JOURNAL OF MODERN SCIENCE

4/64/2025

www.jomswsge.com

DOI: 10.13166/jms/216052

Franciszek Kutrzeba
Gdańsk University of Technology, Poland
ORCID iD: 0000-0003-3641-3299

COMPUTATIONALISM, ARTIFICIAL INTELLIGENCE, AND EMOTIONS

Abstract

The article critiques the reliance on computationalism and functionalism in strong artificial intelligence (AI), emphasizing the inadequacies of comparing the human brain to conventional computers. It discusses computationalism's view that the brain functions similarly to a computer, proposing that mental states can exist independently of a biological substrate. This notion, while appealing, is challenged by evidence of the brain's complex and dynamic nature, including its distributed information storage and the intricate processes governing neuronal activity. The role of emotions in cognition is explored, challenging the dichotomy between reason and emotion, and showcasing how emotions are integral to decision-making and problem-solving. The development of emotional artificial intelligence (EAI) is highlighted, demonstrating its potential applications and associated ethical considerations. Additionally, the article delves into the ongoing debate concerning consciousness, advocating for a nuanced understanding of the mind-body relationship and emphasizing the interconnectedness of neural processes and embodied experiences. Ultimately, the article asserts that simplistic analogies between brains and computers fail to capture the complexity of human cognition and the embodied nature of consciousness, calling for a re-evaluation of AI's theoretical foundations.

Keywords: *emotions, cognition, morality, neural networks, artificial intelligence, computationalism*

1. Introduction

Researchers advocating for strong artificial intelligence (AI) have often relied on a theory called computationalism – the computational theory of mind (CTM). The theory proposes that the brain is functionally similar to a computer. Accordingly, the mind is an interface for the brain as an operation system for the computer's hardware. Both computers and the brain store information in memory and process it systematically to solve problems (Arkoudas and Bringsjord, 2014).

The logical extension of computationalism is a functionalist way of thinking, according to which the mind is not a biological phenomenon found in nature and thus does not need a biological structure to materialize. Mental states are not determined by a biological structure but by the functional role that they play in the system as a whole. According to functionalism, there is nothing to prevent the software of the mind from being run on a machine made of electronic components or even transferring the human mind to such a machine. Indeed, the majority of functionalists believe that it is only a matter of time before computer hardware and programs corresponding to the human brain and mind are designed (Searle, 2012; Churchland, 2004; Arkoudas and Bringsjord, 2014).

Comparing the brain to a computer may seem tempting, but the analogy may fall short. According to philosopher and neuroscientist Churchland (2004), the division into hardware and software is not appropriate to describe the diverse divisions of the human nervous system. Nor does the modular structure of the computer, where one part processes and the other stores, correspond to the structure of the brain or nervous system, although there are areas of functional specialization in the brain. In the brain, information is not stored in a memory organ but appears to be distributed across multiple neurons, and structures that process information can also transform to store information.

Neurons in the brain are significantly more complex and dynamic than computer transistors or neurons in artificial neural networks. The life cycle of neurons includes growth, development, atrophy, and death. New neurons are also produced in certain parts of the brain. They create and strengthen or weaken and break connections with other neurons. Changes occur in the neuronal importing branches, i.e., dendrites, receptors, ion channels, and cell membrane, and the likelihood of neurotransmitter secretion may change (Boden, 2016).

In terms of computational complexity, a single neuron can correspond to even a small computer. Scientists have managed to classify neurons into a few different types, although their function is today generally defined as transmitting and/or receiving signals, or inhibiting the transmission. A recent research by Kinman et alia (2025) suggests, that highly-specialized neurons – the so called 'ovoid cells' – play a key-role in object recognition memory. They activate each time we encounter something new, triggering a process that stores those objects in memory and allowing the human mind to recognize them months, or even years later. Ylirönni (2024) recalls that about half of the brain cells are different types of glial cells, which have their own, as yet poorly understood, role in information processing. Nerve cells are living beings that grow, develop, atrophy, and die. The countless connections between them are also in a constant state of change. Each nerve cell is surrounded by thousands of synapses, which contain receptors sensitive to different substances, and the number of different neurotransmitter-receptor combinations is practically infinite. Inter-neuronal communication is based on electrical or chemical signals that travel from the sending neuron to the receiving neuron. Chemicals can also spread regionally, regardless of whether there are connections between neurons or not. In addition to the rate of triggering, the nervous system probably uses several other communication strategies, such as the timing of triggering relative to other neuronal events. The majority of brain cells are not neurons, but glial cells that, among many other functions, are also involved in information transfer. Dendritic activity is also thought to be related to it (Campbell et alia 2015; Boden 2016; Churchland 2004). Thus, either the hypothesis of the distinct function of neurons is untrue or it supports the long used axiom that a system is more than a sum of its entities. Those who compare the brain to a computer describe the function of nerve cells as *calculating*, but nerve cells do not calculate or manipulate symbols based on logical rules, instead they function to maintain their own metabolism (Ylirönni, 2024).

Haikonen (2017) recalls, that despite all the differences, proponents of strong artificial intelligence believe that a computer with the right software is powerful enough to match the mind and brain of a human. Allen Newell and Herbert Simon, who were among the pioneers of artificial intelligence, have put forward the hypothesis of a physical symbol system close to computationalism, according to which symbols are at the core of intelligent behavior. Both the brain and the computer, as physical symbol systems, have all the characteristics necessary for general intelligence. The symbolic behavior of humans is due to the fact that we have the characteristics of a physical symbol system (Newell and Simon, 1976; Haikonen, 2017).

Computationalism, the theory of the physical symbol system, and functionalism together form the theoretical framework of classical or symbolic artificial intelligence, which alone dominated artificial intelligence research for 30 years and remains a strong influencer (Arkoudas and Bringsjord, 2014). According to computationalism, complex ideas are constructed in symbolic structures just as in natural languages; complex sentences are constructed from simpler parts. The theory presupposes that there are basic blocks, primitives, on which the 'language of thought' is built. The thinking itself is, in my view, the manipulation of symbols and images (from the Latin *imaginatio*, meaning *mental image*). However, critics say that manipulating symbols cannot explain intuition, judgment, sensitivity, taste, or imagination, all of which play a vital role in reasoning and problem-solving. In addition, the symbol grounding problem raises a question about the time and place of sense-making; where inside the brain the primitives gain their meaning, and how they can refer to things outside the brain.

1.1. About symbols

Symbols not only gain their meaning in relation to other symbols or the internal spaces of the brain, but they also gain it from the external world (Arkoudas and Bringsjord, 2014). According to artificial intelligence (AI) researcher Pentti Haikonen (2017), the ultimate meaning of symbols comes from sub-symbols, which are direct sensory perceptions, such as visual or auditory perception. While learning the basic vocabulary of a language, a small child makes observations about the environment and associates words with them. A detected thing, such as a spoon, sends a certain signal pattern to the brain, which is a sub-symbol. Experiences of pleasure and pain are also sub-symbolic, as they do not require interpretation but are based directly on the senses. The meaning of words and other symbols stems from the association of sub-symbols. Things that come together associate together in the brain, regardless of there being any real connection. The principle of operation makes it possible to associate words of a language with objects of the outside world, which have no connection with the sound pattern of the word. This creates a basis for natural language and symbolic thinking (Haikonen, 2017).

A digital computer is not a system that uses external symbols to detect external reality, although it can be connected to a camera and a microphone. The information transmitted by the camera and microphone must be converted to digital format using binary numbers – ones and zeros for the computer (Haikonen, 2017). The computer operates with ones and zeros that have no meaning or semantic content. The ones and zeros are just numbers that do not even mean numbers to the computer. However, they represent logical values (true and false). According to philosopher John Searle, a digital computer cannot have an opinion because the program has only syntax but no semantics. The mind has relevant content, unlike symbols handled by a computer (Searle, 2012). The other approach is that the semantics behind the symbols is a defined meaning, materialized through any kind of language, whether programming (formal) conveyed through given formalized channel or natural language, which is the domain of organic forms of life. For Kutrzeba (2022), any living organism is a neurological unit that acquires knowledge about its surrounding and makes use of it. How could there be evolution if animals weren't able to learn? For people, learning happens through a structured and semi-structured language and by developing dialectical synthesis of experience and reflection – an epistemological superposition that Kutrzeba regards as *a superiori* mode of knowledge creation.

Organisms also use less tangible modes of communication with such catalysts and transistors as pheromones and hormones. These chemical cues are crucial for communication among various species. The scents convey information about territory, mating readiness, danger, and even alarm signals. Ants, for example, leave pheromone trails to guide others to food sources. Kutrzeba (2022) argues that all living things communicate and reflect along with the experience of sensing. Furthermore, as far as human beings are considered, the communication and comprehension can be reduced to three major epistemic domains: the subjective [arts], the intersubjective [technologies] and the objective [science]. The fourth would be an inter-objective domain; both when shared mental models exist in a community, and among computers that use the same operation systems and languages. One of the most intriguing questions to me is the problem of demarcating the transcendence between the living and dead matter, and therefore the various intensities of consciousness, and the different modes of communication with the tangible reality between organisms, organs, cells, fungi, bacteria, and viruses.

Searle (2012) claims that thinking is more than manipulating irrelevant symbols, and that is why a digital computer cannot think, regardless of its computing power. The computer is only capable of *simulating* thinking, awareness, and emotion. Simulation is not a real thing, although people tend to ascribe human features to machines. This illusion is referred to as the Eliza influence – a reference to a chatbot developed in the 1960s. Eliza manipulated sentences in a way that made many of its interlocutors think it understood them. It was still merely a matter of processing the text (Hofstadter, 2018). However, if simulation is not real then neither are the arts (film, dance, theatre, music, literature, especially fiction) should be regarded as unreal. Following the logic of Kutrzeba (2022), anything other than 'hard' science is true only on *subjective* (arts) or on *inter-subjective* domains (technology) as science strives for objectivism and universalism; knowledge stemming from natural sciences is true for everybody within a particular paradigm, and, knowledge or wisdom stemming from artistic or technological endeavours works or makes sense only in particular organizational setting. Hence what counts is the cognitive or computing agent's perception of the reality, not the scope of universalism. Thinking and thus perception are according to Kutrzeba (ibidem) a result of action where genetic knowledge (DNA) and enculturation (language and values; socialization) creates consciousness, and knowledge on how to cope with the ongoing experience, as best as the particular individual (experiencing agent) is capable of at a given time.

2. Neural networks

Symbolic artificial intelligence is not the only option in the world of artificial intelligence research. Criticism of the theory paved the way for a new approach, i.e. connectionism, which garnered strong support in the 1980s. Connectionists use artificial neural networks that differ in many ways from traditional artificial intelligence systems. The neural network consists of nodes resembling neurons in the brain with connections between them. The connections have weights that determine the effect of a node on another node. In larger neural networks, i.e. multilayer networks, the units are connected in series in layers. Unlike traditional systems, a neural network does not have a centralized processor but consists of interconnected units. In the system, memory and processing are combined (Arkoudas and Bringsjord, 2014; Haikonen 2017).

Most neural networks are capable of learning. Usually, learning only affects the weights of the connections, but in some systems, learning can also lead to the creation of new connections and the disconnection of old ones. The dynamism of neural networks and their ability to self-organize are features that are not present in symbolic artificial intelligence systems (Boden, 2016). It is also exceptional compared to the old one that the network can continue to operate even if parts of it are damaged. In a neural network, information is not encoded by symbolic structures but is presented as pattern spread across the entire network based on node triggers (Arkoudas and Bringsjord, 2014).

One known method of learning the weights of connections is the back-propagation algorithm used in multilayer networks. Weight adjustment values are usually calculated by a computer, as there is usually no other mechanism in networks to determine them. Due to countercurrent and similar methods, the function of the traditional artificial neural network bears little resemblance to the function of the brain but is essentially only a statistical calculation (Haikonen, 2017). There are other differences; synaptic connections in the brain are unidirectional rather than bidirectional. Brain networks are not organized strictly hierarchically. Artificial neural networks prioritize too much mathematical elegance and power over the brain. On the other hand, artificial neurons are too simple and present in negligible numbers even in the largest networks compared to the brain. Researchers tend to ignore factors associated with brain *wetness,* such as dendritic branch activity, tissue-propagating neuromodulators, and ion transport (Boden, 2016).

The calculations performed by the brain are semantic in nature and life-oriented, unlike that of a machine (Churchland, 2004). Human thinking is not the execution of program instructions but the processing of meanings and moral values – meanings also cause emotions that play an essential and necessary role in cognition (Haikonen, 2017).

3. The role of emotions in cognition

Reason and emotion in Western thinking are juxtaposed as opposites just like day and night, and this division is firmly rooted in culture. The reasoning of the machine is not affected by emotions, or, the problematics of comprehension the reality, what we are experiencing. So the question if artificial intelligence is pure intelligence is still there to be answered. Could a sufficiently powerful machine, due to its insensitivity, exceed human abilities in decision-making and thinking?

It is a common misbelief that intelligence does not require emotions; feelings interfere and interact with problem-solving, decision-making and rationality. Only psychopaths may have deficiencies in this interaction. Pessoa et alia (2019) claim, that *in different vertebrates, we identify shared large-scale connectional systems involving the midbrain, hypothalamus, thalamus, basal ganglia, and amygdala. The high degree of crosstalk and association between these systems at different levels supports the notion that cognition, emotion, and motivation cannot be separated – all of them involve a high degree of signal integration* (Pessoa et al., 2019, p. 1). Reason and emotion are thus intermingled, and a normally functioning person cannot have one without the other. Emotions are rational, or at least appropriate, otherwise, they would never have developed by organisms.

An evolving field focused on enabling AI systems to interpret human emotions through biometric data such as facial expressions, speech, and behavior has been vibrantly researched in the last decades (Assunção et al., 2022; Tretter, 2024). Emotional Artificial Intelligence (EAI) has applications in healthcare (enhancing patient interactions), automotive safety (detecting drowsiness), education (adapting teaching methods), and social robotics (improving human-AI engagement). In computer science, emotions are researched and developed into what is known as Automated human emotion recognition (AHER).

State-of-the-art technology uses brainwave signals to recognize gamers' emotions while playing a game to foresee their affective states. However, these are intrusive methods, *which distract the person from doing normal activities* (Younis et alia, 2024, p.). The question of whether emotional states can be quantitatively measured is still a topic of controversy. *There is a dispute over* *the precise definition of emotion which has raged within psychology since early ideas attempted to provide a concise response to the question*: what emotions are?. Neuroscientist Antonio Damasio has shown that a lack of emotion can just as well lead to irrational behavior. Damasio has studied patients with peculiar brain injuries that suppress their emotions but do not affect their other cognitive abilities. In all studied patients, the lack of emotion had led to, among other things, a decline in decision-making ability (Damasio, 1994).

According to Damasio (ibidem), reason, and emotion are connected in certain parts of the brain. The system of rationality in the brain is not built based on a biological regulatory system, but in conjunction with and stemming from it. Between emotions, a person experiences background feelings from the underlying body states that make living and being feel like something extraordinary in itself; life is a value inherent to anything else, even in the materialistic perspective. Without emotions, a person's representation of themselves does not function normally. Some empirical studies even indicate that the whole reasoning is an inherently emotional process (Damasio, 1994).

Emotions are primary because when bound to the body, they come first in the history of evolution and create a frame of reference for what comes after them (Damasio, 1994). As pinpointed by Goleman (1995) and previously by Maclean in the 1960s (Butler, 2009), binary thinking could stem from the functionality of evolutionarily seen the oldest part of the human brain, the so called reptilian brain (It is composed of the brainstem (medulla, pons, cerebellum, midbrain, globus pallidus, and olfactory bulbs) – the structures that dominate in the brains of snakes and lizards [Ibidem]), and would hence approximate the way the machines of today compute. Interestingly, computing evolved greatly from the binary thinking [computing] due to the introduction of fuzzy logics by Lofti Zadeh (1965). The nature of fuzziness of most of the life processes provokes us to acquire the concept of a superiori reasoning (Kutrzeba, 2022). This is due especially within such social sciences as management and economics, psychology, sociology, and the political science, but also within the discipline of technologies. From the epistemological perspective, the fuzziness of a given scientific or technological faculty indicates an inter-objective approach to resolving every day problems, meaning practically, that a successful life can be achieved by using both a priori and a posteriori methods of thinking – reasoning and ideation.

According to the traditional view, decision-making leads to the best results when emotions are not allowed to influence the process. Thus, rational consideration carefully weighs all possible options and performs a cold, cool cost-benefit analysis based on pure reason. However, this is not necessarily a truly rational action, as shown by the example of the behavior of Damasio's calm and collected patients. The following example illustrates this well; one of the patients was offered two dates for the next appointment, from which he was allowed to choose the one that suited him best. The patient took out his calendar and began to go through the pros and cons of both dates, taking into account all possible options and scenarios, starting with possible weather conditions. After a good half hour, Damasio interrupted his rational calculations and suggested the date himself (Damasio, 1994).

Emotions improve reasoning by automatically reducing the number of options. When a person considers a bad scenario or situation, they experience an unpleasant sensation that acts as a warning signal. Due to this signal, the person can immediately reject the ones causing unpleasant sensations and thus choose from a smaller set of options. Positive knowledge or impression, in turn, acts as an incentive (Churchland, 2004). According to Churchland (2017, p. 5), learning ethical concepts is also considered to require bonding, which arises thanks to emotions and trust. *The neurobiology of attachment and bonding provides a motivational and emotional substructure that allows the scaffolding of social practices, moral*. Functionalists believe that the right hardware is enough to generate emotion. From another perspective, certain biochemical processes, such as hormone secretion or changes in brain neurotransmitters, may be necessary for the onset of emotions. In this case, artificial actors would not have the opportunity to experience them. If biology is not necessary, similar results can be achieved, for example, with simulated hormone systems (Scheutz, 2014).

AI has become a terrific interpreter of human emotions through facial expressions and gestures. Moreover, *Emotion generation and expression mechanisms have been developed to enable AI systems to respond empathetically* (Narimisaei et alia, 2024, p. 4658). Tretter (2024) pinpoints that there are several ethical risks in the deployment of AI recognition of human emotions: cultural, gender, and racial discrepancies; diminished human agency as users may over-rely on AI recommendations; increased complexity in determining responsibility and liability when AI influences decisions, and potential manipulation through AI-driven nudging.

Intelligence is generally considered to be a higher level of cognitive awareness than emotion. However, according to biologist Helena Telkänranta (2015), the opposite can logically be thought of, as the definition of emotion is only met by conscious experiences, while the intellect is not necessarily more conscious than in a calculator or a smartphone. But what is consciousness itself? We ponder, maybe it is a mere epistemic interface?

3.1. The problem of consciousness

According to the prevailing materialist view, the brain produces consciousness. Various substances that affect brain function alter perception and behavior, and brain injuries can even lead to a permanent personality change. It appears that such facts justify the materialistic position, but strictly speaking, the evidence is not conclusive. For example, one might think that the brain is a receiver of consciousness in the same way that a television set is a receiver of transmission. In this paper, however, we do not go deeper into alternative theories and speculations, but assume the general materialist view as a starting point.

Today, relatively much is already known about, for example, the effects of different chemicals (produced by the body and those that come from outside) on different emotions. However, the mere knowledge that a particular chemical causes a particular feeling does not reveal how exactly that feeling is created. We can locate where a particular substance works, but we still cannot explain why it makes a person feel particular emotions. How and why do certain processes in the brain bring about conscious qualitative experiences, among other things? How do neuronal signals turn into experiences? The philosopher David Chalmers has defined this question as a difficult problem of consciousness (Damasio, 1994; Haikonen, 2017).

If we did not have our own experience of consciousness, we would not, when looking at the brain, assume that neural processes give rise to consciousness. When viewed based on mere physical facts, the hypothesis would seem unnecessary, even mystical (Chalmers, 1996). Would an accurate study of brain structure and function solve the problem of consciousness? All sensory observations are qualitative experiences, i.e., qualities in the language use of philosophers, and it seems impossible to describe them. It is assumed that with the development of brain imaging techniques, researchers will be able to see conscious experiences on the screen of a scanning device. However, they still do not see them, for they cannot be seen and must be experienced instead. Measuring instruments do not enter the phenomenon itself but merely produce a limited description of it, expressed in symbols or numbers (Hoffmeyer, 2014; Haikonen, 2017). Science cannot describe conscious experiences in the language it uses, which, in principle, only works within the bounds of the third person. Mixing first – and third-person experiences is commonly found in scientific thinking. Merely explaining cognitive functions does not explain the contents of consciousness or solve the problem of consciousness (Hoffmeyer, 2014; Haikonen, 2017).

So what am *I* experiencing? Within the brain, there is no unifying little man or even a functional center that would be *me*. Of course we have an idea of the self, which we typically refer to as the Ego. As Damasio (1994) proposed, the experience of a unified mind possibly arises from the simultaneous neural activity of different regions. However, there is no clarity as to how this happens and what the sufficient conditions for the emergence of consciousness are.

Interestingly, the activity of different areas of the brain is triggered in response to an external stimulus even when a person is not in a conscious state. For example, the so-called *facial area* of the human brain in a permanent vegetative state responds to images of a familiar human in the same way as the brain of a normal human. From this, it can be concluded that neurons in the field of view do not even produce a conscious visual experience (Churchland, 2004). When it comes to consciousness, many scholars suggest that conscious experiences form only the tip of the iceberg in the mind. The majority of the processes of the mind take place in the unconscious.

The experience of the self is inevitably also an experience of the corporeality, and many thinkers say that the mind cannot be thought of without some kind of embodiment (Hoffmeyer, 2014; Damasio, 1994). In the evolution, the body comes before the brain, and without the body, there would be no brain.

The brain and body are often distinguished by their structure and function, but in reality, they react to the environment as a whole and are in constant complex interaction with each other. Many of the neural networks are shaped by the demands of the body, and the prerequisites for a normal mind are the representations of the body they produce. We would hardly have the same opinion today without anchoring in the body (Damasio, 1994).

In general, we consider the central functions of the self to be remembrance, conscious reflection, and similar high-level mental functions, but they are probably extensions and transformations of the initial self-based on the autonomic nervous system and sensorimotor system. According to Churchland, they are at the core of the self with the somatosensory nervous system. The autonomic nervous system directs behavior through the regulation of vital functions and colors experiences with emotion. It plays a crucial role in making an organism a coherent biological entity (Churchland, 2004).

Arkoudas and Bringsjord (2014) recall that the important role of corporality was also raised by artificial intelligence circles already in the 1980s. The disappointment with classical artificial intelligence models encouraged some researchers to move from symbolic tasks to tasks related to sensory perception and motor skills. They believed that only the realization of full corporeality could bind the artificial intelligence system to the world and make it a real actor.

3.2. Machine awareness

Artificial intelligence systems know how to play chess or perform other individual tasks better than humans, but they are not truly intelligent, and do not truly understand what they are doing, or do not even know they exist. Qualitative changes are necessary if true universal intelligence or consciousness is to be achieved. Simply increasing the amount of processing and computing power will not be enough to achieve it (Haikonen, 2017; Boden, 2016).

Particularly challenging areas of artificial intelligence include language, creativity, and emotion Boden (2016). People's understanding of the world and other people is fairly based on tacit knowledge (*the moral spine*) that classical symbolic artificial intelligence cannot achieve as a rule-based system. A set of rules may well describe a cognitive phenomenon and make correct predictions about it, but this still does not mean that these rules are coded in the human mind. However, the biggest technical challenge for both weak and strong artificial intelligence is the problem of relevance. The machine lacks a sense of relevance, i.e. the ability to distinguish the essential from the irrelevant. This is the main reason why the machine does not know how to operate according to common sense (Arkoudas and Bringsjord, 2014; Boden, 2016; Churchland, 2004).

In humans and other animal species, a sense of relevance comes naturally and is tied to survival. Representatives of every species are attuned to perceive in their environment only that what is relevant to them, i.e., enemies, friends, mating partners, food, and dangers. The more rudimentary a species is, the fewer number of carriers of importance in its environment, and the more confident it is to move around the world (Uexküll, 2012). Of course, with the evolution of culture, there is more to human behavior, but these form the basis on which living beings navigate the world. For the machine, however, none of these things exist. It is not attuned to the world around it like living beings, and no part of the information that has come from the world is more relevant to it.

Researchers have addressed the issues raised by developing genetic algorithms that work according to an evolutionary model and by designing neural networks that combine digital and analog features. Some scientists have even connected genuine neurons to their networks (Boden, 2016). Haikonen (2017) argues that the associative neural networks he developed without control algorithms solve the problem of symbol grounding and have the potential to achieve consciousness. Is it possible to develop an artificial mind before knowing how the human mind works? Does copying brain mechanisms lead to an increase in consciousness? So far, we have experiential knowledge concerning only a mind based on biological life, and conscious artificial intelligence will not be developed, at least not in the near future, if ever. According to the wildest threats, superior artificial intelligence will subjugate people in the future. However, a more current and real danger than rebellious robots is that humans rely too much on artificial intelligence and give it decision-making power over various matters.

We do not know where the boundary of consciousness lies in nature, so how could we know when a machine is conscious? Alan Turing, the father of artificial intelligence, replies, *the only way to be sure that a machine is thinking* *is to be that machine itself and feel like you are thinking.* (Sheutz, 2014, p. 249). Artificial intelligence researcher Marvin Minsky commented on feelings in 1983 as follows: *I believe that we will be able to program the machine feelings after we have been able to program the thoughts … I am sure that once we have decided upon what feelings we desire for machines, execution will not be difficult.* In contrast to Kutrzeba (2022), who believes that creativity, intuition, cognition, and ideation are fundamentally stochastic and unique in nature, according to Minsky, there is nothing significant or miraculous about emotions, as there is nothing in intuition, creativity, or other human characteristics. Emotions are like an additional feature that can be programmed into a machine if you want to, but why on earth would anyone want emotional devices instead of smart devices? As humans become a luxury good, this might well be the future, and not the least for the elderly, but a remedy for our ageing societies.

4. Emotions, reasoning and the morality

Ylirönni (2024) mentions that emotions, when uncontrolled, can interfere with rational thinking. However, the overflow of emotions represents only the tip of the iceberg of what emotions are as a whole. Neuroscientist Antonio Damasio (1994) has shown that a lack of emotion can also lead to irrational behavior. Damasio had studied patients with peculiar brain injuries that suppress their emotions but do not affect their other cognitive abilities. In all studied patients, the lack of emotion has led to, among other things, a decline in decision-making ability. According to Damasio (ibidem), reason and emotion are connected in certain parts of the brain. The *system of rationality* of the brain is not built on top of a biological regulatory system, but in conjunction with and stemming from it.

Emotions improve reasoning by automatically reducing the number of options, which is commonly known as heuristics. When a person has a bad scenario or situation in their mind, they experience an unpleasant sensation that acts as a warning signal. Due to this signal, the person can immediately reject the ones causing unpleasant sensations and thus choose from a smaller set of options. Positive knowledge, in turn, acts as an incentive.

In addition to supporting reasoning, emotions have other important tasks that will not be readily apparent. David Hume, a classic of philosophy, believed that moral judgments, norms, and values are based on emotion: If we consider an act to be morally wrong, it is due to feelings of aversion aroused by the act. However, in different cultures, the same act can provoke different reactions. People cannot be seen in isolation from culture. Although there are at least 10 universally recognized feelings – happiness, sadness, fear, anger, surprise, disgust, contempt, confusion, sexual desire, relief – emotional reactions seem to be mainly learned through imitation; they strongly depend on enculturation. According to Hume, what all healthy people have in common is the ability to empathize, and this creates the foundation for an organized community and society.

Empathic behavior has been observed in non-human animals, even as distant from us as rats. A team of researchers led by neurobiologist Peggy Mason conducted an experiment in which a rat was placed in the vicinity of another rat trapped in a tank. The free rat learned to save its companion by opening the tank door. It did not open the tank when it was empty or if there were toy rats inside. Even when the alternative to releasing a fellow rat was to enjoy chocolate, rats typically rescued their companion first and then shared the chocolate with it (Bartal et alia, 2011).

It is difficult to teach morality to a rat, let alone human morality, to an insensitive machine. Due to feelings, people are inherently moral actors. Artificial intelligence does not have the biological tendencies, emotions, or culture in which to grow, making it difficult to operate and make decisions in a social environment where many problems are linked to ethical issues. Two groups of people, namely psychopaths and those diagnosed with autism, lack a community-important capacity for empathy. The lack of empathy in psychopaths has been explained by a limited ability to feel fear, sadness, and other negative emotions. Without emotions, it is easy to be indifferent to another's suffering. Empathy can be a hindrance to personal success in a society that places more emphasis on individual achievements than cooperation. Thus, even psychopaths can thrive and reach high positions (Babiak, 1995). Cool and calculating psychopaths are common sense people with no moral restraints. Some studies propose that successful psychopaths possess intact or enhanced neurobiological functions, leading to superior cognitive abilities. These abilities enable them to achieve goals through covert, nonviolent methods, distinguishing them from unsuccessful psychopaths who may have cognitive and emotional deficits leading to overt violent behaviors (Gao et alia, 2020, Ene et alia, 2022). An interesting question is: to what extent is sociopathy a social construct?

References

Arkoudas, K., & Bringsjord, S. (2014). Philosophical foundations. W K. Frankish & W. M. Ramsey (Red.), *The Cambridge handbook of artificial intelligence* (s. 655-680). Cambridge University Press.

Assunção, G., Patrão, B., Castelo-Branco, M., & Menezes, P. (2022). An overview of emotion in artificial intelligence. *IEEE Transactions on Artificial Intelligence*, *3*(6), 867–886. https://doi.org/10.1109/TAI.2022.3155946

Babiak, P. (1995). When psychopaths go to work: A case study of an industrial psychopath. *Applied Psychology: An International Review*, *44*(2), 171–188. https://doi.org/10.1111/j.1464-0597.1995.tb01073.x

Ben-Ami Bartal, I., Decety, J., & Mason, P. (2011). Empathy and pro-social behavior in rats. *Science*, *334*(6061), 1427–1430. https://doi.org/10.1126/science.1210789

Boden, M. (2016). *AI – its nature and future*. Oxford University Press.

Butler, A. B. (2009). Triune brain concept: A comparative evolutionary perspective. W L. R. Squire (Red.), *Encyclopedia of neuroscience* (s. 1185–1193). Academic Press. https://doi.org/10.1016/B978-008045046-9.00984-0

Campbell, N. A., Reece, J. B., Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., & Jackson, R. B. (2015). *Biology: A global approach* (10. wyd.). Pearson.

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford University Press.

Churchland, P. S. (2004). *Neurofilosofia*. Terra Cognita.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. Avon Books.

Ene, I., Wong, K. K.-Y., & Salali, G. D. (2022). Is it good to be bad? An evolutionary analysis of the adaptive potential of psychopathic traits. *Evolutionary Human Sciences*, *4*, e37. https://doi.org/10.1017/ehs.2022.36

Gao, Y., Schug, R. A., Huang, Y., & Raine, A. (2020). Successful and unsuccessful psychopathy. W A. R. Felthous & H. Saß (Red.), *The Wiley international handbook on psychopathic disorders and the law* (2. wyd., s. 357–379). Wiley-Blackwell. https://doi.org/10.1002/9781119159322.ch26

Goleman, D. (1995). *Emotional intelligence: Why it can matter more than IQ*. Bantam Books.

Haikonen, P. O. (2017). *Tietoisuus, tekoäly ja robotit*. Art House.

Hoffmeyer, J. (2014). *Biosemiootika: Uurimus elu märkidest ja märkide elust*. Tallinna Ülikooli Kirjastus.

Hofstadter, D. (2018). The shallowness of Google Translate. *The Atlantic*. https://www.theatlantic.com/technology/archive/2018/01/the-shallowness-of-google-translate/551570/

Kinman, A. I., et al. (2025). Atypical hippocampal excitatory neurons express and govern object memory. *Nature Communications*. https://doi.org/10.1038/s41467-025-56260-8

Kutrzeba, F. (2022). *Skills mismatch in the context of technological change*. Gdansk University of Technology Publishing House.

MacLean, P. D. (1998). The history of neuroscience in autobiography (Vol. 2, s. 242–275). Society for Neuroscience.

Narimisaei, J., Naeim, M., Imannezhad, S., Samian, P., & Sobhani, M. (2024). Exploring emotional intelligence in artificial intelligence systems: A comprehensive analysis of emotion recognition and response mechanisms. *Annals of Medicine and Surgery*, *86*(8), 4657–4663. https://doi.org/10.1097/MS9.0000000000002315

Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, *19*(3), 113–126. https://doi.org/10.1145/360018.360022

Pessoa, L., Medina, L., Hof, P. R., & Desfilis, E. (2019). Neural architecture of the vertebrate brain: Implications for the interaction between emotion and cognition. *Neuroscience & Biobehavioral Reviews*, *107*, 296–312. https://doi.org/10.1016/j.neubiorev.2019.09.021

Scheutz, M. (2014). Artificial emotions and machine consciousness. W K. Frankish & W. M. Ramsey (Red.), *The Cambridge handbook of artificial intelligence* (s. 740–760). Cambridge University Press.

Searle, J. R. (2012). Can computers think? W D. J. Chalmers (Red.), *Philosophy of mind: Classical and contemporary readings* (2. wyd., s. 560–563). Oxford University Press.

Telkänranta, H. (2015). *Millaista on olla eläin?* Suomalaisen Kirjallisuuden Seura.

Tretter, M. (2024). Equipping AI-decision-support-systems with emotional capabilities? Ethical perspectives. *Frontiers in Artificial Intelligence*, *7*, 1398395. https://doi.org/10.3389/frai.2024.1398395

Uexküll, J. von (2012). *Omailmad*. Ilmamaa.

Ylirönni, A. (2024). *Ajasta iäisyyteen – Tieteen, filosofian ja uskonnon näkökulmia kuolemaan*. Basam Books.

Younis, E. M. G., Mohsen, S., Houssein, E. H., & et al. (2024). Machine learning for human emotion recognition: A comprehensive review. *Neural Computing and Applications*, *36*, 8901–8947. https://doi.org/10.1007/s00521-024-09426-2

Internet sources:

Hofstadter, Douglas (2018). *The Shallowness of Google Translate*. Last visited, 06.03.2025: https://www.theatlantic.com/technology/archive/2018/01/the-shallowness-of-google-translate/551570/

Queensland Brain Institute (2025). *Types of Neurons*. Last visited, 06.03.2025: https://qbi.uq.edu.au/brain/brain-anatomy/types-neurons

Churchland, Patricia (2017). The First Neuroethics Meeting: Then and Now. https://patriciachurchland.com/wp-content/uploads/2020/05/2017-The-Brains-Behind-Morality.pdf