AI ethics*. MIT Press.

Coelho Mollo, D., & Millière, R. (2023). The vector grounding problem. ArXiv, abs/2304.01481.

Colburn, T., & Shute, G. (2007). Abstraction in computer science. *Minds and Machines, 17*(2), 169–184. https://doi.org/10.1007/s11023-007-9061-7

Crane, T. (2015). *The mechanical mind: A philosophical introduction to minds, machines and mental representation*. Routledge.

Cross, N. (2006). *Designerly ways of knowing*. Springer.

Cross, N. (2011). *Design thinking: Understanding how designers think and work*. Berg.

Dasgupta, S. (2016). *Computer science: A very short introduction* (Vol. 466). Oxford University Press.

Deacon, T. W. (2003). Universal grammar and semiotic constraints. *Studies in the Evolution of Language, 3*, 111–139.

Deacon, T. W. (2007). Shannon–Boltzmann–Darwin: Redefining information (Part I). *Cognitive Semiotics, 1*(s1), 123–148.

Deacon, T. W. (2011). *Incomplete nature: How mind emerged from matter*. WW Norton & Company.

Deguchi, A., Hirai, C., Matsuoka, H., Nakano, T., Oshima, K., Tai, M., & Tani, S. (2020). What is society 5.0? In *Society 5.0: A people-centric super-smart society* (pp. 1–23). Springer Singapore. doi: 10.1007/978-981-15-2989-4_1.

Del Ser, J., Osaba, E., Molina, D., Yang, X.-S., Salcedo-Sanz, S., Camacho, D., Das, S., Suganthan, P. N., Coello Coello, C. A., & Herrera, F. (2019). Bio-inspired computation: Where we stand and what's next. *Swarm and Evolutionary Computation, 48*, 220–250. 10.1016/j.swevo.2019.04.008.

Dellermann, D., Ebel, P., Söllner, M., & Leimeister, J. M. (2019). Hybrid intelligence. *Business & Information Systems Engineering, 61*, 637–643.

Dennett, D. C. (1987). *The intentional stance*. MIT Press.

Denning, P. J., & Tedre, M. (2019). *Computational thinking*. MIT Press.

Dreyfus, H. L. (2007). Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. *Philosophical psychology, 20*(2), 247–268.

Egan, F. (2014). How to think about mental content. *Philosophical Studies, 170*(1), 115–135. https://doi.org/10.1007/s11098-013-0172-0

Eppe, M., Gumbsch, C., Kerzel, M., Nguyen, P. D., Butz, M. V., & Wermter, S. (2022). Intelligent problem-solving as integrated hierarchical reinforcement learning. *Nature Machine Intelligence, 4*(1), 11–20.

Ericsson, K. A., Hoffman, R. R., Kozbelt, A., & Williams, A. M. (2018). *The Cambridge handbook of expertise and expert performance*. Cambridge University Press.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47*(1), 273–305.

Feltovich, P. J., Prietula, M. J., & Ericsson, K. A. (2006). Studies of expertise from psychological perspectives. In *The Cambridge handbook of expertise and expert performance* (pp. 41–67). Cambridge University Press. https://doi.org/10.1017/CBO9780511816796.004.

Floridi, L. (2008). The method of levels of abstraction. *Minds and Machines, 18*(3), 303–329.

Floridi, L. (2009). Philosophical conceptions of information. In G. Sommaruga (Ed.), *Formal theories of information: From Shannon to semantic information theory and general concepts of information* (pp. 13–53). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-00659-3_2.

Fodor, J. A. (1975). *The language of thought* (Vol. 5). Harvard University Press.

Fortnow, L. (2012). The enduring legacy of the Turing machine. *The Computer Journal, 55*(7), 830–831. https://doi.org/10.1093/comjnl/bxs073

Frankish, K., & Ramsey, W. (Eds.). (2012). *The Cambridge handbook of cognitive science*. Cambridge University Press. http://oro.open.ac.uk/22094/.

Freed, S. (2013). Practical Introspection as Inspiration for AI. In V. C. Müller (Ed.), *Philosophy and theory of artificial intelligence* (pp. 167–177). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-31674-6_12.

Friston, K. J., Ramstead, M. J., Kiefer, A. B., Tschantz, A., Buckley, C. L., Albarracin, M., … & René, G. (2022). Designing ecosystems of intelligence from first principles. arXiv preprint arXiv:2212.01354.

Fu, K., Moreno, D., Yang, M., & Wood, K. L. (2014). Bio-inspired design: An overview investigating open questions from the broader field of design-by-analogy. *Journal of Mechanical Design, 136*(11), Article 111102.

Fukuyama, M. (2018). Society 5.0: Aiming for a new human-centered society. *Japan Spotlight, 27*(5), 47–50.

Gabriel, I. (2020). Artificial intelligence, values, and alignment. *Minds and Machines, 30* (3), 411–437. https://doi.org/10.1007/s11023-020-09539-2

Gamma, E., Helm, R., Johnson, R. E., & Vlissides, J. (2009). *Design patterns: Elements of reusable object-oriented software*. Addison-Wesley Professional.

Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*. The MIT Press. doi: 10.7551/mitpress/2076.001.0001.

Gibbons, A. S., Botturi, L., Boot, E., & Nelson, J. (2008). Design languages. In *Handbook of research on educational communications and technology* (pp. 633–645). Routledge.

Goel, A. K. (1997). Design, analogy, and creativity. *IEEE Expert, 12*(3), 62–70. https://doi.org/10.1109/64.590078

Goel, A. K. (2013). A 30-year case study and 15 principles: Implications of an artificial intelligence methodology for functional modeling. *AI EDAM, 27*(3), 203–215.

Goel, A. K., & Bhatta, S. R. (2004). Use of design patterns in analogy-based design. *Advanced Engineering Informatics, 18*(2), 85–94. 10.1016/j.aei.2004.09.003.

Goel, A. K., & Hancock, W. (2021). Multifunctional and domain independent? A meta-analysis of case studies of biologically inspired design. *Design Science, 7*, Article e24. https://doi.org/10.1017/dsj.2021.23

Goodwin, C. (2015). Professional vision. In *Aufmerksamkeit: Geschichte-Theorie-Empirie* (pp. 387–425). Springer.

Graeff, E., Maranzana, N., & Aoussat, A. (2020). Biological practices and fields, missing pieces of the biomimetics' methodological puzzle. *Biomimetics*, *5*(4), 62. https://www.mdpi.com/2313-7673/5/4/62.

Gruber, P., Bruckner, D., Hellmich, C., Schmiedmayer, H.-B., Stachelberger, H., & Gebeshuber, I. C. (2011). *Biomimetics—Materials, structures and processes: Examples, ideas and case studies*. Springer Science & Business Media.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena, 42* (1–3), 335–346.

Haugeland, J. (1997). *Mind design II: Philosophy, psychology, and artificial intelligence*. MIT Press.

Helms, M., Vattam, S. S., & Goel, A. K. (2009). Biologically inspired design: Process and products. *Design Studies, 30*(5), 606–622.

Hodges, A. (2013). Did church and turing have a thesis about machines?. In *Church's thesis after 70 years* (pp. 242–252). De Gruyter.

Hollnagel, E., & Woods, D. D. (2005). *Joint cognitive systems: Foundations of cognitive systems engineering*. CRC Press.

Jaakkola, H., Thalheim, B., & Henno, J. (2022). About the essence of intelligence-will artificial intelligence (ever) cover human intelligence?. In *Information modelling and knowledge bases XXXIII* (pp. 19–43). IOS Press.

Jackson, M. (2013). Formalism and intuition in software engineering. In *Perspectives on the future of software engineering: Essays in Honor of Dieter Rombach* (pp. 335–347).

Jiang, S., Hu, J., Wood, K. L., & Luo, J. (2021). Data-Driven design-by-analogy: State-of-the-art and future directions. *Journal of Mechanical Design, 144*(2). https://doi.org/10.1115/1.4051681

Kannengiesser, U., & Gero, J. S. (2015). Is designing independent of domain? Comparing models of engineering, software and service design. *Research in Engineering Design, 26* (3), 253–275.

Kar, A. K. (2016). Bio inspired computing – A review of algorithms and scope of applications. *Expert Systems with Applications, 59*, 20–32. 10.1016/j.eswa.2016.04.018.

Karvonen, A., & Saariluoma, P. (2023). Cognitive mimetics and human digital twins – Towards holistic AI design. In *ERCIM news/cognitive AI & Cobots*, (132). https://ercim-news.ercim.eu/images/stories/EN132/EN132-web.pdf.

Karvonen, A. (2020). Cognitive mimetics for AI ethics: tacit knowledge, action ontologies and problem restructuring. In *Culture and computing: 8th International conference, C&C 2020, held as part of the 22nd HCI international conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings 22*.

Kujala, T., & Saariluoma, P. (2018). Cognitive Mimetics for Designing Intelligent Technologies. Advances in Human-Computer Interaction, 2018, 9215863. https://doi.org/10.1155/2018/9215863.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences, 40*, Article e253. https://doi.org/10.1017/S0140525X16001837

LeCun, Y. (2022). A path towards autonomous machine intelligence version 0.9. 2, 2022–06-27. *Open Review, 62*.

Lee, J., & Dewhurst, J. (2021). The mechanistic stance. *European Journal for Philosophy of Science, 11*(1), 20. https://doi.org/10.1007/s13194-020-00341-6

Lenau, T., Metze, A.-L., & Hesselberg, T. (2018). Paradigms for biologically inspired design. *SPIE.* , Article 10593. https://doi.org/10.1117/12.2296560

Lepora, N. F., Verschure, P., & Prescott, T. J. (2013). The state of the art in biomimetics. *Bioinspiration & Biomimetics, 8*(1), Article 013001. https://doi.org/10.1088/1748-3182/8/1/013001

Liao, S.-H. (2005). Expert system methodologies and applications—a decade review from 1995 to 2004. *Expert Systems with Applications, 28*(1), 93–103.

Lieto, A. (2021). *Cognitive design for artificial minds*. Routledge.

Linja, J., Hämäläinen, J., Nieminen, P., & Kärkkäinen, T. (2023). Feature selection for distance-based regression: An umbrella review and a one-shot wrapper. *Neurocomputing, 518*, 344–359. 10.1016/j.neucom.2022.11.023.

Littman, M. L., Ajunwa, I., Berger, G., Boutilier, C., Currie, M., Doshi-Velez, F., Hadfield, G., Horowitz, M. C., Isbell, C., & Kitano, H. (2021). *Gathering strength, gathering storms: The one hundred year study on artificial intelligence (AI100) 2021 study panel report*. Stanford University. Retrieved: September, 16, 2021.

Liu, G. K.-M. (2023). Perspectives on the social impacts of reinforcement learning with human feedback. *arXiv preprint arXiv:2303.02891*.

Longuet-Higgins, C. (1973). *Lighthill report: Artificial intelligence: A paper symposium*. U. S. a. T. F. Council. http://www.chilton-computing.org.uk/inf/literature/reports/lighthill_report/contents.htm.

Lucas, J. (1996). Minds, machines and Gödel: A retrospect. *Artificial Intelligence: Critical Concepts, 3*, 359–376.

Moreno, D. P., Hernandez, A. A., Yang, M. C., Otto, K. N., Hölttä-Otto, K., Linsey, J. S., Wood, K. L., & Linden, A. (2014). Fundamental studies in Design-by-Analogy: A focus on domain-knowledge experts and applications to transactional design problems. *Design Studies, 35*(3), 232–272.

Muggleton, S., & Chater, N. (2021). *Human-like machine intelligence*. Oxford University Press, Incorporated.

Myllylä, M. T., & Saariluoma, P. (2022). Expertise and becoming conscious of something. *New Ideas in Psychology, 64*, Article 100916. 10.1016/j.newideapsych.2021.100916.

Newell, A., & Simon, H. A. (1972). *Human problem solving* (Vol. 104). Englewood Cliffs, NJ: Prentice-Hall.

Newell, A., & Simon, H. (1976). Computer science as empirical inquiry: symbols and Search. *Communications of the ACM, 19*.

Nilsson, N. J. (2009). *The quest for artificial intelligence*. Cambridge University Press. doi: 10.1017/CBO9780511819346.

Oulasvirta, A., Kristensson, P. O., Bi, X., & Howes, A. (2018). *Computational interaction*. Oxford University Press.

Parnas, D. L., & Clements, P. C. (1986). A rational design process: How and why to fake it. *IEEE Transactions on Software Engineering, 2*, 251–257.

Peffers, 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

Penrose, R. (1990). Précis of The emperor's new mind: Concerning computers, minds, and the laws of physics. *Behavioral and Brain Sciences, 13*(4), 643–655.

Petre, M., van der Hoek, A., & Bowers, D. S. (2019). Software design as multiple contrasting dialogues. In *Psychology of programming interest group 30th annual conference*. Newcastle University. http://oro.open.ac.uk/66486/.

Piccinini, G. (2004). Functionalism, computationalism, and mental contents. *Canadian Journal of Philosophy, 34*(3), 375–410.

Piccinini, G., & Scarantino, A. (2010). Computation vs. information processing: Why their difference matters to cognitive science. *Studies in History and Philosophy of Science Part A, 41*(3), 237–246.

Piccinini, G., & Scarantino, A. (2011). Information processing, computation, and cognition. *Journal of Biological Physics, 37*, 1–38.

Pitt, D. (2022). Mental representation. In E. N. Zalta, and U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Fall 2022 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2022/entries/mental-representation/.

Qian, L., & Gero, J. S. (1996). Function–behavior–structure paths and their role in analogy-based design. *AI EDAM, 10*(4), 289–312.

Ralph, P. (2018). The two paradigms of software development research. *Science of Computer Programming, 156*, 68–89. 10.1016/j.scico.2018.01.002.

Rapaport, W. J. (2012). Semiotic systems, computers, and the mind: How cognition could be computing. *International Journal of Signs and Semiotic Systems (IJSSS), 2*(1), 32–71.

Rheinfrank, J., & Evenson, S. (1996). Design languages. In T. Winograd (Ed.), *Bringin design to software*. The ACM Press.

Rosen, R. (1999). *Essays on life itself*. Columbia University Press. http://ebookcentral.proquest.com/lib/jyvaskyla-ebooks/detail.action?docID=909006.

Russell, S. J., & Norvig, P. (2010). *Artificial intelligence a modern approach*. Pearson Education Inc.

Saariluoma, P. (1995). *Chess players' thinking: A cognitive psychological approach*. Psychology Press.

Saariluoma, P. (1997). *Foundational analysis: Presuppositions in experimental psychology* (Vol. 2). Psychology Press.

Saariluoma, P., & Karvonen, A. (2023). Theory languages in designing artificial intelligence. In *AI & Society*. Springer Science and Business Media LLC. doi: 10.1007/s00146-023-01716-y.

Saariluoma, P., Cañas, J. J., & Leikas, J. (2016). *Designing for life: A human perspective on technology development*. Springer.

Saariluoma, P., Kujala, T., Karvonen, A., & Ahonen, M. (2018). Cognitive mimetics: Main ideas. In *International conference on artificial intelligence*.

Saariluoma, P., Cañas, J., & Karvonen, A. (2020). Human digital twins and cognitive mimetic. In *International conference on human interaction and emerging technologies*.

Saariluoma, P., Myllylä, M., & Karvonen, A. (2022). Mental contents in designing AI ethics. In *Culture and computing: 10th international conference, C&C 2022, held as part of the 24th HCI international conference, HCII 2022, virtual event, June 26–July 1, 2022, Proceedings*.

Saariluoma, P., Karvonen, A., & Sorsamäki, L. (2022). Human digital twins in acquiring information about human mental processes for cognitive mimetics. In *Information modelling and knowledge bases XXXIII* (pp. 163–176). IOS Press.

Saariluoma, P., Karvonen, A., Wahlstrom, M., Happonen, K., Puustinen, R., & Kujala, T. (2019). Challenge of tacit knowledge in acquiring information in cognitive mimetics. *International conference on intelligent human systems integration*.

Saariluoma, P., & Rauterberg, M. (2016). Turing's error-revised. *International Journal of Philosophy Study, 4*.

Samsonovich, A. V. (2012). On a roadmap for the BICA challenge. *Biologically Inspired Cognitive Architectures, 1*, 100–107. 10.1016/j.bica.2012.05.002.
Sayre, K. (1987). Cognitive science and the problem of semantic content. *Synthese*, 247–269.
Schön, D. A. (2017). *The reflective practitioner: How professionals think in action*. Routledge.
Schumann, J. F., Srinivasan, A. R., Kober, J., Markkula, G., & Zgonnikov, A. (2023). Using models based on cognitive theory to predict human behavior in traffic: A case study. arXiv preprint arXiv:2305.15187.
Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(3), 417–424.
Sequoiah-Grayson, S., & Floridi, L. (2022). Semantic conceptions of information. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2022 ed.). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2022/entries/information-semantic.
Sheridan, T. (2017). *Modeling human system interaction: Philosophical and methodological considerations, with examples* [Book]. Wiley. https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1442613&site=ehost-live.
Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal, 27*(3), 379–423.
Silver, D., Singh, S., Precup, D., & Sutton, R. S. (2021). Reward is enough. *Artificial Intelligence, 299*, Article 103535.
Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology, 41*(1), 1–20.
Simon, H. A. (1995). Artificial intelligence: An empirical science. *Artificial Intelligence, 77*(1), 95–127.
Simon, H. A. (2019). *The sciences of the artificial*.
Song, H., & Fu, K. (2022). Design-by-analogy: Effects of exploration-based approach on analogical retrievals and design outcomes. *Journal of Mechanical Design, 144*(6). https://doi.org/10.1115/1.4053683
Sun, R. (2001). *Duality of the mind: A bottom-up approach toward cognition* (1st ed.). Psychology Press.
Sutton, R. (2019). The bitter lesson. *Incomplete Ideas (blog), 13*(1).
Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT Press.
Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *Journal of Mathematics, 58*(345–363), 5.
Vattam, S. S., Helms, M. E., & Goel, A. K. (2010). A content account of creative analogies in biologically inspired design. *AI EDAM, 24*(4), 467–481. https://doi.org/10.1017/S089006041000034X
Vermaas, P. E., & Dorst, K. (2007). On the conceptual framework of John Gero's FBS-model and the prescriptive aims of design methodology. *Design Studies, 28*(2), 133–157. 10.1016/j.destud.2006.11.001.
Vincent, J. F. V., Bogatyreva, O. A., Bogatyrev, N. R., Bowyer, A., & Pahl, A.-K. (2006). Biomimetics: Its practice and theory. *Journal of the Royal Society Interface, 3*(9), 471–482. 10.1098/rsif.2006.0127.
Von Eckardt, B. (2012). The representational theory of mind. In K. Frankish, & W. M. Ramset (Eds.), *The Cambridge handbook of cognitive science*. Cambridge University Press.
Wiener, N. (1985). *Cybernetics: Or control and communication in the animal and the machine (second edition)*. The MIT Press.
Wilden, A. (1987). *The rules are no game: The strategy of communication*. Routledge & Kegan Paul Ltd.
Wittgenstein, L. (1953). *Philosophical investigations*. Macmillan.
Zador, A., Escola, S., Richards, B., Ölveczky, B., Bengio, Y., Boahen, K., Botvinick, M., Chklovskii, D., Churchland, A., Clopath, C., DiCarlo, J., Ganguli, S., Hawkins, J., Körding, K., Koulakov, A., LeCun, Y., Lillicrap, T., Marblestone, A., Olshausen, B., … Tsao, D. (2023). Catalyzing next-generation Artificial Intelligence through NeuroAI. *Nature Communications, 14*(1), 1597. https://doi.org/10.1038/s41467-023-37180-x