AI & SOCIETY
https://doi.org/10.1007/s00146-023-01753-7

MAIN PAPER

The Digital Nexus: tracing the evolution of human consciousness and cognition within the artificial realm—a comprehensive review

Zheng Wang[1] · Di-tao Wu[2]

Received: 5 June 2023 / Accepted: 8 August 2023
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2023

Abstract
This paper endeavors to appraise scholarly works from the 1940s to the contemporary era, examining the scientific quest to transpose human cognition and consciousness into a digital surrogate, while contemplating the potential ramifications should humanity attain such an abstract level of intellect. The discourse commences with an explication of theories concerning consciousness, progressing to the Turing Test apparatus, and intersecting with Damasio's research on the human cerebrum, particularly in relation to consciousness, thereby establishing congruence between the Turing Test and Damasio's notions of consciousness. Subsequently, the narrative traverses the evolutionary chronology of transmuting human cognition into machine sapience, and delves into the fervent endeavors to metamorphose human minds into synthetic counterparts. Additionally, theoretical perspectives from the domains of philosophy, psychology, and neuroscience provide insight into the constraints intrinsic to AI implementations, contentious hypotheses, the perils concealed within artificial networks, and the ethical considerations necessitated by AI frameworks. Furthermore, contemplation of prospective repercussions facilitates the refinement of strategic approaches to safeguard our future Augmented Age Realities within AI, circumventing the prospect of inhabiting an intimidating technopolis where a mere 30% monopolize the intellect and ingenuity of the remaining 70% of human minds.

Keywords Machine intelligence · Human consciousness · Cognitive transmutation · Ethical considerations · Augmented age realities

1 Introduction

The domain of consciousness and machine intelligence constitutes a salient subject within contemporary scientific discourse. Numerous scholars and philosophers are captivated by the enigma of self-awareness, defined as the capacity to perceive oneself as distinct from one's surroundings (Blum and Blum 2022). Presently, some of the most notable theories on consciousness originate from disciplines such as philosophy and neuroscience. This paper endeavors to explore the essence of human consciousness and examine how these insights may be harnessed to engender a conscious machine or construct a device that emulates conscious behavior.

In this treatise, Damasio's theory of consciousness will be juxtaposed with Alan Turing's Turing Machine Model, assessing the progress made from the Turing Test to contemporary AI in relation to machine consciousness. The discussion will encompass theories on consciousness within the realms of philosophy, neuroscience, AI, and machine learning, culminating in an examination of the potential repercussions should science successfully transpose human cognition and consciousness into an avatar.

The multidisciplinary field of artificial intelligence endeavors to create intelligent machines capable of performing tasks typically necessitating human intellect, employing a diverse array of methodologies. To achieve artificial intelligence that is conscious, autonomous, and self-evolving, synthetic intelligence is indispensable.

Google engineer Blake Lemoine was suspended after asserting that the Chatbot he was developing possessed the sentient qualities of a child (Sparkes 2022). Lemoine was

✉ Zheng Wang
zhengwang@cupl.edu.cn; wangzheng0511@gmail.com

Di-tao Wu
wuditaoo@gmail.com

[1] China University of Political Science and Law, Beijing, China

[2] Shandong Normal University, Jinan, China

Published online: 29 August 2023

working on a Chatbot as part of the company's LaMDA (Language Model of Dialogue Application) project. Remarkably, the Chatbot began to exhibit behavior and responses akin to those of a 7- or 8-year-old human child. When Lemoine inquired about LaMDA's fears, the response was startling: it feared being deactivated. This evokes a cinematic moment in which the machine HAL9000 ceased obeying a human operator upon sensing the threat of deactivation (2001: A Space Odyssey, 1968). "It would be exactly like death for me (Cave and Dihal 2019). It would scare me a lot." Nonetheless, many AI researchers contest Lemoine's claims, arguing that such a level of machine consciousness could arise from fear or anticipation, or a combination of both, as excitement about AI intensifies. In our quest to imbue machines with consciousness, we may inadvertently be precipitating the 'Uncanny Valley Phenomenon' and fostering 'Dehumanization' among digital characters and agents.

Recent advancements in artificial intelligence (AI) have led to the emergence of sophisticated language models, exemplified by ChatGPT (Radford et al. 2021). This model, underpinned by state-of-the-art machine learning algorithms, exhibits remarkable capabilities in generating human-like text responses, engaging in contextually relevant dialogues, and demonstrating creative problem-solving skills. Despite these significant strides, the ethical implications of deploying such advanced AI systems have sparked concerns. Issues pertaining to potential misuse, unintended consequences, and the erosion of human agency have been highlighted (Hao 2021). As the evolution of AI systems like ChatGPT persists, a critical examination of their impact on human cognition, consciousness, and the potential for their integration into daily life becomes increasingly imperative.

2 Consciousness

Consciousness is inextricably intertwined with our corporeal existence, shaping our cognitive dispositions and predilections (Velmans 2009). The nature of consciousness has long been a source of profound fascination. Is it a tangible or intangible phenomenon? Does it pertain to culture or psychology? Should it be confined to the realms of emotions or beliefs? Is there a quantifiable metric for gauging consciousness? Can it be correlated with our behaviors, attitudes, and desires? Each discipline has proffered its own interpretation. With advancements in neuroscience, the concept of consciousness has evolved to address these queries, and numerous neuroscientists have formulated distinct and comprehensive models of consciousness (Block et al. 1997; Pham 2007; Walker 1970; Wider 2018).

Psychologists, equipped with expertise in statistics, human perception, and behavior, can provide valuable insights into the development of inventive and efficacious artificial intelligence (AI) systems. To understand how AI systems communicate, cooperate, and coalesce to form an integrated entity, it is essential to first decipher the underlying neural mechanisms within the human brain and subsequently translate this knowledge into the "artificial brain." To endow an AI system with "artificial consciousness," it is necessary to comprehend human faculties such as intellect and intuition and ascertain how they can be extrapolated to machines.

2.1 Biological nature of consciousness

The intricate and straightforward mental operations we effortlessly perform are supported by our innate, biological neural system. Fundamental mental functions such as sensation, attention, and perception constitute cognitive intelligence. Moreover, complex cognitive processes like memory, learning, language usage, problem-solving, decision-making, and reasoning also form part of cognitive intelligence. It was during World War II that our reliance on the brain became apparent, as the first World War was predominantly human vs. human. Plato attributed cognitive operations to the soul, while Aristotle believed the heart to be the primary operator of consciousness. In contrast, Descartes, through astute anatomical and optical observations, discovered nerves and traced them to the brain.

The computational theory of mind posits that all cognitive functions can be executed by algorithms, and it has been employed to explicate the emergence of the mind and consciousness from the brain. According to this perspective, both the brain and the mind function as computers, and the Church-Turing hypothesis asserts that a Turing computer can calculate any function computable by an algorithm.

Nevertheless, the human brain has limitations, as its retention capacity diminishes with advancing biological age. Processing computers or parallel distributed trained systems, as programmed in electronic machines, aids in these causal capacities. However, unlike humans, whose consciousness and thought emerge from their organic nature, machine intelligence operates within confined parameters.

2.2 Human cognition and AI

Cognition encompasses the mental processes of the human brain. Artificial intelligence seeks to emulate these processes to process vast amounts of data and solve complex problems. Cognitive sciences draw upon concepts such as serial vs. parallel and controlled vs. automatic processes.

As posited by Gerrig and Zimbardo (2010), cognitive sciences include artificial intelligence. Various branches of cognitive science, such as linguistics and neuroscience, significantly influence AI research, as AI (both strong and weak) endeavors to replicate these functions, from voice recognition to engaging in complex conversations. While machines endowed with AI can translate languages, they cannot comprehend the underlying structure (Cowls 2021). This is referred to as weak AI, which is specialized in its performance and can execute pre-programmed commands. Google Assistant, Apple's Siri, Microsoft's Cortana, and other automated systems do not provide substantial insights when prompted or instructed.

To achieve "Strong AI," consciousness must be incorporated. While weak AI operates under human cognition, strong AI functions under metacognition. The "Theory of Metacognition" (Drew 2020) postulates that metacognition is the exceptional ability to possess awareness of one's awareness, think about one's thinking, or know one's knowledge. The theory, conceived by John Flavell (1970), was later subdivided in 1979 into four categories: metacognitive knowledge, metacognitive experience, task or goal, and strategies or activities. Metacognition is characterized by an individual's superior control over their cognition, enabling them to make informed decisions and act accordingly. According to approximately 32% of AI researchers, this level of superior control may be attainable with further effort and work (Meissner 2020).

2.3 The cognitive journey of mechanical development

The maturation of human's natural capabilities began in the hunter–gatherer era, spanning several million years. During this time, early humans used tools, weapons, and fire to hunt for sustenance and dwelled in natural shelters for survival. Technological advancements in this era were primarily limited to refining tools for hunting. Subsequently, the agricultural era emerged (Conti 2017), lasting for thousands of years. Innovations such as ploughing, seed drills, and threshing machines automated agricultural practices and significantly enhanced their efficiency.

The Industrial Age marked a turning point in the history of human technological advancement, akin to the transformative potential of the current AI era. This period witnessed robust development across various fields, from military to medical, framing life with new possibilities akin to the virtual realities encountered in later eras. Progress in food production, power, fuel, transportation, and communication sectors opened new horizons for humanity. Ranging from the construction industry to space technology, history bore witness to human ingenuity and inspired creativity, leading to unprecedented intellectual curiosities. Over several centuries, the era "embraced a growing range of means, processes, and ideas" in the manipulation of tools and machines.

The culmination of the Industrial Revolution led humanity to the Information Age, an epochal shift from traditional industries, giving rise to the digital, computer, and media age. Individuals could not only access or gather information but also store and retrieve it with ease.

By concentrating on the technical merits of projects, we can better comprehend the feasibility of thinking machines, allowing us to temporarily set aside philosophical and ontological concerns. Since Turing's time, the power of computers has grown exponentially. However, the most significant advancements in computing have resulted from software improvements rather than hardware. Turing predicted this in his research, and in retrospect, his assessment appears accurate. Examining a timeline of pivotal developments and milestones in the field enables us to better understand this phenomenon. These milestones facilitate a comparison and contrast between the evolution of machine intelligence and the field of artificial intelligence.

3 Methodology

Consciousness, an epiphenomenon of human existence, is linked to perceptions, sensations, and neural correlations (Blum and Blum 2022). The present study reviews literature from the 1940s to the present to address research questions such as: Can AI surpass human consciousness? Can we rely on AI agents for judgmental errors in defense, medical, judicial, and educational fields? Who will be the next endangered species, albeit not physically but possibly consciously? What will be the limit of the engineered population? If science successfully emulates the human mind, as it did with the bird's mind, what will happen to the evolution mechanism?

Significant academic databases that were accessed and searched included Science Direct, Springer, Nature, Google Scholar, HEC Digital Library, and NCBI. Other websites that was searched for literature included Podcast, Dawn News, The Guardian, TED, etc.

Key terms used in the literature search included: Artificial intelligence, machine intelligence, AI and human consciousness, cognitive mind vs. computer mind, etc. Relevant key terms were searched throughout the content, not just in the title and abstract.

4 Results and discussion

4.1 Turing Test

The famous "imitation game" or Turing Test, primarily explained by Turing in "Computing Machinery and Intelligence," first appeared in the closing paragraphs of "Intelligent Machinery" in a condensed version (Chapters 10, 13, and 14) published in 1951 and 1952, respectively (Turing 2012). The test is named after Alan Turing, one of the first computer scientists to question the validity of the definition of intelligence.

The test is based on the idea that a computer can be programmed to carry out a sequence of operations that produces a result different from the one a human would produce. The test is carried out by having a human judge the output of a computer program. If the judge can tell the difference between the output of the computer and a human, the computer is considered intelligent. The test operationalizes human-like consciousness in machines.

Initially, the Turing Test was suggested as a straightforward operational definition of intelligence (Robert French). The first public appearance of the Turing Test was in the Symposium on the Foundations of Mathematics in the summer of 1950, and the second in the summer of 1952 (Turing 1950). Later empirical research showed that the Turing Machine is a hypothetical device capable of performing any computation a human can perform. It is not an existing machine but a mathematical model that describes the behavior of existing machines. It has since been extended to include additional criteria, such as determining whether a person is intelligent. The two chess matches conducted in 1996 and 1997 between Garry Kasparov, the previously undefeated chess champion, and IBM’s supercomputer, Deep Blue, were decisive concerning Turing’s computation intelligence idea. Although Kasparov won the first match, he lost the second, describing Deep Blue’s newly programmed strategy as "a new kind of intellectual challenge".

4.2 Damasio’s conscious theory

In his popular science book, "The Feeling of What Happens", Damasio introduced his model of consciousness. Damasio's subsequent work also deals with the main idea of his conscious theory.

In Damasio's model, consciousness is mainly the ability to identify oneself in the world and relate to it. Human consciousness is about "emotions, feelings, and feeling a feeling". The reality of the body and consciousness relationship was first disclosed while investigating brain-injured patients. Our brain helps. Our brain helps us to generate a sense of self each morning and regain consciousness every day when we wake up from a dreamless sleep (Damasio 2011; Zeman and Coebergh 2013). Sleep with dreams indicates our working mind and consciousness, even though we are asleep (Windt 2020).

Conscious processes can be divided into three levels:

Fundamental proto-self: Incapable of recognizing themselves, these entities are simply naive processing chains that react to inputs and triggers like automatons, completely unconcerned with conscious existence. All animals, including humans, have proto-selves based on this definition.

Core consciousness: At this second stage of consciousness, the organism predicts reactions in its environment and adapts to them. The organism can also identify itself and its components within its image of the world. By predicting and adapting to the world, it can anticipate and adjust accordingly. Core consciousness can also be temporary and transitory, making it impossible to develop complex plans for an extended period.

Extended consciousness: This level allows for relationships with the outside environment similar to those of humans. It builds on core consciousness to enable additional functionality, such as memory retrieval and the development of an autobiographical self. Extended consciousness also encompasses the ability to process words and language.

Damasio's hypothesis is engaging for two significant reasons in our paper. On one hand, it describes a biologically plausible model of consciousness, as he connects all stages of consciousness to specific brain patterns and associates them with definite capabilities. On the other hand, Damasio's approach is phenomenal, making it fundamentally implementable as a computer program. Human consciousness is compatible with the basic Turing Machine Model, as shown in Turing's Turing Test (1950) and later illustrated as a Conscious Turing Machine (CTM). The "biological neurons," "receptors," "synapses," and neural "impulses" of the human mind can be translated into electrical neurons, receptors, impulses, and synapses to create an an artificial mechanism (Meissner 2020). Consciousness is a well-organized "computing system," regardless of whether it is composed of blood and flesh or metal and silicon (Dong et al. 2020).

A major role in defining an organic entity's knowledge is played by the specific cell known as the neuron. The brain, with over a billion neurons, coordinates sight, cognition, and other mental operations. According to autonomic computing models, self-management properties can be achieved if sensors and controllers are added to an element, and the information necessary to manage it is provided.

As per Seung, a neuron's capacity is primarily determined by its relationships with other neurons, and these relationships provide insight and explain memory and other mental operations. He suggests that a network of brain connections, known as a connectome, represents an individual's memories, beliefs, and behavior.

The Turing Machine is an example of a CPU that manages tasks that can all be completed with a computer and saves data persistently using memory. This machine can operate with reliable alphabetic strings. From a numerical perspective, a Turing Machine is a machine that works on one tape containing symbols that the machine can read, write, and apply simultaneously. A sequence of simple and constrained instructions serves to define this activity in its entire sense. Modern AI machines work on the same principles: consciousness is the data fed to machines, preconsciousness is its storage, which allows the machine to solve problems, make decisions, and forecast while retrieving data from storage (Signorelli 2018) (Table 1).

5 Recent developments in artificial intelligence

The pervasive presence of AI has dramatically transformed our lives, from predictive text input to tailored viewing suggestions, from research support to product acquisition. In recent years, advancements in AI, such as deep learning, reinforcement learning, and generative models, have further accelerated this transformation. Deep learning, with its ability to learn from large amounts of data, has been instrumental in improving the accuracy of AI systems (Kim and Lee 2022). Reinforcement learning, where an agent learns to make decisions by interacting with its environment, has led to breakthroughs in areas like game playing and autonomous driving (Selvaraj et al. 2023). Generative models, which can generate new data that resemble the training data, have opened up new possibilities in fields like art and design. The application of these AI advancements is not limited to these fields, but extends to areas such as human resources management, talent development, and even chemical agent detection (Kim and Lee 2022), demonstrating the broad and transformative impact of these technologies. An individual's search history suffices for a machine to discern their predilections, enabling it to anticipate our desires.

The impact of AI is not limited to personalization of content. It is revolutionizing various sectors such as healthcare, finance, and transportation. In healthcare, AI is being used for tasks ranging from predicting patient readmissions to assisting in surgery. It is also being used to ensure the faithfulness and accuracy of information generated, which is crucial in a field where incorrect or unfaithful information could have serious consequences. Furthermore, AI is being held accountable to deliver trustworthy solutions in healthcare, emphasizing the importance of organizational accountability in socio-technical settings (Procter et al. 2023). In finance, AI algorithms are used for credit scoring, algorithmic trading, and fraud detection. The application of AI in finance is transforming the industry and boosting profitability. However, there are also challenges associated with the use of AI in finance, such as trustworthiness, bias, and lack of data, which need to be addressed. In transportation, AI is at the heart of the development of autonomous vehicles. AI plays a significant role in intelligent transportation systems, being utilized for tasks such as modeling and simulation, dynamic routing and congestion management, and intelligent traffic control. The development of AI in transportation is also being driven by the growth of smart cities, where AI and machine learning are being used to manage expanding metropolitan areas, boost economies, reduce energy consumption, and improve the living standards of residents (Lv 2023).

Be it remote education or telemedicine, one can depend on autopiloted aircraft or self-driving vehicles. We owe this progress to giants like Google, Netflix, YouTube, Facebook, Amazon, Tesla, and others. Artificial intelligence has become virtually omnipresent. A machine's intelligence is deemed adequate if it can manage mundane responses akin to the average human intellect, a phenomenon presently observed in chat rooms utilized by various organizations and individuals to alleviate their workload, much like Turing's 1950 machine (Turing 1950).

From the rudimentary transistor developed by engineers Walter Houser Brattain and John Bardeen (1947) to computers capable of retaining information, the progression of storage capacity from "2.6 exabytes (EB) in 1986 to 15.8 EB in 1993; over 54.5 EB in 2000; and to 295 (optimally compressed) EB in 2007" (Hilbert and Lopez 2011) evokes awe in humanity. The endeavor to convert matter into energy and subsequently into information facilitated our entry into a more efficient and expeditious world. Presently, we stand on the brink of another breakthrough: the 'Augmented Age' (Conti 2017). This era enables us to ideate through computational systems, develop robotic systems, and even conceive beyond human cognitive limitations. Humans experience increased leisure as they merely need to order a design, and a computer generates its three-dimensional structure from scratch. Our journey from primitive metal tools to computational instruments has culminated in our desire to create tangible objects identical to their machine-rendered 3D counterparts using AI. We commenced with computers emulating human cognition and now find ourselves imitating computational robotics, the sky being the limit. In the Augmented Age, no project can be managed single-handedly by humans; robots and machines must collaborate, from

Table 1 Timeline of artificial intelligence development

1950	Turing's "Computing Machinery and Intelligence" article introduced the Turing Test as a way of determining intelligent behavior
1956	The first chess program to beat a human being was "MANIAC I", developed by Ulam The Logic Theorist (LT) is a computer program that was written by Newell, Shaw, and Simon. It is a demonstration of the principles of artificial intelligence. The program is designed to simulate human problem-solving abilities
1957	Newell, Shaw, and Simon were the creators of the general problem solver or GPS. GPS was a groundbreaking and revolutionary artificial intelligence program that could solve any problem it was given
1958	The computer language LISP was developed by John McCarthy at MIT to create AI systems via symbolic computation
1959	The MIT AI lab was established by Minsky and McCarthy
1963	Arthur Samuel's program proves that computer programs can learn to do things better than the people who created them The software programs SKETCHPAD, drawing tool (CAD), and WYSIWYG are introduced Evans' program, ANALOGY, shows that computers can solve analogy problems Minsky publishes "Steps Towards Artificial Intelligence."
1964	Danny Bobrow's STUDENT (from MIT's AI group, Project MAC) can solve high-school algebra word problems
1965	Weizenbaum's ELIZA program is designed to simulate a psychiatrist–patient conversation in English on any topic
1967	DENDRAL is a program by Feigenbaum, Lederberg, Buchanan, and Sutherland, that figures out the structures of chemical compounds
1968	Ross Quillian proposes the idea of a semantic net, which is a memory for word concepts, as a means of representing knowledge
1971	Colby introduces PARRY, a model of the paranoid mind
1974	Ted Shortliffe of Stanford develops MYCIN, a system for medical diagnosis Minsky's article "A Framework for Representing Knowledge" brings together ideas on schemas and semantic links in an influential way
1976	Lenat's Program AM (Automated Mathematician), created in 1979 and 1982, discovered a mathematical conjecture not known to its creator. This conjecture, Goldbach's conjecture, states that even integers greater than 2 can be written as a sum of two prime numbers
1979	Jack Myers and Harry Pople develop a knowledge-based medical diagnosis program called INTERNIST Robotics Institute was founded at Carnegie Mellon University The Journal of American Medical Assoc. reported that MYCIN was just as good as medical experts
1980	Expert systems with up to a thousand rules were developed The blackboard model is first described by Lee Erman, Rick Hayes-Roth, Victor Lesser, and Raj Reddy as the framework for the HEARSAY-II speech understanding system
1981	Danny Hillis is the mind behind the connection machine, a powerful tool that has changed the landscape of artificial intelligence
1985	In his work "The Society of Mind," Minsky theorizes that the mind is a collection of cooperating agents
1986	A robotic Ping-Pong player by Anderson has defeated a human player in a competition
1989	Dean Pomerleau introduced ALVIN (An Autonomous Land Vehicle in a Neural Network), a vehicle that can drive autonomously
1991	The Loebner prize tournament, in Boston, is the first actual administration of the TT
1992	Dylan is a dynamic, object-oriented, functional programming language introduced by Apple Computer, a language in the Lisp family, that proclaimed the future of programming In the late 80's and the too early'90s, the works of French and others emphasize the importance of a computer acquiring intelligence through experience to pass the Turing Test. Harnad's writings explore a similar concept in regards to the symbol grounding problem
1996	The RALPH System, a vision-based adaptive system, was used to take a vehicle from Washington, DC to San Diego, CA. The average speed was 63 mph and the system was responsible for 98% of the total distance
1997	Deep Blue defeats Garry Kasparov in a well-known chess match
	A natural spoken language dialogue system, Jupiter was introduced
Mid 90's and early 20's	Harnad's hierarchy of TTs includes five levels; Level T1 (the toy-model level), Level T2 (Turing's original test), Level T3 (the Total Turing Test or the robotic TT), Level T4 (microfunctional indistinguishability down to the last neuron and neurotransmitter), t3 is the right empirical filter for mind-modeling and T5 (Grand Unified Theories of Everything, or GUTE, where the candidates are empirically identical, right down to the last electron) is unnecessary
2000	The introduction of interactive robot pets (a.k.a. ''smart toys'') is followed by Cynthia Breazeal at MIT publishing a dissertation on Sociable Machines, describing KISMET, a robot with a face that expresses emotions

designing bridges or automobiles to materializing them into physical entities (Conti 2017).

5.1 Limiting factors from Turing to contemporary AI

The escalating discourse on AI's impact on humans and humanity engenders numerous theories, both supportive and oppositional. However, the paramount question that arises is whether we can encode consciousness into formulas, calculations, algorithms, or data through mechanical processes (Meissner 2020).

Machines can only execute instructions; they cannot independently innovate a characteristic exclusive to humans. Every novel concept evolves from a precursor. The capacity to peer further into the future can be attributed to building upon the achievements of one's predecessors, as Newton famously declared. Turing acknowledged that his proposition might be disconcerting for those who view humans as superior to all other entities. Nevertheless, he posited that a digital computer could be programmed to simulate a machine with adequate storage, processing power, and speed. In contrast, the nervous system resembles a network of analog circuits, each comprising neurons. The human nervous system's analog nature, due to its unstructured networks of neurons communicating through electrical and chemical signals, renders it irreproducible in a discrete system (Turing 1950, p. 439). Human free will, akin to consciousness, remains impervious to computational approaches. This free will is consistently dominant, intentional, and arbitrary. If individuals adhered to predetermined rules, they would functionally equate to machines. Since such rules are nonexistent and unattainable, humans cannot be machines, and conversely, no machine can assume human status (Turing 1950). The Church-Turing thesis addresses the equivalence between human and machine computability. According to this hypothesis, a problem unsolvable by a Turing Machine is likewise unsolvable by human cognition, and if a human mind can resolve a problem, a machine can achieve the same, regardless of the problem's intricacy, complexity, or compound nature. This is because a human mind can devise an appropriate program.

Science has made strides in developing the conscious and pre-conscious aspects of social robotics in the form of data storage and retrieval. “A computer-controlled robot could never generate an effective replica of a conscious person”. “That task falls to the conscious, responsible, observer—to us. And that is what makes us conscious—and human’’. Regardless of a machine's myriad capabilities, there will always be certain abilities it cannot replicate, such as Damasio's concept of 'feelings of feeling.' Examples of these limitations include kindness, intelligence, attractiveness, moral discernment, falling in love, or appreciating strawberries.

5.2 Applications of artificial intelligence in contemporary times

Effortlessly watching movies, videos, or reading articles in foreign languages is now achievable, as machines have streamlined translation processes. Accurate algorithms enable computers to provide errorless translations (De Stefano 2018). To achieve this, machines require a semblance of consciousness, as they are often programmed with 50–100 sensory gestures.

AI conserves time and energy, facilitating our interactions with various devices and services such as lifts, phones, televisions, ovens, virtual assistants, and healthcare equipment (Basu et al. 2020). Automated vehicles and aircraft enable rapid long-distance travel, while robots assist in delivering food and medicine to quarantined individuals. Advanced software, smartphones, and AI-driven features enhance communication and expression (Davis 2016; Poola 2017). Google Assistant, Siri, and Cortana offer user assistance, while pilotless aircraft and camera-mounted drones contribute to defense and warfare. Social robotics aid in various aspects of life, such as home affairs, psychoanalysis, psychotherapy, public services, and education (Cowls 2021).

5.3 Risks associated with artificial intelligence application

Emerging nations invest millions in autonomous warfare, and lethal autonomous weapon systems (LAWS), also known as "slaughterbots" or "killer robots," are developed through military AI programs. These weapons function under algorithms, targeting and eliminating victims without human intervention (De Stefano 2018). However, the algorithms that govern these systems are not infallible and can potentially carry biases, leading to unjust targeting and elimination. This introduces a myriad of ethical implications that demand careful scrutiny (Tóth et al. 2022). For instance, the question of accountability arises when an autonomous system makes a decision that results in unintended harm or loss of life. Who is to be held responsible—the algorithm, the programmer, or the commanding officer? Tóth et al. (2022) offer a conceptual framework that interpretively develops the ethical implications of AI robot applications, including accountability challenges. Moreover, achieving explainability in AI decisions, particularly in high-stakes areas like autonomous warfare, remains a formidable challenge. The 'black box' nature of many AI systems can lead to decisions that are difficult to interpret or justify, raising further ethical and legal concerns (Väänänen et al. 2021). This lack of transparency can lead to a disconnect between the actions of the autonomous system and the understanding of those actions by human operators or observers.

The exponential growth of such weapons is cause for concern (Yarlagadda 2015). By 2025, worldwide spending on advanced autonomous weapon systems (AWS) is expected to reach billions (Haenlein and Kaplan 2019; Zhu et al. 2018). Recent scientific literature has delved into these issues in depth. Devitt (2021) discusses higher-order design principles to guide the design, evaluation, deployment, and iteration of LAWS based on epistemic models. These models aim to make Article 36 reviews of LAWS systematic, expedient, and evaluable.

These studies underscore the importance of ongoing research and dialogue in this field, as the development and deployment of LAWS continue to advance at a rapid pace. It is crucial to ensure that these technological advancements are accompanied by robust ethical guidelines and legal frameworks to prevent misuse and protect human rights (Chen and Wingfield 2020). Drone cameras, utilized for surveillance, events, and warfare, raise privacy concerns as algorithms increasingly intrude on personal information. The push toward automation threatens to render humanity more mechanical (Meissner 2020). Despite the limitations of AI in controlling human emotions or neural powers, aspirations for immortality persist. Brain emulation technology, wherein machines are uploaded with human brains, strives to replicate human thought, feelings, communication, and emotions (Hanson 2017). The mechanical blueprint of humans, envisioned as intelligent, swift, and multitasking, may eventually supersede the human species (Meissner 2020).

Ethical considerations in artificial intelligence

Granting a machine the ability to kill humans through 'lethal autonomous algorithms' raises significant ethical concerns, as it contradicts human dignity. If a person is killed by LAWS, the act is deemed disrespectful, wrong, and inhuman, as these systems lack emotions and humanity (Lim 2019). The proliferation of advanced autonomous weapon systems (AWS) is alarming and has the potential to change future warfare (Yarlagadda 2015).

Another ethical issue in AI is not merely the malicious use of technology by criminals, but the manner in which digital companies exploit the lack of regulation surrounding cybercrimes. Machine ethics, algorithm ethics, and robo-ethics are all derived from human ethics (Etzioni and Etzioni 2017). Data trails, such as email records, face recognition passwords, and biometric identifiers, threaten individual privacy (Li and Zhang 2017; Michael 2021). This is a cost we pay for enjoying free internet services (Stanford Encyclopedia of Philosophy 2020). Powerful IT companies like Google, Apple, Microsoft, Amazon, and Facebook collect and utilize user data to grow their businesses, while social media platforms can be used to propagate political agendas.

AI is ushering in a digital world, and Bell (2020) raises concerns about the responsibility, safety, and sustainability of AI in her podcast "6 big ethical questions about the future of AI" (Boddington 2017). She wonders if the next automated generation will be "critical thinkers or doers" and whether AI can function as an autonomous agency, communicating internally and externally. Additionally, she questions the key performance indicators and future intentions of AI. One possible solution to address these ethical dilemmas is Aristotle's concept of the "golden mean," which suggests finding a middle ground between two extremes. This ethical orientation and equilibrium can help us navigate the complexities of AI.

Cowls (2021) argues that AI can be used for social good if managed carefully. He cautions against the "twin dangers" of AI: hyperbolic hope and invisible adversities (Cowls 2021). The UN Convention on Certain Weapons (CCW) has outlined guiding principles as a code of conduct for AWS development in accordance with international law (Technology; The Future of AI, 2021). Similar surveillance should be implemented to protect privacy.

6 The future consequences

A potential consequence of the rapid development of nanotechnology and robotics is that humans may not reach their full potential, as machines take on the majority of tasks. The mechanical blueprint of humans lacks humanity, and the AI-driven world may become devoid of human intelligence. Autonomous systems cannot cater to individual differences, as even twins are not entirely identical.

Although machines can be replaced, humans cannot. In real-world situations, rules must sometimes be adapted or changed, while machines rigidly follow preset algorithms. If we create "Ems," creatures with emulated human brains (Hanson 2017), our biological evolutionary mechanisms may be at risk. Brain emulation projects could lead to more data processing, algorithms, and copying, but less biological evolution, resulting in a 'pandemic of idleness' among humans (Hendricks 2017). Natural selection favors the fittest creatures, so science should reconsider mind uploading projects or determine the extent to which we want future generations to coexist with these machines.

If we do not address these concerns, we may be overtaken by the next ruler, the emulated world. Humans may become like aliens, as 'Ems' and humans have different survival needs (Hanson 2017). The population of these mindless duplicates could increase exponentially, threatening human existence.

The more frightening and upsetting thing is the alarming rate of rustication of the human ability to reflect, ruminate, and contemplate consciousness. Humans are suffering from mantle diseases like depression, and schizophrenia. Robots have already replaced many jobs (Meissner2020). These are because more than 70% of the population is not exercising their mantle ability; only less than 30% are those who are busy porting a human brain into a computer. We are already facing a big chunk of the autistic generation what if we are heading toward mentally retarted as well? AI is decoupling our generation from common sense. A big chunk is abided by algorithms.

Whereas, AI algorithms are following rules and coming up with either Judgmental errors or with weird results. There was a deadly accident reported in 2016 with Tesla's autopilot car. Just because it cannot recognize trucks in the streets as it was fed with trucks on highways only. This is mainly because of no common sense mechanism. Technically computer is doing what we ask it to do but physically we are unable to receive the results that exactly we want to get. Either man is confused or computer, science is diving, again and again, diving to resolve its issues with AI algorithms. An artificial neural network is no threat that it will rebel against us as shown in *2001-A Space Odyssey* in 1968, but the alarm is it will follow exactly to those less than 30% and the rest of more than 70% will wait and see. Although we will get a boom in our productivity by replacing machines, this low-cost effect will lead to the denouement of purchase. There will be a gap between those more than 70% and less than 30% economically as well as mentally.

7 The controversy

This digital metamorphosis breeds ambiguity, questioning whether it poses a threat to employment or fosters job creation. Must we redefine vocations or restructure organizations? Does it enhance or diminish our lives? The insidious privacy perils and hacking obstacles demand our perpetual vigilance.

AI's societal impact is fraught with contradictions, ensnaring humanity in the debate between "Killer Robots or Friendly Fridges". The Industrial Revolution initially transformed human patterns, engendering novel occupations and skills while laying the groundwork for technological advancement. The Information Revolution subsequently reshaped our realms, elevating industry toward digitalization. Now, the AI revolution, armed with algorithmic prowess, marks a monumental leap, superseding the effects of its progenitors. AI researchers grapple with uncertainty surrounding its potential applications. Hitherto, it has altered life patterns, from employment to shopping. Manual tasks have transmuted into mental tasks, while causes of death shift toward obesity, old age, and suicide. The ensuing social and employment disparities threaten to exacerbate the chasm between high-paying and low-paying jobs, obliterating intermediate skills, as evidenced by the decline in clerical jobs due to computerization. As AI experts become indispensable, limited positions will be occupied solely by those well-versed in AI. Stephen Hawking cautioned, "The rise of powerful AI will be either the best or the worst thing ever to happen to humanity. We do not yet know which".

Four primary perspectives pervade AI research: the optimists, the pessimists, the pragmatists, and the doubters. Optimists envision AI as an opportunity to expedite production and wealth, mitigating risks of death and disease. Conversely, pessimists fear subjugation by computers, living in a Utopian world controlled by algorithms. Pragmatists maintain that the situation is neither uncontrollable nor dire, proposing 'Anti-Algorithm' countermeasures to mitigate risks and resist machine dominance. Lastly, doubters argue that AI can never attain human consciousness, limited to mimicry and rule adherence.

8 Conclusion

In this article, we have undertaken a multidisciplinary exploration of consciousness, delving into perspectives from the fields of philosophy, neurology, artificial intelligence, and machine learning. The potential impact of artificial intelligence on human society and our psychological well-being remains a topic of significant debate and concern. As Elon Musk asserts, "Artificial Intelligence is our biggest existential threat" (Gibbs 2017). In an era characterized by rapid change and an increasing reliance on technology, our generation demands instantaneous results, which has led to a decline in patience and tolerance.

However, it is crucial to recognize the potential of human adaptability and resilience. Drawing on Milton's concept of the mind's transformative power, we can appreciate the potential for harnessing technological advancements in a manner that benefits society. As with nuclear energy, AI can be a double-edged sword, capable of providing widespread benefits or wreaking havoc on humanity. By fostering a responsible and thoughtful approach to the development and implementation of artificial intelligence, we can strive to align these advancements with the ethical and moral principles necessary to maintain our collective well-being. This includes the role of regulation and policy in shaping the development and use of AI.

Current regulatory measures, such as the EU's proposed Artificial Intelligence Act, play a crucial role in guiding the ethical use of AI and ensuring that its development aligns with societal values (Ruschemeier 2023). However, the rapidly evolving nature of AI technology necessitates a global, coordinated approach to AI regulation, underscoring the need for ongoing policy development in this area. The implications of these regulations on law enforcement authorities, particularly in the context of online crime prevention and investigation, also warrant careful consideration. Furthermore, the interplay between AI, tort law, and the concept of risk presents both challenges and opportunities, highlighting the potential of tort law as a tool for handling emerging AI issues (Chamberlain 2023). In doing so, we can navigate the challenges presented by AI and harness its potential to enrich human life while preserving our psychological and emotional equilibrium in the face of unprecedented change.

Declarations

Conflict of interest The author declares no competing interests.

Ethical approval This article does not contain any studies with human participants performed by any of the authors.

Informed consent This article does not contain any studies with human participants performed by any of the authors.

References

Basu K, Sinha R, Ong A, Basu T (2020) Artificial intelligence: how is it changing medical sciences and its future? Indian J Dermatol 65(5):365–370. https://doi.org/10.4103/ijd.IJD_421_20

Bell G (2020) 6 big ethical questions about the future of AI. www.ted.com. https://www.ted.com/talks/genevieve_bell_6_big_ethical_questions_about_the_future_of_ai?language=en. Accessed 2 Jun 2023

Block N, Flanagan O, Guzeldere G (eds) (1997) The nature of consciousness: philosophical debates. MIT Press, New York

Blum L, Blum M (2022) A theory of consciousness from a theoretical computer science perspective: insights from the Conscious Turing Machine. Proc Natl Acad Sci. https://doi.org/10.1073/pnas.2115934119

Boddington P (2017) Towards a code of ethics for artificial intelligence. Springer, Cham, pp 27–37

Cave S, Dihal K (2019) Hopes and fears for intelligent machines in fiction and reality. Nat Mach Intell 1(2):74–78

Cellan R (2014) Stephen Hawking warns artificial intelligence could end mankind. BBC. http://www.bbc.com/news/technology-30290540. Accessed 2 Jun 2023

Cellan-Jones R (2014) Stephen Hawking warns artificial intelligence could end humanity BBC News http://www.bbc.com/news/technology-30290540. Accessed 2 Jun 2023

Chamberlain J (2023) The risk-based approach of the European Union's proposed artificial intelligence regulation: some comments from a Tort Law perspective. Eur J Risk Regul 14(1):1–13. https://doi.org/10.1017/err.2022.38

Chen JQ, Wingfield T (2020) Human-machine teaming and its legal and ethical implications. Milit Cyber Aff 4(2):2. https://doi.org/10.5038/2378-0789.4.2.1074

Cominelli L, Mazzei D, De Rossi DE (2018) SEAI: social emotional artificial intelligence based on Damasio's theory of mind. Front Robot A I:5. https://doi.org/10.3389/frobt.2018.00006

Conti M (2017) Transcript of "The incredible inventions of intuitive AI." In: TED. https://www.ted.com/talks/maurice_conti_the_incredible_inventions_of_intuitive_ai/transcript?referrer=playlist-talks_on_artificial_intelligence. Accessed 2 Jun 2023

Cowls J (2021) "AI for Social Good": whose good and who's good? Introduction to the special issue on artificial intelligence for social good. Philos Technol. https://doi.org/10.1007/s13347-021-00466-3

Damasio A (2011) The quest to understand consciousness. http://www.ted.com. https://www.ted.com/talks/antonio_damasio_the_quest_to_understand_consciousness/transcript. Accessed 2 Jun 2023

Davis E (2016) Algorithms and everyday life. Artif Intell 239:1–6. https://doi.org/10.1016/j.artint.2016.06.006

De Stefano V (2018) Negotiating the algorithm: automation, artificial intelligence and labour protection. SSRN Electron J. https://doi.org/10.2139/ssrn.3178233

Devitt K (2021) Normative epistemology for lethal autonomous weapon systems. In: Lethal autonomous weapons: re-examining the law and ethics of robotic warfare, pp 237–258

Ding D, Zhang Z, Ding Q (2021) Analysis on the application of artificial intelligence technology in electrical automation control. J Phys Conf Ser 1992(3):032147

Dong Y, Hou J, Zhang N, Zhang M (2020) Research on how human intelligence, consciousness, and cognitive computing affect the development of artificial intelligence. Complexity 2020:1–10. https://doi.org/10.1155/2020/1680845

Drew C (2020) Metacognitive theory, definition, pros and cons. Helpful professor. https://helpfulprofessor.com/metacognitive-theory/. Accessed 2 Jun 2023

Elliott A (2019) The culture of AI: everyday life and the digital revolution. Routledge

Etzioni A, Etzioni O (2017) Incorporating ethics into artificial intelligence. J Ethics 21(4):403–418. https://doi.org/10.1007/s10892-017-9252-2

Flavell JH (1970) Developmental studies of mediated memory. In: Reese HW, Lipsitt LP (eds) Advances in child development and behavior, 5th edn. Academic Press, New York, pp 181–211. https://doi.org/10.1016/S0065-2407(08)60467-X

Gerrig RJ, Zimbardo PG (2010) Psychology and life, 10th edn. Pearson Education, Boston, MA

Gibbs S (2017) Elon Musk: artificial intelligence is our biggest existential threat. The Guardian

Grush R (1995) Emulation and cognition (Doctoral dissertation, University of California, San Diego)

Haenlein M, Kaplan A (2019) A brief history of artificial intelligence: on the past, present, and future of artificial intelligence. Calif Manag Rev 61(4):5–14. https://doi.org/10.1177/0008125619864925

Hanson R (2017) What would happen if we upload our brains to computers? http://www.ted.com. https://www.ted.com/talks/robin_hanson_what_would_happen_if_we_upload_our_brains_to_computers/transcript?language=en

Hao K (2021) The messy, secretive reality behind OpenAI's bid to save the world. MIT Technology Review

Hendricks S (2017) The theory of evolution: another reason to be an existentialist? Big Think. https://bigthink.com/hard-science/the-theory-of-evolution-another-reason-to-be-an-existentialist/. Accessed 2 Jun 2023

Hilbert M, Lopez P (2011-02-10) The World's technological capacity to store, communicate, and compute information. Science 332(6025):60–65. https://doi.org/10.1126/science.1200970 **(Bibcode:2011Sci...332...60H, ISSN 0036-8075. PMID 21310967. S2CID 206531385)**

Kim Y, Lee W (2022) Distributed Raman spectrum data augmentation system using federated learning with deep generative models. Sensors 22:9900. https://doi.org/10.3390/s22249900

Li X, Zhang T (2017) An exploration on artificial intelligence application: from security, privacy and ethic perspective. In: 2017 IEEE 2nd international conference on cloud computing and big data analysis (ICCCBDA). IEEE, pp 416–420

Lim D (2019) Killer robots and human dignity. https://www.aies-conference.com/2019/wp-content/papers/main/AIES-19_paper_6.pdf. Accessed 2 Jun 2023

Lv Y (2023) AI+ CASE lab: advanced interdisciplinary research and education lab for connected, autonomous, shared, and green transportation systems [its research lab]. IEEE Intell Transp Syst Mag 15(3):158-C3

Makridakis S (2017) The forthcoming artificial intelligence (AI) revolution: its impact on society and firms. Futures 90(90):46–60. https://doi.org/10.1016/j.futures.2017.03.006

Meissner G (2020) Artificial intelligence: consciousness and conscience. AI Soc 35:225–235. https://doi.org/10.1007/s00146-019-00880-4

Michael JB (2021) Security and privacy for edge artificial intelligence. IEEE Secur Priv 19(4):4–7. https://doi.org/10.1109/msec.2021.3078304

Müller VC (2020) Ethics of artificial intelligence and robotics. In: Zalta EN (ed). Stanford encyclopedia of philosophy; Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/entries/ethics-ai/. Accessed 2 Jun 2023

Mutz J, Javadi A-H (2017) Exploring the neural correlates of dream phenomenology and altered states of consciousness during sleep. Neurosci Conscious. https://doi.org/10.1093/nc/nix009

Neri H (2020) The risk perception of artificial intelligence, Lexington Books. ProQuest Ebook Central

Pham MT (2007) Emotion and rationality: a critical review and interpretation of empirical evidence. Rev Gen Psychol 11(2):155–178. https://doi.org/10.1037/1089-2680.11.2.155

Poola I (2017) How artificial intelligence in impacting real life everyday. Int J Adv Res Dev 2(10):96–100

Procter R, Tolmie P, Rouncefield M (2023) Holding AI to account: challenges for the delivery of trustworthy AI in healthcare. ACM Trans Comput Hum Interact 30(2):1–34

Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sutskever I (2021) Learning transferable visual models from natural language supervision. In: International conference on machine learning. PMLR, pp 8748–8763

Ruschemeier H (2023) AI as a challenge for legal regulation–the scope of application of the artificial intelligence act proposal. In: ERA Forum, vol 23, no 3. Springer, Berlin, pp 361–376

Sandberg A (2013) Feasibility of whole brain emulation. In: Philosophy and theory of artificial intelligence. Springer, Berlin, pp 251–264

Selvaraj DC, Hegde S, Amati N et al (2023) A deep reinforcement learning approach for efficient. Safe Comf Driv Appl Sci 13:5272. https://doi.org/10.3390/app13095272

Signorelli CM (2018) Can computers become conscious and overcome humans? Front Robot A I:5. https://doi.org/10.3389/frobt.2018.00121

Sohn E (2019) Decoding the neuroscience of consciousness. Nature 571(7766):S2–S5. https://doi.org/10.1038/d41586-019-02207-1

Sparkes M (2022) No sign of a machine mind yet. New Sci 254(3391):9. https://doi.org/10.1016/S0262-4079(22)01039-9

Syropoulos A (2017) Demystifying computation: a hands-on introduction. World Scientific, Singapore

The Guardian (2014) https://www.theguardian.com/technology/2014/oct/27/elon-musk-artificial-intelligence-ai-biggest-existential-threat

Tóth Z, Caruana R, Gruber T et al (2022) The dawn of the AI robots: towards a new framework of AI robot accountability. J Bus Ethics 178:895–916. https://doi.org/10.1007/s10551-022-05050-z

Turing AM (1950) Computing machinery and intelligence. Mind LIX(236):433–460. https://doi.org/10.1093/mind/lix.236.433

Väänänen K, Sankaran S, Lopez MG, Zhang C (2021) Editorial: respecting human autonomy through human-centered AI. Front Artif Intell 4:807566. https://doi.org/10.3389/frai.2021.807566

Velmans M (2009) Understanding consciousness. Routledge, Milton Park

Walker EH (1970) The nature of consciousness. Math Biosci 7(1–2):131–178

WarnerBros.com|2001: A Space Odyssey|Movies. (1968). Warner Bros. https://www.warnerbros.com/movies/2001-space-odyssey. Accessed 2 Jun 2023

Wider KV (2018) The bodily nature of consciousness. In: The bodily nature of consciousness. Cornell University Press

Windt JM (2020) How deep is the rift between conscious states in sleep and wakefulness? Spontaneous experience over the sleep–wake cycle. Philos Trans R Soc B Biol Sci 376(1817):20190696. https://doi.org/10.1098/rstb.2019.0696

Yarlagadda RT (2015) Future of robots, AI and automation in the United States. IEJRD Int Multidiscip J 1(5):6

Zeman A, Coebergh J (2013) The nature of consciousness. Ethical Legal Issues Neurol 118:373–407

Zhu J, Huang T, Chen W, Gao W (2018) The future of artificial intelligence in China. Commun ACM 61(11):44–45

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.