Computers and Education: Artificial Intelligence 3 (2022) 100107

Contents lists available at ScienceDirect

Computers and Education: Artificial Intelligence

journal homepage: www.sciencedirect.com/journal/computers-and-education-artificial-intelligence

Human and artificial cognition

George Siemens [a,b,*], Fernando Marmolejo-Ramos [a], Florence Gabriel [a], Kelsey Medeiros [c], Rebecca Marrone [a], Srecko Joksimovic [a], Maarten de Laat [a]

[a] University of South Australia, Australia
[b] University of Texas, Arlington, USA
[c] The University of Nebraska, Omaha, USA

ARTICLE INFO

Keywords:
Cognition
Artificial intelligence
Human-machine collaboration
Knowledge processing

ABSTRACT

Predictions of the timelines for when machines will be able to perform general cognitive activities that rival humans, or even the arrival of “super intelligence”, range from years to decades to never. For researchers in the education sector, the potential future state of AI, while provocative, is secondary to important shorter-term questions that influence how AI is integrated into learning and knowledge practices such as sensemaking and decision making. AI is not a future technology. It is already present in our daily lives, often shaping, behind the scenes, the types of information we encounter. It is, therefore, important to consider immediate questions surrounding the dynamics of human-machine interactions. In this paper, we focus on the relationship between human and artificial cognition and treat these as separate systems, each with distinct strengths and capabilities. We adopt a functional view (i.e., discrete tasks) of the activities that artificial cognition completes and those that are best handled by humans. This creates a foundation to then evaluate models for how these two cognitive systems interact and the mechanisms for coordination that are required. In doing so, we create a basis for future researchers to develop testable hypotheses regarding the impact of artificial cognition on knowledge processes such as learning, sensemaking, and decision making. Our evaluation provides insight for researchers regarding the optimal relationship between which cognitive activities should be handed off to the machine, which should remain the domain of human performance, and how these two should then be integrated when outputs are passed from one cognitive system (human or artificial) to the other.

1. Introduction

Forecasts of the impact of artificial intelligence on society favour the extreme and dramatic. The anticipated outcomes range from AI-fuelled mass unemployment (Bordot, 2022) to vastly reduced work weeks (TUC - Trade Union Congress, 2018), to robots for everyone (MIT, 2022), to sentient agents that are humanity’s final invention, with contradicting severity of outcomes offered by tech gurus such as Bill Gates and Elon Musk and celebrity scientists such as the late Stephen Hawking. The emotional, fatalistic, and utopian proclamations rarely leave room for nuanced consideration, presenting both researchers and the public with unclear expectations of the real potential of AI. AI, however, is not a future problem or opportunity. It is one that is already deeply woven into our daily lives (Stone et al., 2016). Banking security, personal assistants (Siri, Alexa, Google), predictive text while searching, and content and friend recommendations on social media are a daily, often behind the scenes, presence for internet users. The type of information people see, and the kinds of ideas that they interact with, and the communities they belong to are shaped by algorithms.

While much of the dialogue is future focused with fatalistic tones, AI is a present-day mundane technology already impacting us in subtle ways in all areas of our lives and increasingly in our learning, sensemaking, and decision making (complex knowledge processes). Over the past several decades, as robots took over many manual tasks on factory floors, jobs in the domain of cognition and knowledge work (medical staff, legal scholars, scientists) were often seen as immune from threats of automation or displacement. Today, however, these jobs are anticipated to be impacted as much, if not to a greater degree, than traditional routine labour. Lawyers, doctors, financial advisors, and middle managers now find their work at risk of automation (Caron, 2019; Fenwick &

* Corresponding author. University of South Australia, Australia.
E-mail addresses: George.Siemens@unisa.edu.au (G. Siemens), Fernando.Marmolejo-Ramos@unisa.edu.au (F. Marmolejo-Ramos), Florence.Gabriel@unisa.edu.au (F. Gabriel), kelseymedeiros@gmail.com (K. Medeiros), Rebecca.Marrone@unisa.edu.au (R. Marrone), Srecko.Joksimovic@unisa.edu.au (S. Joksimovic), Maarten.deLaat@unisa.edu.au (M. de Laat).

https://doi.org/10.1016/j.caeai.2022.100107
Received 15 August 2021; Received in revised form 7 November 2022; Accepted 7 November 2022
Available online 8 November 2022

Vermeulen, 2020; Goldhahn et al., 2018). As AI moves into the domain of human knowledge processes, important considerations arise regarding the type of processes that should be retained by humans and those that should be done by AI.

The critical question confronting all sectors of society is one of *interaction between human and artificial cognitive systems during complex knowledge processes such as learning, sensemaking, and decision making.* Some researchers have argued that work categories will not be entirely automated, but that specific parts of jobs will be handled computationally (Brynjolfsson et al., 2018). This is already evident in environments where coordination of complex tasks is required, such as the work of pilots. Much of the monitoring of an aircraft, including the anticipation of turbulence, is handled by automated systems. While automation does not necessarily mean AI, automation does provide an early indicator of what human-machine interactions will look like in the future where collaboration and coordination between these two systems is required (Weick & Roberts, 1993). Similarly, processing large quantities of real-time data, such as in financial markets, has created an ecosystem where humans are no longer able to intervene meaningfully in real time (Johnson et al., 2013). In this sense, technology improves the capability of pilots and financial analysts for example, to better perform their work by doing highly complex computation out of scope of human capability and even awareness, but still directly impacting knowledge processes.

It is for this reason that we argue that AI is too broad a term to be practically useful for researchers, educators, government, and even business leaders. A focus on cognition, as opposed to intelligence, moves the conversation around human-machine interactions from a high-level, philosophical debate, to one where concrete and discrete cognitive processes and activities can be identified and practically allocated to the agent best suited to complete it. This produces two cognitive systems (human and artificial) with each doing cognitive work in different ways, at different scale, and at a different pace. Integration and coordination between the outputs of both systems is an area in need of additional research. What are the principles whereby humans integrate the outputs of artificial cognition back into the human knowledge system so ongoing coordination and subsequent actions are meaningfully translated into the domain of human decision making? Similarly, what are the guiding principles that determine how human cognitive activities are integrated back into artificial cognitive systems and vice versa? These questions point to a substantial area of needed research because clear roles and relationships between these two cognitive systems have not yet been systematically theorized, resulting in limited guiding frameworks of knowledge processes where these systems intersect, as illustrated in Fig. 1.

With the prevalence of emerging technologies that encroach on activities once the exclusive domain of humans, it is time to conceptualize models of intersection between human and artificial cognition. To evaluate how both human and artificial cognition will engage in knowledge processes, and ultimately impact society, we begin by reviewing basic concepts regarding human cognition and elaborate on what artificial cognition could be like (section 2). We then review how AI has advanced (section 3) and the possible effects it might have on human knowledge systems and consider instances of interactions between human and artificial cognition (AC) (sections 4 and 5). Finally, we consider ways that AC might impact society in the long run, emphasizing the ongoing research challenges of integrating AC outputs into human knowledge practices and systems (section 6).

2. Human and artificial cognition

The focus of this paper is on the intersection that arises between human and artificial cognition and the coordinating mechanisms that are required for outputs from one system to be passed on to the other. There is limited consensus on the definition of *human cognition* (Bayne et al., 2019). We commence by defining cognition as the **sensory processes, general operations, and complex integrated activities involved in interacting with information.** *Sensory processes* include vision, perception, and attention. *General operations* involve language, memory, recognition, recall, information seeking and management behaviours. *Complex integrated activities* include reasoning, judgement, decision making, problem solving, sensemaking, and creativity.

Within each of these three primary areas of cognition, various elements can be handled by either human cognition or artificial cognition. De Winter and Dodou (2014) detail one of the earliest attempts, in 1951, to allocate cognitive tasks to machines or humans: Fitts List. This list allocated functions based on which agent, human or machine, would be best able to perform the task. A total of 11 statements were offered around actions such as automation (machine) and deductive reasoning (human). More recently, Roth et al. (2019), detailed the allocation of tasks across technology systems in general. Our approach is to emphasize which *distinct* cognitive task can be handled or augmented by technology and which should remain in the domain of human

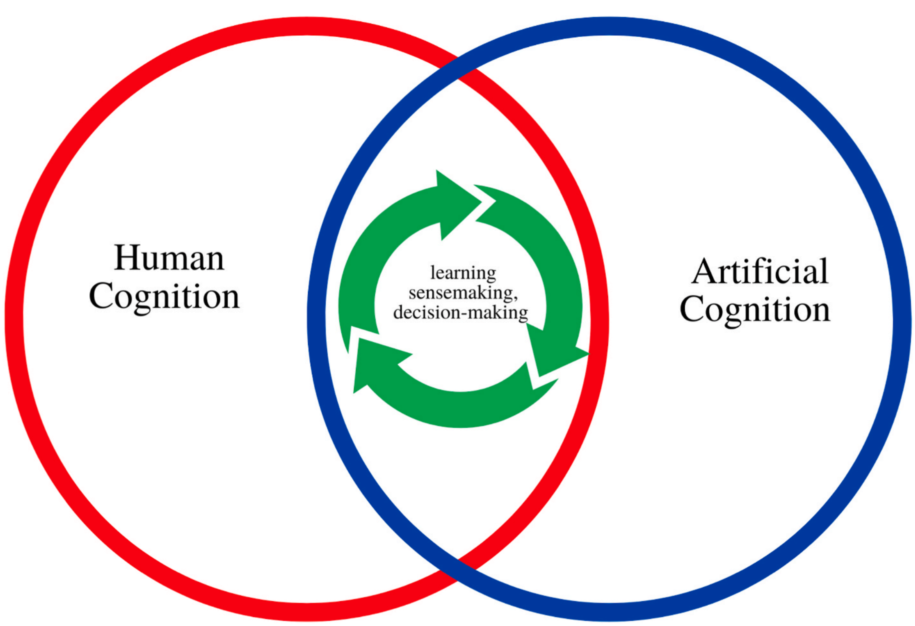

Fig. 1. Space of coordination between human cognition and artificial cognition.

performance – the intersection of overlap detailed in Fig. 1. An important point to emphasize: evaluation models of human or artificial cognitive processing should involve more than assessing efficiency and productivity of outputs. Considerations of bias, ethics, suitability, and long-term impacts on individuals and society are equally consequential. For example, the bias embedded in AI systems will carry over into human systems. Surfacing the ways that AI already impacts complex knowledge processes is key to counteracting its influence.

2.1. What could artificial cognition look like?

To better grasp the relationship between human and artificial cognition, it is essential to understand human cognition. However, to arrive at a definition, it is necessary to briefly review some of the historical antecedents of our proposed definition. Doing so, will enable us to propose the relationship between human and artificial cognition. Attempts to articulate the nature of artificial cognition have recently been undertaken. Spring boarding from the work by Ritter et al. (2017), Taylor and Taylor (2021) define artificial cognition as the use of experimental psychology for understanding, evaluating, and explaining machine learning (ML) algorithms benchmarked via human performance. This new research line is also akin to the emerging field of 'anthropomorphic ML' that aims to produce ML algorithms that learn the way humans do (Angelov & Gu, 2018). Our interest is not in artificial cognition as defined by Taylor and Taylor where the focus is on using psychology research methods to understand and explain AI; our interest is in those aspects of AI that are cognition-like and share overlapping capabilities with humans. In this sense, our interest is to understand the nature of human-like cognitive processes being conducted by automated or AI systems that outperform, interact with, or even change human cognition.

Theories of cognition have a broad range, but often settle on two analogies: (1) those using the computer as an analogy and (2) those resorting to the brain as an analogy. Research using the computer analogy understands cognition as information processing, while research focused on the brain analogy is based on cognition as networked, situated, and dynamic. Additionally, while the computer-analogy approach is task-oriented in that it is about inputs and outputs occurring in specific tasks, the brain-analogy approach aims to match the internal structure of models to the internal structure of (neuro) biologically inspired systems. In other words, the first approach is functional and specific, while the second is structural and general (see Bermúdez, 2014; Lieto, 2021). Theorists have argued for the need to have a hybrid model where both symbolic AI (computational) and neural networks (brain-based) contribute to artificial intelligence (Marcus & Davis, 2019).

A key difference between the approaches is that while computer-based models of cognition use abstract and amodal symbols, the brain-based models argue that cognition operates with modal symbols grounded in sensorimotor experience. This second approach is commonly known as embodied cognition and recognizes that cognition is often extended into objects and artifacts outside of just the human brain. This difference in approaches translates into what is known as the symbol grounding (SG) problem (see Harnad, 1990); that is, how symbols (e.g. words and concepts) gain meaning. A computational account cannot provide a response to the SG problem as in that approach abstract symbols are connected to other abstract symbols. Meaning finds no grounding as there is no referent.

Harnad (1990) proposed that the SG problem could be solved via sensory projections of distal objects and events (iconic representations) and learned or innate ability to detect invariant features of those objects and events (categorical representations). Such a theoretical solution was later formalised by (Barsalou, 1999) in his perceptual theory of knowledge (i.e. perceptual symbol system; PSS). Barsalou relied on cognitive neuroscience to argue that cognition is not about manipulating amodal symbols handled in modular systems away from the brain's systems for action, perception, and introspection. Instead, under the PSS view, cognition happens when perceptual and motor and bodily systems are engaged during interactions with the environment and multi-modal mental simulations ensue (see Barsalou, 2008). About the same time the PSS view was being proposed, a core cognitive system was being redefined by Glenberg (1997), who argued that memory was for storing concepts intrinsically linked to patterns of actions occurring to meet environmental requirements. Regardless of the specific approach, networks as a conceptual framework now underpin most prominent cognitive models (see Barbey, 2018) and the prevalence of networks and connections, both biologically (the human brain) and computationally (neural networks), is central to our formulation of HAC.

Theoretical cognitive models, or computational implementations based on the two analogies described earlier, are instantiated in cognitive architectures (CA). The two most well-known are the Adaptive Control of Thought - Rational (ACT-R) and SOAR, but there are many others (for a comprehensive review see Kotseruba & Tsotsos, 2020). A key difference between CAs is that symbolic-inspired CAs do not resort to perceptual inputs; instead, they use direct data (e.g., ACT-R, SOAR). Biologically inspired CAs resort to multiple sensory and motor inputs via physical and simulated sensors (e.g., iCub, BBD, DAC, MDB). The biologically-inspired CAs have shown great promise in that they are akin to how the human brain achieves feature integration (Zmigrod & Hommel, 2013) by accounting for information coming from the senses (i.e. vision, audition, etc). Overall, CAs and cognitive models indicate that an artificial-cognitive agent mimicking human cognition should, at minimum, i) go beyond simple pattern recognition by building causal models of the world (e.g. via predictive coding), ii) ground concepts (formed during learning) in physical and social environments (Marmolejo-Ramos et al., 2017), and iii) be able to generalise (e.g. as in Harnad, 1990, categorical representations) (Lake et al., 2017). A fourth element is that of emotion processing; suggesting that an AI agent should iv) be able to produce and understand emotions in a meaningful way. Although emotions are not commonly associated with definitions of cognition, neuroscientific (Holstege, 1992) and experimental evidence (Marmolejo-Ramos et al., 2020) points to the close coupling between sensory and motor systems for the processing of emotions. Hence, it is being argued that AI agents should implement emotion processing capabilities to approximate social cognition (Larue et al., 2018; Pessoa, 2017; Ziemke & Lowe, 2009).

Connectionism is an approach used to explain cognition that is neither fully embodied nor fully computational. Connectionism is not fully computational as it does not resort to symbolic computations, and it is not embodied as it does not implement sensory and motor systems in the brain. However, precisely because connectionist models are inspired by the structure and function of neurons (i.e., a low-level structure of the brain), it closely mimics the way the brain processes information (groups of neurons are represented as networks and their synapses as connections). Such neuronal approaches allow connectionist models to define learning as the storage of information from the environment in the form of neural networks (i.e., as connections between neurons). A key appeal of connectionism is that it tolerates building computational implementations that can be used as testbeds of specific cognitive processes. Interestingly, the cognitive processes investigated can come from frameworks seemingly different from connectionist models. For example, Flusberg et al. (2010) demonstrated from a connectionist approach (via artificial neural networks) behavioral findings favouring the embodied conceptual metaphor. More recently, there have been proposals to blend neural networks with Bayesian probabilistic models of perception (McClelland, 2013) and generative models (as in 'active inference'; aka the free energy principle; see Constant et al., 2021). Adopting a networked view, more akin with current proposals that regard human neuroarchitecture as combinatorial, complex, and dynamic (Pessoa et al., 2022) thus, would allow proposing a domain-general definition of cognition that considers a) sensory processes, b) general operations (e.g., via simple learning rules), and c) complex integrated activities involved in interacting with information (e.g., via distributed representations and top-down/bottom-up interactive processing). Thus, our definition above emphasizes complex interconnective cognitive processes (see also Bayne et al., 2019, for other definitions). We believe this definition allows taking the best from embodied and symbolic perspectives in a way that artificial and human cognitive processes can be talked about using similar terminology and concepts. It also sets a foundation, by normalizing and coordinating cognitive processes, for determining which system, human or artificial, should perform which tasks and activities.

Based on the above review of distinct cognition processes, AI is concerned more with the machines and programs that can learn on their own, without being explicitly programmed by people. Artificial cognition, in contrast, is concerned with distinct cognitive processes that are automated or handled by a machine or program. This contrasts with Taylor & Taylor's focus on artificial cognition as contributing to explainable AI, but does align with researchers who focus on artificial cognition as extending and augmenting human cognition (see Raisamo et al., 2019). While specific definitions are still emerging, our primary focus is to move away from an emphasis on *intelligence* and use *cognition* instead to *describe processes conducted by an artificial agent and the outputs then being integrated back into a human knowledge system* (and vice versa). As indicated in Fig. 1, and detailed below, conceptual models are required that permit human and artificial cognition to work together in solving complex tasks, regardless of the broader conceptual frameworks (computational, embodied, or connectionist) adopted by researchers.

3. Review of artificial cognition technologies and innovations

Despite the progress made by AI researchers, a major question in the field remains: "Can machines develop cognitive abilities comparable to the human brain?". A recurring science-fiction theme sees the machines attaining, and ultimately surpassing, human intelligence often with dire consequences as memorably depicted in films such as Ex Machina or The Terminator. This type of (super)human level of artificial cognition is commonly referred to as artificial general intelligence (AGI), where machines can reason autonomously and solve problems beyond those they were trained to solve. Although the research into AGI is active and ongoing (for an example see; Kotseruba & Tsotsos, 2020), current efforts fall well short of human-like intelligence. The pursuit of narrower domains of artificial cognition (including fields such as computer vision, natural language processing, decision making) have been more successful in developing intelligent models. In this section we will discuss the latest trends in narrow artificial cognition and their impacts on research and society.

3.1. Current trends in artificial cognition

A major strength of current artificial-cognitive tools is their ability to investigate complex patterns from large amounts of data and derive reasonable "rules" for taking new data and making decisions, predictions, or adaptations. This data-driven approach has allowed artificial-cognition researchers to produce models with superhuman performance in fields such as image classification, game-playing, and text comprehension. In what follows, we discus deep learning, as an arguably crucial technology in the development of artificial cognition. Further, we reflect on two most used applications of artificial cognition – computer vision and natural language processing. These two fit into the first two components of our three component cognition definition: sensory processes and general operations (the third being complex integrated activities which is generally not evident in current AI research and would more accurately fit into the domain of AGI).

3.1.1. Deep learning

Most recent advances in artificial cognition stem from developments in deep learning and big data, and the rise of deep learning has been advanced by Graphics Processing Units (GPUs) and the availability of unprecedented amounts of data for training neural networks (Bengio et al., 2021). Exploiting a particular form of compositionality that allows for features in one layer to be combined in more abstract feature sets in the next layer is a driving principle that allows deep learning models to outperform previous approaches (Bengio et al., 2021).

Most of the traditional applications of deep learning have been built on supervised models. That is, they were trained to minimise prediction error against some known target. **Reinforcement Learning** (RL) differs in that the target of interest is often abstract from the decisions available to the machines at any one point. For example, consider a game of chess. At any point in the game, the actor (human or machine) has the option to move any of their pieces in any valid way. This move may happen a long way from the end of the game, yet we want to have some way of learning how this move affects the outcome; win, lose or draw. RL is a means of training a deep learner to associate end-game outcome with all the moves that happened in the earlier stages. Over the course of many iterations, it learns what course of action it should take at any one time to maximise its chance of victory.

In 2016, a team from DeepMind (now part of Google) famously trained a model, AlphaGo, through RL that was able to beat a top-rated player in the game of Go. Go had been a difficult game for computers to learn due to the exponentially large number possible moves available at most points in the game (Silver et al., 2016). As an RL model, AlphaGo learned from an archive of Go games and by playing against itself to learn policies to maximise its chances of victory. In doing so, it discovered strategies that had been unknown to human competitors. Since then, DeepMind developed AlphaZero, which, after being trained solely via "self-play" (i.e. no information from the archives), reached a superhuman level of play, not only in Go, but also chess and shogi (Silver et al., 2018).

The above referenced examples would fit under the *general operations*, as defined in the proposed view of cognition. Beyond game-playing, DeepMind has also developed an AI system called AlphaFold to predict the 3-D structure of proteins based solely on their amino acid sequences (Evans et al., 2018), which is also aligned with *general operations*, as it involves processing information. Predicting protein structures is a task that has challenged scientists for decades and is hoped to provide vital insights to diagnose and treat diseases caused by misfolded proteins, such as Alzheimer's and Parkinson's. It opens new doors for drug discovery and may potentially reduce the costs associated with experiments, but also presents important questions regarding the ways that these discoveries are presented to scientists for integration into human decision making.

Another family of advanced deep learning architectures is the **Generative Adversarial Network** (GAN). A GAN is a deep learning system whereby one model learns to generate content (images, text, etc.) while a second model simultaneously learns to detect content produced by the first model among real content (Goodfellow et al., 2014; Karras et al., 2020). The generators produced by training a GAN can be applied to a wide range of complex tasks (Gui et al., 2021). So far, this process has primarily been used in the image domain (e.g. photo inpainting (Yu et al., 2018), ageing/de-ageing filters (Antipov et al., 2017), cross-domain image transfer [e.g. generate paintings from photos and vice versa; Zhu et al., 2017], and creating super resolution images from low resolution images (Ledig et al., 2017), all aligned with the *sensory processes*, as viewed in our definition of cognition.

Despite the promise of GANs, their widespread use has already highlighted their potential to blur the line between real and fake images and videos. In 2019, Nvidia released StyleGAN, a generator capable of producing highly realistic images of human faces (https://thispersondoesnotexist.com/). More controversial is the proliferation of deepfakes, where people's faces can be digitally replaced with other people's and have unfortunately already been used for fake news, revenge pornography and social engineering scams.

3.1.2. Computer vision

Computer vision is one of the fields that benefited most from the recent developments in deep learning architectures and involves training machines to process and recognise images and video files. Representing primarily the sensory processes, this encompasses image classification ("This is an image of a Siamese cat"), object detection ("There are balloons at the following coordinates in this image"), pose detection ("This man is crouching") and more. Modern computer vision systems are almost universally built from convolutional neural networks (CNN; LeCun et al., 2015). These are artificial neural networks (loosely) inspired by the mammalian visual system whereby patches of pixels are "pooled" together when analysing an image (Fukushima, 1980), and have been the basis of most state-of-the-art computer vision since AlexNet won the Imagenet 2012 competition (the *de facto* benchmark for image classification) with a 9% improvement over the previous (non-CNN) winner (Krizhevsky et al., 2012).

Computer vision can be applied to tasks as varied as facial recognition (Wright et al., 2008), autonomous vehicle navigation (Bimbraw, 2015), optical character recognition (Islam, N., Islam, Z., & Noor, 2017) or even estimating poverty from satellite images (Jean et al., 2016). With improving models, data and hardware, computer vision applications will become more available and more widespread. A topical example is the adoption of facial recognition systems by law enforcement agencies around the world.

3.1.3. Natural language processing (NLP)

Natural language processing involves the use of machines to extract meaning from text and speech that would be understandable to a human and is not structured like code. NLP has a long history that predates artificial cognition and is primarily aligned with general cognitive operations. Much recent work in this field has embraced deep learning. It has been used, for example, for summarizing text (e.g., finding keywords or topics), named-entity recognition (e.g. in the phrase "Jack hit his head yesterday", we know that "Jack" is a person, "yesterday" is a date, etc.) and identifying emotions.

User-generated content (social media, blogs, Wikipedia, etc.) has been a key driver for recent work into artificial-cognition powered NLP. These huge quantities of unstructured text have allowed practitioners to train data-driven artificial-cognition models of language that are showing state-of-the-art performance in the GLUE and SuperGLUE challenges (two of the most widely cited benchmarks of NLP performance; Wang et al., 2018, 2019). These models can accurately extract meaning from language structures such as negation, nested clauses, metaphors, idioms, polysemous words, etc. That non-artificial cognition, hand-coded NLP systems often struggle to cope with.

In 2019, OpenAI's GPT-2 (Generative Pretrained Transformer 2 [Radford et al., 2019], since superseded by GPT-3 [Brown et al., 2020]) made headlines with a story about four-horned unicorns residing in South America. What made this story remarkable was that it was entirely machine-generated yet still syntactically and (mostly) semantically coherent. Around the same time as GPT-2, there was the publication of Google's BERT (Bidirectional Encoder Representations from Transformers; Devlin et al., 2018), a model with a similar so-called Transformer architecture (Vaswani et al., 2017). These models were initially trained with a conceptually simple objective to predict either the next word in a sentence (GPT-2) or the missing words from a masked sentence (BERT), and the rules the models learnt from these general tasks have turned out to transfer very well to other, more specific language tasks such as sentiment analysis and question-answering. These models have become the basis for most of the top-ranking models on the GLUE and SuperGLUE leaderboards, exhibiting superhuman performance (as of July 2021, the human baselines were ranked 16th and 4th on GLUE and SuperGLUE, respectively).

3.2. What has artificial cognition ever done for us?

Understanding the value of artificial cognition lies in answering the question 'what has artificial cognition ever done for us?' This question can be addressed through three key developmental steps (Van den Bosch & Bronkhorst, 2018). The first step is the process of humans assisting machines to develop, or train, the artificial-cognition capabilities. This first developmental step is *uni-directional* as it involves humans solely training the artificial agent. The second step is machines and other artificial agents assisting humans to facilitate data processing, analysis, and decision-making. This phase is considered *bi-directional* as the human can ask the machine to perform a task, and the machine can respond. The third phase is human and machine collaboration i.e., working as a team. During this *collaborative* phase, humans and machines are both aware of each other's states and outputs from each cognitive system can be passed to the other for ongoing processing.

Developing artificial-cognition capabilities involves training the AI to interpret and learn from data to perform a task accurately (our functional focus on cognition emphasizes determining most suitable human or machine task allocation). Once the artificial-cognitive system has been accurately trained it can assist humans and amplify human cognition across a growing number of fields (the second stage of artificial cognition). Scientific research across various disciplines has long used artificial cognition and machine learning capabilities. Algorithms can be applied to perform complex calculations, search through large repositories to find relevant data and generate models. Algorithms can detect new patterns in large datasets that would otherwise go unnoticed. For example, data-driven algorithms have been developed to discover new astronomical phenomena, such as finding new pulsars from existing data sets (Morello et al., 2014) and correctly classifying galaxies (Banerji et al., 2010). In the field of particle physics, CERN hosted a competition in 2018 to develop an AI system to quickly reconstruct particle trajectories of collision products based on the complex patterns left on the detector of the Large Hadron Collider (Amrouche et al., 2021). In the health sector, deep-learning algorithms have been developed to assist medical diagnoses by classifying images of skin lesions as either benign or as malignant melanomas (Haenssle et al., 2018). These models outperformed most dermatologists and missed fewer melanomas. Artificial-cognitive systems have been deployed in hospitals, such as the Trauma Reception and Resuscitation System (TR&R) currently in use at the Alfred Hospital in Melbourne.[1] This system[2] was designed to support highly trained medical staff to stabilise and treat major trauma patients when they arrive at the hospital. TR&R monitors patient data and provides recommendations displayed on a large screen visible to all medical staff. It acts as a decision support system and has helped reduce errors of omissions and the time spent by patients in the Intensive Care Unit. In education, artificial-cognitive systems are being used to personalise learning for individual students (Luckin & Cukurova, 2019). These systems aim to transform how students learn and assist educators by closely monitoring and assessing students' progress. One such system is Carnegie Learning's AI-based platform, which provides research-based curricula, textbooks and mathematics learning adaptive software that facilitate active learning for middle school students through university (Ritter et al., 2015).

Thirdly, human and artificial cognition can work collaboratively. Known as 'Human-in-the-loop' (HITL) hybrid-augmented intelligence is defined as an intelligent model that requires human interaction. In this type of system, the human is always a part of the process and consequently influences the outcome of the interaction. To be truly collaborative, humans and machines require shared cognition. That is, both the

[1] https://a2i2.deakin.edu.au/projects/trauma-reception-and-resuscitation-system/.

[2] This is strictly-speaking a rule-based decision support tool rather than a deep learning system.

human and the machine must have insight into and understand each other's knowledge. Additionally, both the human and the machine must understand the task. Current artificial-cognitive systems are solely uni-directional; however, some systems are now becoming bi-directional (e.g., artificial cognition that can detect bias and report these errors back to the human). Moving into this third phase of true HAC collaboration has not yet been realised however researchers are exploring opportunities in numerous fields including; defense, health and aviation.

4. Interactions between human and artificial cognition

As depicted in Fig. 2, humans and AI interact in 3 ways: 1) the construction of data sets (H), the building of the algorithms and AI agents (H1 & H2), and 3) during decision-making or performance-related tasks (e.g. Human-in-the-loop - HITL). Big data are used to feed algorithms that, in turn, are at the core of AI agents. There are four important aspects to note: i) big data revolves around human-related states, processes, and events, ii) such data is the substance of any algorithm, iii) algorithms are the drivers of AI agents, and iv) algorithmic/AI behaviours and outputs have effects on how new data are built and how humans relate to ADA-technologies in general. H1 and H2 are a subset of humans with highly specialised skills pertinent to algorithms, data, and AI (ADA).

Most (if not all) artificial-cognition technologies depend on large data sets with content related to human states, processes, and each represent sample events. As such, the first way that humans and artificial cognition interact is in the feeding of human-based information into a dataset. For example, although the accuracy of artificial cognition trained to detect skin cancer (Esteva et al., 2017) depends on the algorithm used (e.g., deep convolutional neural network), the algorithm itself relies quite heavily on the size of the data set used to train the algorithm (in the study by Esteva et al. the data set consisted of ~130,000 labelled clinical images. That is, the algorithm can output an educated guess that depends directly on the exemplars it has been exposed to (see Marcus, 2018, for a discussion on the limitation of deep learning).

Once the dataset is constructed, HAC interact in the algorithm development process. At this stage humans decide the way the algorithm is designed, the level of explainability and transparency, and the way the algorithm is trained (see Courtland, 2018, for a discussion). It is worth emphasizing that all algorithms depend directly on extant statistical and mathematical algorithms that are combined into machine learning. In turn, an algorithm's performance depends on the machine learning technique used and its fine tuning thereof (e.g., the performance of a deep learning algorithm is influenced by the number of layers and nodes; see Kliegr et al., 2020 for a review of ML applied to behavioral sciences).

The outputs of artificial-cognition technologies that use these data and algorithms, in turn, impact humans and the societal environment in which they live (see Appendix 3 in Whittlestone et al., 2019, for a succinct discussion on this matter). This situation has motivated discussions around the concept of 'human-in-the-loop' (HITL; see also section 4.2). Under the HITL premise, artificial-cognitive agents are checked by humans with specific skills (see Fig. 2 above) thus improving and regulating artificial-cognition technologies. A recent proposal consists of extending the loop so that a larger group of stakeholders are linked to the loop. The 'society-in-the-loop' premise consists of adding a social contract to HITL (Rahwan, 2018). Most recently, the impact of AI on society has been brought to the fore by accusations from a whistleblower at Meta, formerly Facebook. Frances Haugen, the whistleblower, argued that Zuckerberg and his algorithm-driven social network, have "unitaleral control over 3 billion people" and are using it to perpetuate harm. In this vein, the data and algorithms used to build artificial-cognition capabilities represent yet another way HAC interact.

The third way HAC interacts as represented in Fig. 2 is in what's known as human augmentation. Raisamo et al. (2019, p. 132) define human augmentation as "an interdisciplinary field that addresses methods, technologies and their applications for enhancing sensing, action and/or cognitive abilities of a human. This is achieved through sensing and actuation technologies, fusion and fission of information, and artificial intelligence (AI) methods." That is, the role of augmented cognition is to support human information processing related to sensory memory, working memory, executive functions, and attention. These authors highlight that augmented cognition provides access to an extended memory and virtually unlimited knowledge. Although some of the technologies required for augmented cognition are constantly undergoing development (e.g. VR headsets), the easy integration between human's cognitive capabilities (e.g., emotions and sensorimotor processes) and those technologies is also warranted (see Raisamo et al., 2019). The goal of human augmentation is to extend people's cognitive abilities to improve processes such as decision making and creative problem solving.

4.1. *The trouble with Human-AI interactions*

Although doomsday scenarios often focus on robot overlords growing sentience, perhaps more central to modern discussions of artificial cognition is the role *human* sentience plays in our interaction with artificial cognition, and the subsequent and cyclical impact on human cognition. Increasing attention is being paid to the role of bias in artificial cognition, with a specific focus on the impact of this bias on human decision making and performance. For instance, when looking at image data, bias and error is introduced when images are labelled by humans. Tommasi et al., 2017 highlight how images are labelled differently in different datasets (see Fig. 3), introducing confusion and lack of clarity.

Such bias in data is not limited to images only; it can be introduced from the moment data about humans are collected and recorded. The Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) risk-assessment algorithm in the U.S.A. exemplifies this point. COMPAS was designed to estimate the likelihood of criminals reoffending, and consequently inform judges sentencing decisions (e.g. bail or sentence). COMPAS was fed with a data set containing the results of a 137-item questionnaire (or 137 features in machine learning lingo) and was given to ~2000 individuals (Brennan et al., 2009). However,

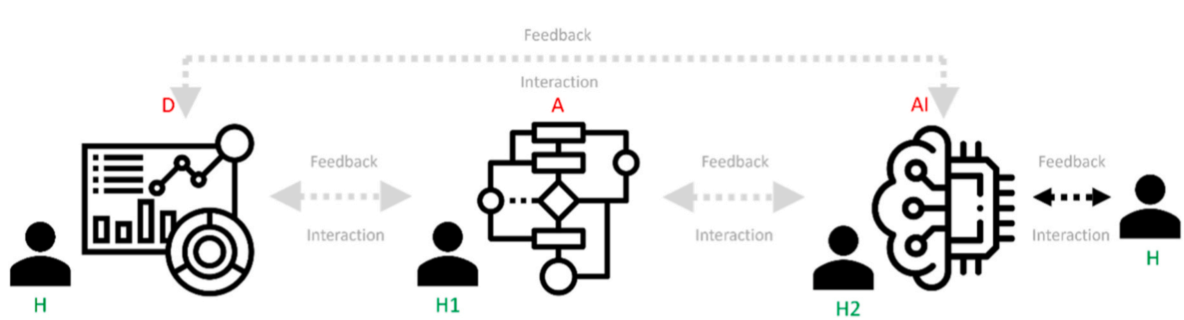

Fig. 2. Relationship among data (D), algorithms (A) and AI (AI) (Icons are from www.flaticon.com).

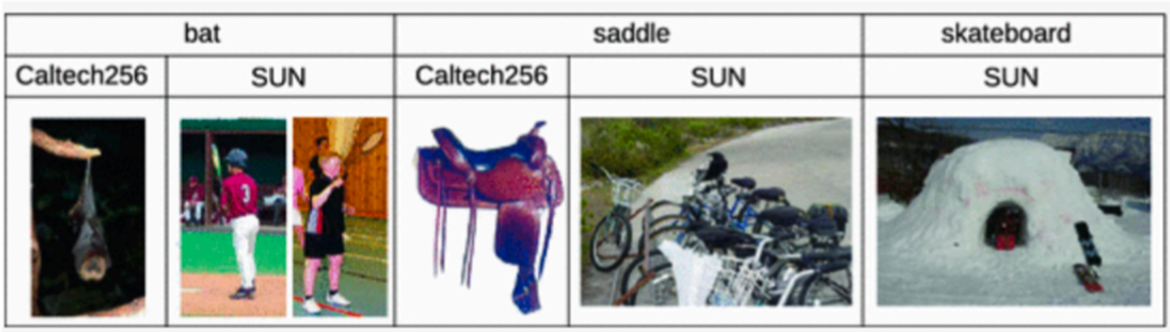

Fig. 3. Sets of images and their labels according to the datasets containing them. For example, in English language the concept 'bat' refers to a mammal but also it refers to an item. Note: Caltech256 and SUN are two large data sets of images (Source: Fig. 2.3 in Tommasi et al., 2017).

there were serious issues with COMPAS as the results suggest that the algorithm's predictive capability isn't accurate enough (as the same results can be obtained via a simple logistic regression or even human prediction; see Dressel & Farid, 2018). Additionally, the data itself were highly biased. Specifically, the data fed to the COMPAS algorithm came from heavily policed areas that also correlated with high representation of minorities and high poverty. In other words, the data were already contaminated by decades of biased police practices (Kirkpatrick, 2017). This bias in the COMPAS data has led to black people being twice as likely as white people to be scored with higher recidivism scores (interestingly, the same individuals showed a lower likelihood to reoffend as compared to white individuals).

Concerns of human cognition centre on the challenges of artificial cognition-enabled parsing of data and the creation of sub-communities on social media that shape and change the belief systems of individuals based on promotion of sensationalized stories and the propagation of misinformation. Early indications are that these social media spaces and algorithm-driven interactions cause negative impacts on human cognition and attention (Berryman et al., 2018). As such, concerns are increasing regarding the psychological harm that results from AI gaming human attention and emotion systems.

5. Human and artificial cognition in practice

The defining challenge for researchers, then, becomes one of developing models that provide guidance on when human cognition should be applied to a challenge, when artificial cognition should be used, and how coordinating between the two is handled. This is not a new challenge for researchers. In 1951, Fitts detailed a series of functions that were unique to humans and those that were better handled by machines. The 11-item list focused on broad activities (such as "exercise judgement" and "perform repetitive, routine tasks"). Today's list should include specific cognitive functions, instead of broad cognitive activities. Rather than focusing on the viability of artificial general intelligence, this work addresses the current reality where machines and humans are already jointly engaged in knowledge processes and then sharing the output with the other system. As such, models can be developed where the three dimensions of our definition of cognition (sensory processes, general operations, and complex integrated activities) can be undertaken by a human or artificial cognitive agent.

For example, the role of artificial cognition in creative efforts has become an increasingly popular topic of research and discussion (e.g., Cropley & Marrone, 2021). Although creative problem solving is a complex integrated task, examining more detailed processes and activities that make up the complex task reveals that some creative activities may be better suited to artificial cognition and others to human cognition (Medeiros et al., 2023). Using the three dimensions of cognition, we can better understand how to successfully integrate artificial and human cognition in creative work. Sensory processes involved with identifying the problem, the first process of creative problem solving (Mumford et al., 1991), may be completed by both human and machine. Humans may be best suited to detect the emotional component of a problem and to understand why a solution is needed. Artificial cognition, however, may be well-equipped to gather data related to current market trends which human cognition may then process and identify potential gaps (or problems). Regarding general operations, artificial cognition may be useful for sourcing large amounts of information about the problem at hand more effectively and quickly than humans. However, reviewing and summarizing this information is likely a task more suitable to human cognition. Human cognition is also best suited for the complex integrated tasks involved in creativity such as brainstorming and evaluating proposed solutions. Although artificial cognition may assist in these processes by providing prompts or recording meetings, these are largely sensory processes which facilitate the complex integrated activities being carried out by human cognition.

6. Ongoing research needs

Research in basic AI and human cognition continues at an unprecedented pace, accelerating our understanding of both the human and artificial cognitive systems. Our proposed definition of cognition, models of HAC overlap, as well as HITL represent promising and needed areas of ongoing research. As detailed above, the needed research at the intersection of these systems is twofold: how to functionally allocate different tasks and activities and how to integrate the outputs from each system during the completion of complex knowledge activities.

There are a series of stages that need to be considered and that each have different attributes. When a project or knowledge work is initiated in teams, someone generally decides regarding allocation of tasks. This is the *task allocation* stage. For example, someone could decide that the meeting needs to be recorded and transcribed rather than rely handwritten or typed notes. Increasingly, this stage can be automated behind the scenes and is outside of human awareness. During the COVID-19 pandemic, as work and learning moved online, large technology providers experienced an explosion in data around human collaboration. This resulted in automated summaries of tasks to do (integrating email, platforms such as Microsoft Teams, and other digital traces individuals left as they collaborated in teams). Then, after task allocation, the *knowledge processing* stage is initiated. This is where the problem is solved, and solutions are generated. During this stage, depending on the role of AC, there may be *augmentation* activities in real time that address the challenges being confronted, enlarges the discussion, provides insights to what's happening, or what might happen next. Throughout this process, and especially at the conclusion of knowledge work, there is an *integration* stage where the work of human and artificial cognition is integrated so decisions can be made, and actions undertaken. These are high level views of cognitive work and at each stage, research is required to assess the actual impact and outputs of integration. Additionally, the emotional or affective attributes experienced by team members when engaged with artificial agents, reveal another important area of ongoing research.

7. Conclusion & future research

Society is facing growing concerns including disinformation campaigns, responding to climate change, ongoing response to the COVID-19 pandemic, and complexity from globalization and emerging economies. In each instance, artificial cognitive processes will be utilized in identifying responses and finding solutions. Understanding how artificial cognition will contribute to solving problems will require more sophisticated models of integration and coordination. Progress in artificial cognition, human cognition, and neuroscience research will change coordination models, especially as the domain of what artificial-cognition tools can do evolves and increasingly encroaches on human capabilities.

As artificial cognition makes progress in domains that remain uniquely human – creativity, sensemaking, and emotional and affective states – existing views of artificial cognition as a support to human cognition will require upgrading to reflect artificial cognition as an active and equal partner. A better understanding is needed around human and artificial cognition (HAC) collaboration from a cognitive domain specific lens. For example, what are the underlying cognitive processes of creativity and sensemaking and how are these expressed when humans and machines work together?

Additionally, HAC will be expressed differently in different domains. Military applications will differ from the needs in education or healthcare settings. Similarly, social systems that involve collaboration, decision making, or teaching and learning will also utilize more nuanced collaboration models. For robotics completing a constrained task on a factory floor, collaboration is bounded. For an artificial-cognitive system collaborating in a changing crisis scene such as an emergency room, the points of human and machine integration will be less defined and more subject to contextual factors.

The future of knowledge work is one where humans and artificial agents collaborate. This collaboration is already occurring at discrete cognitive tasks that are rapidly integrated into human cognitive work. Routine activities such as composing an email or tracking trends on social media demonstrate the interplay of artificial cognition conducting tasks and their corresponding impact on how people connect with one another or engage with that content. Understanding how this integrated work is coordinated requires conceptual and theoretical frameworks that are validated through research. The various contexts in which shared work occurs – healthcare, business, education – will require unique and specific approaches as well. By shifting the exploration of how humans and machines interact to the cognitive - rather than intelligence – level, researchers and practitioners can begin to practically explore and implement ideal configurations for cognitive work, while waiting for the long rumoured artificial general intelligence to arrive.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Amrouche, S., Basara, L., Calafiura, P., Emeliyanov, D., Estrade, V., Farrell, S., … Vlimant, J. R. (2021). The tracking machine learning challenge: Throughput phase. *arXiv preprint arXiv:2105.01160*.

Angelov, P. P., & Gu, X. (2018). Toward anthropomorphic machine learning. *Computer, 51*(9), 18–27.

Antipov, G., Baccouche, M., & Dugelay, J. L. (2017). Face aging with conditional generative adversarial networks. In I. International (Ed.), *Conference on image processing (ICIP)* (pp. 2089–2093). IEEE.

Banerji, M., Lahav, O., Lintott, C. J., Abdalla, F. B., Schawinski, K., Bamford, S. P., & Szalay, A. (2010). Galaxy zoo: Reproducing galaxy morphologies via machine learning. *Monthly Notices of the Royal Astronomical Society, 406*(1), 342–353.

Barbey, A. K. (2018). Network neuroscience theory of human intelligence. *Trends in Cognitive Sciences, 22*(1), 8–20.

Barsalou, L. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577–660.

Barsalou, L. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617–645.

Bayne, T., Brainard, D., Bryne, R., Chittka, L., Clayton, N., Heyes, C., Mather, J., Ölveczky, B., Shadlen, M., Suddendorf, T., & Webb, B. (2019). What is cognition? *Current Biology, 29*, R608–R615.

Bengio, Y., Lecun, Y., & Hinton, G. (2021). Deep learning for AI. *Communications of the ACM, 64*(7), 58–65.

Bermúdez, J. L. (2014). *Cognitive Science: An introduction to the science of the mind.* Cambridge University Press.

Berryman, C., Ferguson, C. J., & Negy, C. (2018). Social media use and mental health among young adults. *Psychiatric Quarterly, 89*(2), 307–314.

Bimbraw, K. (2015). Autonomous cars: Past present and future a review of the developments in the last century the present scenario and the expected future of autonomous vehicle technology. In *12th international conference on Informatics in control, Automation and robotics (ICINCO)*. IEEE.

Bordot, F. (2022). Artificial intelligence, robots and unemployment: Evidence from OECD countries. *Journal of Innovation Economics Management, 37*(1), 117–138.

Brennan, T., Dieterich, W., & Ehret, B. (2009). Evaluating the predictive validity of the COMPAS risk and needs assessment system. *Criminal Justice and Behavior, 36*(1), 21–40.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., & Amodei, D. (2020). Language models are few-shot learners. *arXiv preprint*.

Brynjolfsson, E., Mitchell, T., & Rock, D. (2018). What can machines learn and what does it mean for occupations and the economy? *AEA Papers and Proceedings, 108*, 43–47.

Caron, M. S. (2019). The transformative effect of AI on the banking industry. *Banking & Finance Law Review, 34*(2), 169–214.

Constant, A., Clark, A., & Friston, K. J. (2021). Representation wars: Enacting an armistice through active inference. *Frontiers in Psychology, 11*, 598733.

Courtland, R. (2018). The bias detectives. *Nature, 558*, 357–360.

Cropley, D. H., & Marrone, R. L. (2021). *Automated scoring of figural creativity using a convolutional neural network*. PsyArXiv. https://psyarxiv.com/8qe7y/.

De Winter, J. C., & Dodou, D. (2014). Why the Fitts list has persisted throughout the history of function allocation. *Cognition, Technology & Work, 16*(1), 1–11.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). *Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv*.

Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances, 4*.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*(7639), 115–118.

Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., … Senior, A. W. (2018). De novo structure prediction with deeplearning based scoring. *Annual Review of Biochemistry, 77*(363–382), 6.

Fenwick, M., & Vermeulen, E. P. (2020). *Future lawyers: A roadmap for living & working with artificial intelligence*. Available at SSRN.

Fitts, P. M. (1951). *Human engineering for an effective air navigation and traffic control system*. Washington, DC: National Research Council.

Flusberg, S. J., Thibodeau, P. H., Sternberg, D. A., & Glick, J. J. (2010). A connectionist approach to embodied conceptual metaphor. *Frontiers in Psychology, 1*(197).

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics, 36*(4), 193–202.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences, 20*(1), 1–55. https://doi.org/10.1017/s0140525x97000010

Goldhahn, J., Rampton, V., & Spinas, G. A. (2018). Could artificial intelligence make doctors obsolete? *BMJ, 363*.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., & Bengio, Y. (2014). *Generative adversarial networks. arXiv preprint*.

Gui, J., Sun, Z., Wen, Y., Tao, D., & Ye, J. (2021). *A review on generative adversarial networks: Algorithms, theory, and applications*. IEEE Transactions on Knowledge and Data Engineering.

Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., … Zalaudek, I. (2018). Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. *Annals of Oncology, 29*(8), 1836–1842.

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

Holstege, G. (1992). The emotional motor system. *European Journal of Morphology, 30*(1), 67–79.

Islam, N., Islam, Z., & Noor, N. (2017). *A survey on optical character recognition system*.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science, 353*(6301), 790–794.

Johnson, N., Zhao, G., Hunsader, E., Qi, H., Johnson, N., Meng, J., & Tivnan, B. (2013). Abrupt rise of new machine ecology beyond human response time. *Scientific Reports, 3*, 2627. https://doi.org/10.1038/srep02627

Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., & Aila, T. (2020). Training generative adversarial networks with limited data. *Advances in Neural Information Processing Systems, 33*, 12104–12114.

Kirkpatrick, K. (2017). It's not the algorithm, it's the data. *Communications of the ACM, 60*(2), 21–23.

Kliegr, T., Bahnik, S., & Furnkranz, J. (2020). *Advances in machine learning for the behavioral sciences*. American Behavioral Scientist.

Kotseruba, L., & Tsotsos, J. (2020). 40 years of cognitive architectures: Core cognitive abilities and practical applications. *Artificial Intelligence Review, 53*, 17–94.
Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 1097–1105.
Lake, B. M., Ullman, T. D., Tenenbaum, J., & Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioural and Brain Sciences, 40*. https://doi.org/10.1017/S0140525X16001837
Larue, O., West, R., Rosenbloom, P., Dancy, C., Samsonovich, A., Petters, D., & Juvina, I. (2018). Emotion in the common model of cognition. *Procedia Computer Science, 145*, 740–746.
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444.
S. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., & Acosta, A. (2017). Photo-realistic single image super-resolution using a generative adversarial network *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4681–4690
Lieto, A. (2021). *Cognitive design for artificial minds*. Routledge.
Luckin, R., & Cukurova, M. (2019). Designing educational technologies in the age of AI: A learning sciences-driven approach. *British Journal of Educational Technology*.
Marcus, G. (2018). *Deep learning: A critical appraisal*. https://arxiv.org/abs/1801.00631.
Marcus, G., & Davis, E. (2019). Rebooting AI: Building artificial intelligence we can trust. *Vintage*.
Marmolejo-Ramos, F., Khatin-Zadeh, O., Yazdani-Fazlabadi, B., Tirado, C., & Sagi, E. (2017). Embodied concept mapping: Blending structure-mapping and embodiment theories. *Pragmatics and Cognition, 24*(2), 164–185.
Marmolejo-Ramos, F., Murata, A., Sasaki, K., Yamada, Y., Ikeda, A., Hinojosa, J. A., Watanabe, K., Parzuchowski, M., Tirado, C., & Ospina, R. (2020). Your face and moves seem happier when I smile. Facial action influences the perception of emotional faces and biological motion stimuli. *Experimental Psychology, 67*(1), 14–22.
McClelland, J. L. (2013). Integrating probabilistic models of perception and interactive neural networks: A historical and tutorial review. *Frontiers in Psychology, 4*(503).
Medeiros, K. E., Marrone, R. L., Joksimovic, S., Cropley, D., & Siemens, G. (2023). Artificial and human creativity. Handbook of organizational creativity (R. Reiter-Palmon & S. T. Hunter, Eds.). Elsevier..
MIT. (2022) Retrieved November from: https://www.media.mit.edu/groups/personal-robots/overview/.
Morello, V., Barr, E. D., Bailes, M., Flynn, C. M., Keane, E. F., & van Straten, W. (2014). SPINN: A straightforward machine learning solution to the pulsar candidate selection problem. *Monthly Notices of the Royal Astronomical Society, 443*(2), 1651–1662.
Mumford, M. D., Mobley, M., Reiter-Palmon, R., Uhlman, C., & Doares, L. (1991). Process analytic models of creative capacities. *Creativity Research Journal, 4*(2), 91–122. https://doi.org/10.1080/10400419109534380
Pessoa, L. (2017). Do intelligent robots need emotion? *Trends in Cognitive Sciences, 21*(11), 817–819.
Pessoa, L., Medina, L., & Desfilis, E. (2022). Refocusing neuroscience: Moving away from mental categories and towards complex behaviours. *Philosophical Transactions of the Royal Society B, 377*(1844), Article 20200534.
Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Blog, 1*, 8.
Rahwan, I. (2018). Society-in-the-loop: Programming the algorithmic social contract. *Ethics and Information Technology, 20*(5–14).
Raisamo, R., Rakkolainen, I., Majaranta, P., Salminen, K., Rantala, J., & Farooq, A. (2019). Human augmentation: Past, present and future. *International Journal of Human-Computer Studies, 131*(131–143).
Ritter, S., Barrett, D. G., Santoro, A., & Botvinick, M. M. (2017). Cognitive psychology for deep neural networks: A shape bias case study. In *International conference on machine learning* (pp. 2940–2949).
Ritter, S., Carlson, R., Sandbothe, M., & Fancsali, S. E. (2015). *Carnegie learning's adaptive learning products*. 8th International Conference on Educational Data Mining.
Roth, E. M., Sushereba, C., Militello, L. G., Diiulio, J., & Ernst, K. (2019). Function allocation considerations in the era of human autonomy teaming. *Journal of Cognitive Engineering and Decision Making, 13*(4), 199–220.
Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., & Dieleman, S. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature, 529*(7587), 484.
Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., & Lillicrap, T. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science, 362*(6419), 1140–1144.
Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., … Teller, A. (2016). *Artificial intelligence and life in 2030: The one hundred year study on artificial intelligence*.
Taylor, J. E. T., & Taylor, G. W. (2021). Artificial cognition: How experimental psychology can help generate explainable artificial intelligence. *Psychonomic Bulletin & Review, 28*(2), 454–475.
Tommasi, T., Patricia, N., Caputo, B., & Tuytelaars, T. (2017). A deeper look at dataset bias. In G. Csurka (Ed.), *Domain adaptation in computer vision applications. Advances in computer vision and pattern recognition*. Springer.
TUC - Trade Union Congress. (2018). *A future that works for for working people*. Congress House. Retrieved November from https://www.tuc.org.uk/sites/default/files/FutureofWorkReport1.pdf.
Van den Bosch, K., & Bronkhorst, A. (2018). *Human-AI cooperation to benefit military decision making*. NATO.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 5998–6008.
Wang, A., Pruksachatkun, Y., Nangia, N., Singh, A., Michael, J., Hill, F., & Bowman, S. (2019). *Superglue: A stickier benchmark for general-purpose language understanding systems*.
Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. R. (2018). Glue: A multi-task benchmark and analysis platform for natural language understanding. *arXiv*.
Weick, K. E., & Roberts, K. H. (1993). Collective mind in organizations: Heedful interrelating on flight decks. *Administrative Science Quarterly*, 357–381.
Whittlestone, J., Nyrup, R., Alexandrova, A., Dihal, K., & Cave, S. (2019). *Ethical and societal implications of algorithms, data, and artificial intelligence: A roadmap for research*. Nuffield Foundation.
Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., & Ma, Y. (2008). Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 31*(2), 210–227.
Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., & Huang, T. S. (2018). Generative image inpainting with contextual attention. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5505–5514.
Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proceedings of the IEEE International Conference on Computer Vision*, 2223–2232.
Ziemke, T., & Lowe, R. (2009). On the role of emotion in embodied cognitive architectures: From organisms to robots. *Cognitive Computation, 1*, 104–117.
Zmigrod, S., & Hommel, B. (2013). Feature integration across multimodal perception and action: A review. *Multisensory Research, 26*(1–2), 143–157.