2021 ASEE ANNUAL CONFERENCE
Virtual Meeting | July 26–29, 2021 | Pacific Daylight Time
Paper ID #33915

Artificial Intelligence Paradigms and the Future of Learning: What a Partial Review of Half a Century of AI Conceptualization Suggests

Joseph Maloba Makokha, Stanford University

Joseph Makokha was born, raised and educated in Kenya. He obtained a BSEE degree from the University of Nairobi before moving to the United States, where he earned two masters degrees in education before starting his doctoral studies in mechanical engineering at Stanford University focussing on design. He researches human collaboration with artificial intelligence (AI), with the goal of understanding how to design AI that augments humans on thinking tasks.

Artificial Intelligence Paradigms And The Future Of Learning - What A Partial Review Of AI Conceptualization Over Half A Century Suggests

introduction

Over the past half century, artificial intelligence (AI) research and practice has delivered successes and setbacks that when closely examined, reveal strengths and weaknesses in the patterns of conceptual frameworks applied in designing AI applications. An example of such a framework is “AI as a human tutor”, which has influenced AI for learning over the past decades. In their review of evolution and revolution in Artificial Intelligence In Education (AIED), Roll & Wylie[1] highlight this metaphor, and suggest that it has “run its course” as a useful paradigm. In terms of our understanding of AI, it has been difficult to conduct a holistic examination of the disciplines that comprise the AI space in the past, due to differences in methods specific to the fields. However, recent advances in tools, as well as increased collaborations among diverse practitioners from the humanities, bioengineering, computer science and others is leading to better ways of exploring this space. Why does AI matter in engineering education? First, we begin with the premise that AI can support learning activity during the problem-solving stage by introducing “surprise”, which has been identified as an important trigger for framing, re-framing and redefining the problem[2]. Together with ambiguity, this “surprise” keeps the practitioner (or learner) from routine behavior - which is likely to lead the learner to known, predictable solutions - and creates possibilities for novelty as they take the reframed problem and experiment with it to discover what consequences and implications can be derived. Secondly, trust in AI is gaining attention as many applications come into widespread public use, with recent audits revealing biases[3] that mostly emanate from inadequate design frameworks, in addition to other complex reasons. It is therefore imperative to develop appropriate frameworks that support learning while also mitigating human biases that have been shown to erode trust in AI[4].

Using research and intellectual property publications from the first and last decades of the past fifty years (these two decades present the most notable developments relative to other decades), we examine AI paradigms and applications that arise from them, and relate these to the kinds of problems they solve in the real world. We then develop a framework for guiding the mapping of specific tasks assigned to AI with the best conceptual models, capabilities and roles; for situations where humans collaborate or interact with AI. This framework also considers the ways we currently train machines (using examples from humans), and takes trust and bias into consideration, as these are critical components in ensuring widely acceptable AI applications for education and other domains. This framework will inform the design of effective AI applications, leading to improved attitudes towards machines that may take the place of humans, or work besides them in diverse environments including engineering education settings.

human-inspired artificial intelligence

For generations, the way humans think has inspired engineers, tinkerers, researchers and enthusiasts to attempt building "thinking" machines. Such machines tend to mimic how people learn and how they apply the knowledge to act in the world, hence an understanding of ways people learn has been central to their development. How do people learn? This question has endured among educators, and continues to inspire research to date. Table 1 below shows examples of learning theories applied in the past century. From behaviorist (Maslov[5], Skinner[6]) to psychosocial (Piaget[7], Vygotsky[8]) to cognitive (Erik Erickson[9]) to situated (Wenger & Lave[10], Schon[2]) to information processing (Atkinson & Shiffrin [11]) among others, learning theories have shaped the structure, content and experience around instruction for students and teachers over the years.

Proponent	How people learn
Maslov[5], Skinner[6]	Behaviorist - emphasizing the role of stimulus and response
Piaget[7], Vygotsky[8]	Cognitive - biology and environment lead to progressive reorganization of mental processes
Erickson[9]	Psychosocial - biological and sociocultural factors influence learning
Wenger & Lave[10], Schon[2]	Situated - social engagements provide proper context that helps learning
Atkinson & Shiffrin[11]	Information processing - storage, retrieval and decision making in understanding learning

Table 1. A sample of theories on learning

This aspiration to build thinking machines, together with paradigms on artificial intelligence have shaped the tools and applications created over the past decades. As new theories and knowledge are developed, new paradigms have also emerged as seen from the pattern of shifts between the 1970's and 2010's.

using paradigms to understand AI's evolution

Practitioners in diverse fields define the term "paradigm" in different ways depending on their domains, with slight variations corresponding to norms in their respective fields. We take Kuhn's[12] view which holds that a paradigm provides an open-ended resource that presents a framework of concepts, results and procedures within which subsequent work is structured. A characteristic of paradigms is that they can "shift" with new knowledge or evidence. An example using human flight experience can be represented as shown in Table 2 below. The inspiration may have originated from nature, through birds' ability to swiftly move in air. Legends and mythology from early Greek times suggest that humans aspired to fly like birds[13], but we easily identify this to be a limited though interesting way to enable human transportation. The preceding paradigm shift - possibly from the advent of the steam engine - provides an important conceptualization of powering a machine to transport a human being that also controls it. Mass, supersonic travel, such as enabled by the Concorde represents one form of actualization of this human flight idea. Paradigms therefore present us a lens through which we view the concepts in publications and intellectual property that are covered in this paper.

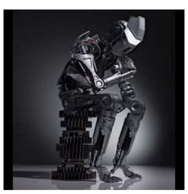
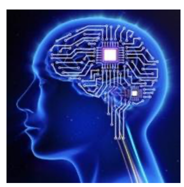
Inspiration	**Paradigm Shift**	**Recent Implementation**
	 Credit: Science Photo Library	 Credit: Associated Press
 Credit: Phillip Toledano	 Credit: Stan Garfield	 Credit: University of Bern

Table 2. Inspiration, paradigm shift, recent implementation of flight and thinking machines

Turning to AI, the desire to create machines with human thinking abilities (which goes back centuries - to Leibniz in the 17th century, or even Aristotle in 4th century BC)[14] can be mapped in a similar way as it evolves from robotic embodiment of the machine; to a focus on cognition; and onto specific cognitive tasks that augment or completely replace humans. The medical diagnostics implementations such as in cancer diagnosis from radiology images[15], where machines perform representative classification tasks like diagnosing a suspicious object as either benign or malignant, is one of many advances in today's AI applications. In examining the recent research and practice of AI, it is evident that paradigms have shifted as a result of several factors including the following:

- fast, affordable computation capabilities
- recognition of the need for inter-disciplinary domain knowledge

- increased opportunities for implementing AI across disciplines, including its commercialization

First, a sample of papers related to AI from the period leading to and during the 1970's. Table 3 below represents summaries of what abilities the authors ascribed to AI, as well as how each ability relates with humans on tasks (replacing humans, augmenting humans, or unsuitable for the task).

BASIS OF THEIR PROPOSAL	RELATIONSHIP TO HUMANS	AUTHORS
Exhibit intuition, insight and learning. Machines that think, that learn, that create … the range of problems which they can handle will be coexistive with the range to which the human mind has been applied.	REPLACE	NEWELL, SHAW, SIMON[16]
Programming a robot with an integrated suite comprising planning systems, models of the world and sensory processing systems enables it to successfully accomplish tasks in the real world settings.	REPLACE	Nilsson, N. J.[17]
Role of AI cased instruction is to make possible a new kind of learning environment, offering instruction in a manner sensitive to learner's strengths, weaknesses and preferred learning style.	REPLACE	Simon, H.A., & Kadane, J.[18]
AI is based on the ability to use vast amounts of diverse knowledge in procedural ways, as opposed to having a few general and uniform principles.	REPLACE	Goldstein, I., & Papert, S.[19]
Cognitive models from human beings are unsuitable for designing artificial intelligence. That AI programs are in fact problems, not solutions to problems.	UNSUITABLE	McDermott, D.[20]
The human mind is an analytic machine	REPLACE	HOBBES, KANT, HUSSERL[16]
Efficient search is key to solving complex problems by using AI, though we need better models to accomplish this.	REPLACE	Pohl, I.[21]
Young learners gain from programming computers using LEGO language, and are able to apply the concepts in learning math.	AUGMENT	Feurzeig, W., Papert, S., Bloom, M., Grant, R., & Solomon, C.[22]
Intelligence is the outcome of receiving and processing information - ie feedback systems are key to intelligence	REPLACE	Norbert Wiener[23]
There are certain kinds of problems that machines are unable to solve due to the ill-defined nature of problems (ill-structured problem).	RELIES ON	Simon, H. A.[24]
Differences between objects of perception (the physical environment) and objects of cognition (mental representations) contributes to errors in representing non-linguistic (symbolic) information in AI.	REPLACE	Pylyshyn, Z. W.[25]

Intelligence is congruent to formal rule-following, and knowledge is entirely the internal representations of reality (ie the mind works by computing algorithmic rules on discrete symbolic representations)	REPLACE	NEWELL & SIMON[26]
While important, it is difficult to construct belief systems models in AI, arising from the differences between belief and knowledge systems.	REPLACE	Ableson, R. P.[27]
A major methodology of AI science is building precise models of cognitive theories, while its fundamental goal is to understand intelligent processes independent of their physical realization.	AUGMENT	Goldstein, I.[28]
By utilizing an information network, we can create AI capable of answering new questions, thereby going beyond pre-programmed questions and anticipated answers provided in advance.	REPLACE	Carbonell, J. R.[29]
By acquiring enough rules, we can create an effective AI tutor that helps physicians, capable of giving the reasons for decisions it makes.	AUGMENT	Shortliffe, E. H., Axline, S. G.[30]

Table 3. A sample of attitudes on AI from the period around the 1970's

Table 3 features some of the attitudes taken by AI researchers and practitioners that will be used in reviewing the applications developed during that decade (1970s), and to highlight overarching design influences and paradigms. While conceptions of machines that can perform intelligent tasks goes back to Descartes and Leibniz, the table covers relatively recent years. A distinction arises regarding artificial general intelligence (AGI) and artificial narrow intelligence (ANI) with a realization that advances in AGI are not advances in ANI. Regarding human capabilities, Dreyfus[31] argued that "computers, who have no body, no childhood and no cultural practice, could not acquire intelligence at all" since human knowledge is tacit and cannot be articulated and incorporated in a computer. This argument held within previous paradigms of AI, but has been obsoleted by Big Data (where mathematical methods are applied to huge amounts of data to find correlations and infer probabilities) and Deep Learning.

Recognizing that there are limits in applying computational techniques to mimic human thinking, AI practitioners and researchers have attempted to advance the field by trying different strategies like breaking down the thinking process into conceptually manageable units. In "What Computers Can't Do"[31], Dreyfus splits AI into Cognitive Simulation (CS) and Artificial Intelligence (AI), which he says leads to two separate but related questions:

- does a human, in "processing information" actually follow formal rules like a digital computer?
- can human behavior, no matter how generated be described in a formal way that can be manipulated by a computer?

The answers that AI designers give to these questions reveal paradigms upon which subsequent applications they develop are likely to be based.

the patents from around the 1970's

Table 4 shows some of the patents with corresponding simplified summaries

TITLE, (Date Filed)	REF NO.	NOTES [Class G06N3/02 Computer systems based on biological models using neural network models]
Syntactic word organizer (1977)	US4156868A	Bell Labs(USA). Word sequencing prediction based on acoustic correspondence with a predetermined sequence.
Arrays of machines such as computers (1978)	US4247892A	P.N.L.(USA). Multi-dimensional arrays of computers for enhanced processing capability - for neural network simulation
Information processing apparatus (1963)	US3310784A	RCA Corp.(USA). Information processing using neural logic functions
Automat with customized intelligence (1978)	DD145436A1	E. Liss(Germany). Understanding or recognizing input information via associative thinking to generate motor control information.
Device for simulating adaptive neuron (1977)	SU708368A1	C.H.(Russia). A device for modeling an adaptive neuron - to model living organisms and subsystems
Device for evaluating psychophysiological characteristics of control system operator (1978)	SU1003130A1	Bekhterev et al. (Russia). Improve biotechnical control of humans operating with machines and hybrid intelligence.
Variable image display apparatus (1978)	GB1605135A	S. of I.(Britain) Producing a face of an unknown person by variable image display - by combining drawings or photographs of facial features selected from a kit.
System and method for increasing memory performance (1970)	US4954951A	H.G.P.(1970) Efficient, rapid solution of high order multivariable polynomials that produce high level AI.
Neuron Information Processing Apparatus (1963)	US3310783A	RCA Corp.(USA). Information processing apparatus for handling information by simulated neural processes.
Signal vector recognition system (1971)	US3727193A	V. Bolie (USA). Automatic categorization of an N-element stimulus pattern by determining degree of similarity between the stimulus vector and various reference vectors based on orthogonality.

Table 4. Selected Patents From 1970's With Summary Descriptions

what publications and patents from the 1970's reveal

These patents from the 1970's (Table 4) reveal an emphasis on computational techniques, with little human element despite the "biological modeling using neural networks" category under which they fall within the patent classification system. They feature neural networks, prediction, patterns, polynomials, associative thinking and other computational aspects. These patents are viewed together with the publications from the same decade (Table 3) on researchers' and practitioners' attitudes towards AI, in order to establish relationships with the applications that emerged around that period.

Among the early work on AI, Newell and Simon[[21]] conceived the Logic Theorist (LT), a program that proved theorems by applying appropriate rules of thumb or heuristics in conducting a search through a "search tree" (which is a representation of the options and end solutions). New advances followed shortly such as those by Weizenbaum[[32]] of MIT, who created Eliza[[32]], a programming language that enables natural language conversation with a computer via keyboard input. One can see the influence of paradigms such as intelligence being congruent to "formal rule following", which ties into the premise that "enough" rules can enable the creation of an effective tutor, or that an information network can help AI answer new questions, going beyond pre-programmed questions and answers.

Publications highlight the differences between human thinking and pure rule-following, such as McDermott[[20]] who argues that cognitive models from humans are unsuitable for designing AI. Other examples are "ill-structured" problems which while computational in nature, present impossibilities to machines - but humans may use intuition to give an answer. Constructing belief systems models also falls under tasks that are difficult for machines. In his book "Computer Power and Human Reason", Weizenbaum[[33]] makes a distinction between machine power and human reasoning in terms of choice and decision, the former being innately human while the later can be purely computational. Hence computation power is different from human reason - comprising Aristotle's prudence (making right decisions in concrete situations) and wisdom (ability to see the whole). In concurring with Weizenbaum, the mathematician Roger Penrose[[34]][[35]] argues that human thinking is not algorithmic and cannot therefore be simply modelled by machine-generated code. Tensions surrounding capabilities of "thinking machines' are also evident from early work by Dreyfus, whose book "What Computers Can't Do in 1972"[[31]] explores the capabilities of AI relative to humans at the time, concluding that machines are ill-equipped to perform cognitive tasks easily done by humans. While some of these limits have been overcome by recent technological advances, those relating to an understanding of the physical world remain unresolved. Influence from these attitudes are evident in the AI paradigms at the time, which emphasized rule-based search through "trees" of predetermined branches that led to final results. A notable feature of these early AI applications is that one could retroactively explain the decision-making that yielded a given solution, unlike current machine learning (ML) based versions that operate within a "black-box"[36] that is difficult to explain, leading to some of today's distrust of AI.

Over the past half century, there have been concerted efforts to develop "thinking" machines that work alongside, or completely replace people on tasks previously considered a preserve of humans. Following the demonstration of computers' capability to manipulate symbols - a step further from numeric manipulation, some AI researchers developed rule-based, algorithmic systems to achieve "intelligence". Building on this symbolic manipulation capabilities, Newell, Simon and Shaw (1958) demonstrated a version of a thinking machine named the Logic Theorist (LT) while noting that their subjects tended to use rules or shortcuts that were not universally correct (humanoid heuristics), and which often helped, even if they sometimes failed[37] [33, p. 6]. They believed that this theory of heuristics (rather than algorithmic) programming would enable them to program computers to exhibit intuition, insight and learning[33, p. 7]. Key aspects of these efforts and their relative successes or failures can be attributed to their mental models, paradigms and design philosophies at the time - the basis upon which the designs of these "thinking" machines were based. The Logic Theorist generated interest and spurred some developments in pursuit of AI. While a promising development, this approach had its limitations partly due to the way humans handle problem solving. In discussing capabilities of computers, Dreyfus'[31] points out one important shortcoming regarding machines' lack of real world experience gained by humans as they grow up starting from childhood. There is no way to imbue such learned experiences into a machine. Next, we present in Table 6 below some recent publications from the 2010's.

publications from the 2010's decade

Table 4 shows some of the patents with corresponding simplified summaries

BASIS OF THEIR PROPOSAL	RELATIONSHIP TO HUMANS	AUTHORS
Proposes an effective computer support tool applying efficient knowledge representation schemes, that helps the product designer make better informed decisions by enabling them to sift through vast quantities of raw data available to the designer.	AUGMENT	Chandrasegaran et al[38]
It is effective to apply predictive modeling concepts relevant to cardiology such as feature selection and frequent pitfalls such as improper dichotomization; common algorithms used in supervised learning and reviews selected applications in cardiology and related disciplines; advent of deep learning and how these methods could be applied to enable precision cardiology and improve patient outcomes.	AUGMENT	Johnson et. al[39]
In government, AI can reduce administrative burdens, help resolve resource allocation problems, and take on significantly complex	AUGMENT	Mehr, H., Ash, H., & Fellow, D.[40]

tasks. Examples are answering questions, filling out and searching documents, routing requests, translation, and drafting documents. Can make government work more efficient while freeing up time for employees to build better relationships with citizens, improving citizen engagement and service delivery.		
Future generations of Wearable Internet of Things (WIoT) promise to transform the healthcare sector, wherein individuals are seamlessly tracked by wearable sensors for personalized health and wellness information—body vital parameters, physical activity, behaviors, and other critical parameters impacting quality of daily life.	REPLACE	Hiremath, S., Yang, G., & Mankodiya, K.[41]
New technologies may enhance the traditional aims of journalism, or may initiate greater interaction between journalists and information and communication technology (ICT) specialists. The enhanced use of computing in news production is related in particular to three factors: larger government datasets becoming more widely available; the increasingly sophisticated and ubiquitous nature of software; and the developing digital economy. Creates foundations for original investigative journalism, increase the scope for new forms of interaction with readers.	AUGMENT	Flew, T., Spurgeon, C., Daniel, A., & Swift, A.[42]
Today, AI has advanced to a stage where on many cognition related tasks it can match and even surpass the performance of humans. But AI is also now achieving extremely high efficiency in practical applications such as speech and object recognition, self-driving cars, intelligent tutoring systems, efficient decision support systems, and in the capacity to detect patterns in Big Data and in constructing accurate models of social behavior. Thus, for the first time in history, we must ask ourselves: "has our monopoly on intelligence, however defined, been challenged?"	REPLACE	Nowak, A., Lukowicz, P., & Horodecki, P.[43]
Artificial Intelligence of the next generation needs to interact with users socially, convincing them in its ability to understand human minds, including emotions. For this to happen, an artificial emotional intelligence is needed, capable of adequate, believable behavior in social emotional interactions, endowing it with fluent describing, in addition to appraisals, somatic markers, feelings, emotions, moods, emotional reactions and biases. Key building blocks that integrate them are moral schemas and semantic maps. The model describes interaction of three factors: plans and commitments, moral and ethical values, and somatic comfort.	REPLACE	Samsonovich, A. V.[44]
Technology is not simply a tool for human intention. It is an actor in the cognitive ecology of immersive humans-with-technology environments. There is fruitful overlap between Artificial Intelligence, Computational Technologies and Mechanical Engineering that is of value to consider in mathematics education.	AUGMENT	Gadanidis, G.[45]
Based on analysis of reported failures of artificially intelligent systems to extrapolate to future AIs, it is likely that both the frequency and the seriousness of future AI failures will steadily increase. A single failure of a super intelligent system may cause a catastrophic event without a chance for recovery. The goal of	REPLACE	Yampolskiy, Roman V., and M. S. Spellchecker[46]

cybersecurity is to reduce the number of successful attacks on the system; the goal of AI Safety is to make sure zero attacks succeed in bypassing the safety mechanisms.		
AI can be used to address many challenges facing America's healthcare system - from disease detection to building predictive models for treatment - thereby improving the quality and lowering the cost of patient care.	AUGMENT	D. B. Neill[47]
The broad use of machine learning makes it important to understand the extent to which machine-learning algorithms are subject to attack, particularly when used in applications where physical security or safety is at risk. We investigate a novel class of attacks on facial biometric systems: attacks that are physically realizable and inconspicuous, that allow an attacker to evade recognition or impersonate another individual through printing a pair of eyeglass frames.	REPLACE	Sharif, M., Bhagavatula, S., Bauer, L., & Reiter, M. K.[48]
Many tasks in robotics involve interactions with physical environments and objects. One of the fundamental components of such interactions is understanding the correlation and causality between actions of an agent and the changes of the environment as a result of the action. Since the 1970s, there have been various attempts to build a system that can understand such relationships. Recently, with the rise of deep learning models, learning-based approaches have gained wide popularity.	REPLACE	Yuke Zhu, Roozbeh Mottaghi, Eric Kolve, Joseph J Lim, Abhinav Gupta, Li Fei-Fei, Ali Farhadi[49]
Recent work has shown that deep neural networks are highly sensitive to tiny perturbations of input images, giving rise to adversarial examples. Though this property is usually considered a weakness of learned models, we explore whether it can be beneficial. We find that neural networks can learn to use invisible perturbations to encode a rich amount of useful information. In fact, one can exploit this capability for the task of data hiding. We jointly train encoder and decoder networks, where given an input message and cover image, the encoder produces a visually indistinguishable encoded image, from which the decoder can recover the original message. We show that these encodings are competitive with existing data hiding algorithms, and further that they can be made robust to noise.	REPLACE	Jiren Zhu, Russell Kaplan, Justin Johnson, Li Fei-Fei[50]

Table 6. A selection of publications from the recent decade (2010’s)

the patents from around the 2010's

Table 7 shows some of the patents with corresponding simplified summaries

TITLE, (Date Filed)	REF NO.	NOTES [Classes G06N3, G06Q10, H01L27, G06N5, G06K19]
Self learning adaptive modeling system (2012)	US9576262B2	Best models are chosen from a set of predictive candidate models automatically built from accumulated data and distributed across multiple levels, based on comparison of performance metrics and models. The best model can then be activated for use in making predictions
Neural network data entry system (2017)	US10789529B2	System that receives context text items from a user, and has a predictor trained to predicts the next item.
Providing Financial-Related, Blockchain-Associated Cognitive Insights Using Blockchains (2016)	US20180165758A1	Providing the financial-related, blockchain-associated cognitive insight from a plurality of data sources comprising financial related data sources and blockchain data sources;
Artificial intelligence (AI) based chatbot creation and communication system (2016)	US20190392285A1	A system for creating and managing an artificial conversational entity using an artificial intelligence (AI) based communications system.
Applications of back-end-of-line (beol) capacitors in compute-in-memory (cim) circuits (2018)	US20190138893A1	The apparatus includes a compute-in-memory (CIM) circuit for implementing a neural network disposed on a semiconductor chip. The CIM circuit includes a mathematical computation circuit coupled to a memory array. The memory array includes an embedded dynamic random access memory (eDRAM) memory array.
Collaboration of audio sensors for geo-location and continuous tracking of health conditions for users in a device-independent artificial intelligence (AI) environment (2017)	US10846599B2	Uses Audio sensors collaborate for geo-location and tracking of health conditions for multiple users, independently geo-located and tracking different users within the AI environment. Sends a command for an AI action, such as contacting a hospital server or adapting to monitor a suspected health condition to classification of a health event of concern, in view of the estimated location.
Artificial intelligence (AI) electrocardiogram (ECG) (2018)	US10729351B1	Receives impulses from a heartbeat, which are compared to a database of subwaveforms and discontinuity points for normal and abnormal patients then outputs an average data value of detection for each interval
Partitioned artificial intelligence for networked games (2012)	US9327194B2	An exemplary system splits the AI into a computationally lightweight server-side component and a computationally intensive client-side component to harness the aggregate computational power of numerous gaming clients.
Methods and apparatus for spiking neural computation (2012)	US9367797B2	Provides methods and apparatus for spiking neural computation of general linear systems. One example aspect is a neuron model that codes information in the relative timing between spikes.

Compositions and methods for diagnosis and treatment of pervasive developmental disorder (2013)	US20150023949A1	A method of assessing whether a subject is afflicted with a pervasive developmental disorder, is predisposed to developing a pervasive developmental disorder, disposition or severity of a pervasive developmental disorder.

Table 7. A Selection Of US Patents From 2010's With Summary Descriptions

what publications and patents from the 2010's reveal

The sample of publications and patents from the 2010's demonstrate a major shift compared to those from the 1970's. Among the most outstanding are:

- a paradigm shift towards teaching machines to predict the future - machine learning, (ML) - seems to displace rule-based AI
- we see a larger variety of contexts in ways AI is applied, such as finance, medicine, security, government, at home and others
- there is a stronger focus on specific implementation of AI in forms that can readily be tested - in radiology, telecoms, general information processing among others
- more patents seem to describe contemporary devices and applications, as opposed to futuristic ones as in the 1970's, like the chatbot and the ECG medical device
- we see more compelling applications that augment humans compared with the 1970's

Let us now examine how AI applications resulting from these shifts have fared relative to those from the 1970's.

From a historical view, a new approach to solving AI using Artificial Neural Networks (ANN) arose as a paradigm in the 1980's, applying our nervous system and brain as a model - as opposed to manipulation of symbols. These ANNs can learn without receiving explicit instructions, a fact that obsoleted Dreyfus'[[31]] argument regarding learning limitations of machines. Subsequently, there arose a new argument for the big data approach – one that goes to the extend of suggesting that we do not have to create AI that thinks like humans, rather we may change our thinking to be like that of computers, in the sense that they find correlations and use this to predict the future. Implicitly, this is the message of Viktor Mayer-Schönberger and Kenneth Cukier's[[51]] Big Data exposition that describes ways in which insights that cannot be derived from a small scale become possible with large data. Notably, this paradigm contrasts causality (old scientific method) with correlation in the sense that predictions are made purely out of correlation patterns that are fine tuned by AI, as opposed to seeking causes of observed phenomena as is common practice in science. This is a major shift in how outcomes are defined, and leads to a new way of optimizing AI by focusing on finding correlations as opposed to causes. It naturally follows that the best models will arise from large quantities of "good" data, the qualification being data that is rich in relevant features associated with the outcome - a simple example being age in relation to physical abilities. In summary, the paradigms represented in the 2010's lean heavily on (unpredictable) machine learning as opposed to searching for solutions based on (predictable) rules.

Early paradigms set out to demonstrate AI capabilities by solving puzzles, and chess was a suitable game for such a task. Since the (Newell and Simon's) Logic Theorist proposal in the 1970's, there had been expectations of machines beating humans on strategy games, then Watson beat the two best Jeopardy players, Ken Jennings and Brad Rutter[52] in 2011. This was after Deep Blue - a forerunner of 1960's MIT and Moscow's Institute of Theoretical and Experimental Physics (ITEP) chess programs - defeated Garri Kasparov[53] in chess. Most recently, Watson has so far turned out AI assistants that perform well in routine tasks in oncology, genomics and clinical trials matching (Strickland, 2019)[54]. AlphGo beat Le Sedol in 2016[55] - a milestone since it used deep reinforcement learning (artificial neural network, ANN based). This demonstrated that computers can handle tacit knowledge, obsoleting Dreyfus' argument (or not, since this "tacit " knowledge is restricted to the idealized world of science, quite different from the human world envisioned by Dreyfus).

Puzzles and games aside, AI has faced several challenges due to the nature of machines - they lack the experience of growing up in the real world, hence do not understand causality. Based on the new ML-centric paradigms that have been prominent in the 2010's, the question arises as to whether causality still matters, or correlation is enough to predict the future with certainty. Researchers like Pearl and Mckenzie[56], who coined the "mini" Turing Test[57] think that in order to create a humanlike intelligence, the computer must be able to master causality. They are motivated by the question "how can machines and people represent knowledge in a way that enables them to access the necessary information swiftly, answer questions correctly, and do it with ease, as a three-year-old child can?" Not surprisingly, they are inspired by a paradigm proposed by Laplace[58], who saw a similarity between a single molecule and the universe, inferring that if we had complete knowledge of the state of the universe at one time, we could in principle determine the state at any previous or successive time[58, p. 6] – a direct result from cause and effect standpoint. This causality aspect provides some insights on possible design characteristics of an effective AI framework.

It appears that the shifts in envisioning and implementation of AI - evident in both publications and patents from the 2010's - have led to applications that apply a different model for gaining knowledge, that is, machine learning. In the recent decades, there arose deep learning neural networks, consisting layers of artificial neurons such that to identify a cat for example, the first layer identifies pixels as light or dark, next layer edges and simple shapes, next one more complex shapes and objects and a fourth one may learn which shapes can be used to identify an object[59]. These have accelerated learning tools for AI systems in image recognition, given the corresponding improvements in computing capabilities. When it comes to decisions, ANNs are able to recalibrate and compute probabilities of future events with high certainty as long as they have received enough “good” data as mentioned above. Notably, this training data is also responsible for the bias aspect that is covered below.

bias in AI – the challenge of machines learning from humans

It is evident that researchers and practitioners have achieved considerable success in creating “thinking machines”. Recent machine learning methods, together with powerful processors have contributed to this success. However, as AI gains mainstream status, some audits reveal problems including “opaqueness” and bias in these systems. This raises an important and related aspect: the concept of agency. Langley[60] highlights two forms of agency, “explainable” and “normative” agency, the former being when AI agents can explain what they have done in ways humans can understand; while “normative agency” stems from agents demonstrating that they follow the norms and implicit societal rules when pursuing explicit goals. He points out one critical requirement in order for AI to gain widespread acceptance - that it must be able to explain itself in a convincing way that it shares our aims. Subsequently, this relates directly to bias and trust in AI systems, since humans are likely to have confidence in applications that explain themselves and demonstrate shared goals with the humans. Noting that humans operate in a complex world, Langley states that while AI can address many issues on legal, moral and normative reasoning, we need better integrated systems that can operate in complex, sometimes ambiguous situations found in the real world. Based on the observations made so far from the past half century, this will likely require a paradigm shift from current practices where we train machines with many examples, then expect them to find correct predictions to future questions, based on the examples.

Another related challenge arises from the very feature that enables modern AI to perform so well on some tasks – correlation. This is because the focus on correlation minimizes or even disregards tacit knowledge, the concept formulated by Polanyi[61], that “..we can know more than we can tell”[61, p. 4]. Importantly, he posits that skills are a prerequisite for knowledge in general, and scientific knowledge in particular - and that skills cannot be learned from text books but from someone who knows the trade. The immediate implication is that machines are unable to learn in the same way that humans do - as they do not grow up around people knowledgeable in trades such as doctors, teachers or engineers. While this is mitigated by machine learning methods that employ big data such as those found in the research from the 2010’s (ex. Johnson et. al[38], Nowak et al[43]), reliance on examples for training data inadvertently contributes to the bias problem, as the training data includes existing biases which are directly adapted by AI.

Finally, a major shortcoming of AI is it’s lack of context and situational awareness, two components that humans often understand. As Dreyfus[31] points out, situations that are presented as facts and figures can easily be interpreted in terms of rule-based information processing - making them suitable for solving using computational mechanisms and AI (however, he emphasizes that while problem solving is the goal in this situation, he disputes the assumption that all intelligent behavior is of problem-solving type). On the other hand, humans encounter and solve many problems that are not fact-based nor figure-based. In addition to this, humans gain awareness through a learning process as they solve problems, first as novices, then gaining skills and experience over time. Competence, according to Dreyfus[62] is gained from a person's attitude towards problem solving (which is determined by abilities on the task, ranging from novice to expert), something that machines are incapable of replicating. These abilities are broken down into levels of competence as shown in Table 7.

Skill Level	Components	Perspective	Decision	Commitment
1. Novice	Context-free	None	Analytical	Detached
2. Advanced beginner	Context-free and situational	None	Analytical	Detached
3. Competent	Context-free and situational	Chosen	Analytical	Detached understanding and deciding. Involved in outcome
4. Proficient	Context-free and situational	Experienced	Analytical	Involved understanding. Detached deciding
5. Expert	Context-free and situational	Experienced	Intuitive	Involved

Table 7. Stages of Human Skill Acquisition[62, p. 50]

Notably, humans consider context, which is difficult for machines to do; and AI's reliance on training data injects existing biases into the system, leading to distrust when outcomes reveal these biases[3]. This suggests a need for new paradigms on machine learning.

towards a framework

Today, AI applications are performing so well that people sometimes find it difficult to discern between human and computer-generated art for example - which demonstrates human bias towards AI, that leads to false assumptions about such artifacts. In a recent study by Gangadharbatia[63] (2021), people attributed representational art to humans and abstract art to AI, mostly based on surface features like signs of physical brush strokes – which unbeknown to the testers, can be easily mimicked by AI. This is an example of ways that AI increasingly performs at acceptable levels of desired experiences, delivering human-like outcomes. However, looking at the past half century, the promises of Human-Artificial Intelligence (H-AI) research compared with resulting outcomes in form of applications and tools, there still exist a gap between expectations and actual results among intended users. This arises from several factors, key among them being the assumptions and guiding principles that creators have about users on one hand, and those of users on the other. Publications and intellectual property spanning this period reveals four general paradigms guiding research and development of AI:

- replacing the entire human with AI
- replacing some part of the human with AI
- augmenting the human with AI, and
- keeping AI out of the loop (it is good for nothing!) Due to bias? Poor performance?

These are represented visually in Fig. 1 below.

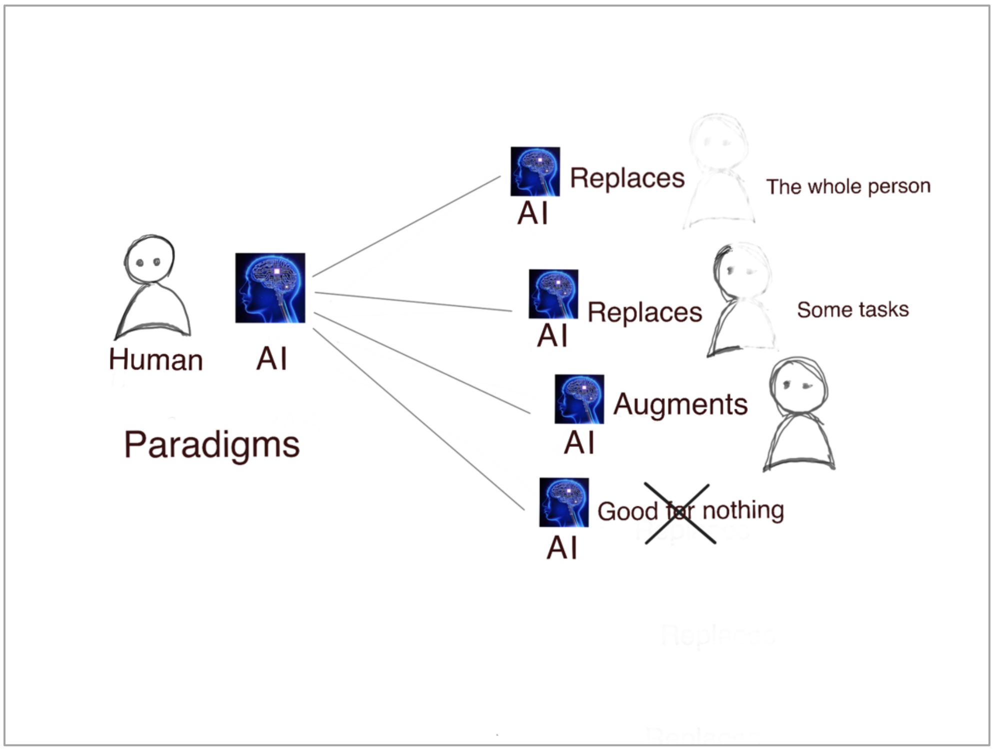


Fig. 1 Paradigms of AI's Role Relative To Humans

Each of these paradigms leads to vastly different conceptualization of the human interaction with AI, affecting how we learn from, with, and by AI; what we trust AI to perform; perceptions of AI among different groups; and how well we view AI's performance relative to our needs, assumptions and expectations. In the human vs AI-generated artwork example above, we can completely replace the human on specific kinds of art. This also applies to some medical diagnostic tasks[56], conversational bots, cars and robots among others. The successes demonstrated from such uses comprise the following characteristics relevant to AI application in engineering education:

- domain understanding of the problem space, including theories, issues and types of available resources such as data
- a determination of whether it makes sense to keep the human in the loop
    - if the human stays in the loop, specify the roles they take on

    - with the human out of the loop, specify an acceptable level of fidelity that the outcome should meet
- clarity in choice of conceptual designs that optimize the available data and computation capabilities in specific use cases

This presents a starting point, but more research is needed in order to inform design of current and future AI applications, as we increasingly narrow the performance gap between humans and machines on common tasks, including learning. The review of publications and intellectual property on AI from the first and last decades of the past half century suggest that while many factors are contributing to the actualization of “thinking machines”, paradigms about AI are a critical in translating AI research into effective, reliable and trustworthy real-world applications for learning, health, automation and other domains.

References

[1] Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. International Journal of Artificial Intelligence in Education, 26(2), p.582-599.

[2] Schön, D. A. (c1983.). The reflective practitioner: How professionals think in action /. Basic Books,.

[3] Osoba, O. A., & Welser IV, W. (2017). *An intelligence in our image: The risks of bias and errors in artificial intelligence*. Rand Corporation., p.14

[4] Hengstler, M., Enkel, E., & Duelli, S. (2016). Applied artificial intelligence and trust—The case of autonomous vehicles and medical assistance devices. *Technological Forecasting and Social Change*, 105, 105–120

[5] Maslow, A. H. (1970). Motivation and personality. New York: Harper & Row.

[6] Skinner, B. F. (1953). Science and human behavior. New York: Macmillan.

[7] Piaget, J. (2013). The child's conception of physical causality. Hoboken: Taylor and Francis.

[8] Vygotsky, L. S. (1997). The collected works of LS Vygotsky: Problems of the theory and history of psychology (Vol. 3). Springer Science & Business Media. p. 30.

[9] Ladd, G. W. (1999). Peer relationships and social competence during early and middle childhood. Annual review of psychology, 50(1), p. 334.

[10] Lave, J., & Wenger, E. (1991). Situated learning: Legitimate peripheral participation. Cambridge university press.

[11] Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In Psychology of learning and motivation (Vol. 2, p. 89-195). Academic Press.

[12] Kuhn, T. S. (1996). The structure of scientific revolutions. 3rd ed Chicago, Ill.: University of Chicago Press.

[13] Graves, Robert. (1957). The Greek Myths. New York: G. Braziller,

[14] https://en.wikipedia.org/wiki/History_of_artificial_intelligence

[15] Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. (2018). Artificial intelligence in radiology. Nature Reviews Cancer, 18(8), 500-510.

[16] Newell, A., Shaw, J. C., & Simon, H. A. (1958). Elements of a theory of human problem solving. *Psychological review*, *65*(3), 151.

[17] Nilsson, N. J. (1969). A mobile automaton: An application of artificial intelligence techniques. Sri International Menlo Park Ca Artificial Intelligence Center.

[18] Simon, H.A., & Kadane, J. (1977). Optimal problem-solving search: All-or-none solutions. Artificial Intelligence, 6(3), 235-247.

[19] Goldstein, I., & Papert, S. (1977). Artificial intelligence, language, and the study of knowledge. Cognitive science, 1(1), 84-123.

[20] McDermott, D. (1976). Artificial intelligence meets natural stupidity. Acm Sigart Bulletin, (57), 4-9.

[21] Pohl, I. (1970). Heuristic search viewed as path finding in a graph. Artificial intelligence, 1(3-4), 193-204.

[22] Feurzeig, W., Papert, S., Bloom, M., Grant, R., & Solomon, C. (1970). Programming-languages as a conceptual framework for teaching mathematics. ACM SIGCUE Outlook, 4(2), 13-17.

[23] Crevier, D. AI : the tumultuous history of the search for artificial intelligence. New York: Basic Books.1993.

[24] Simon, H. A. (1973). The structure of ill structured problems. Artificial intelligence, 4(3-4), 181-201.

[25] Pylyshyn, Z. W. (1978). Imagery and artificial intelligence.

[26] Newell, A., & Simon, H.A. (1961). GPS: A program that simulates human thought. In H. Billings (Ed.), Lernende automaten (p.109-124). Munchen: R. Oldenbourg.

[27] Abelson, R. P. (1979). Differences between belief and knowledge systems. Cognitive science, 3(4), 355-366.

[28] Goldstein, I. (1976). Seymour Papert Marvin Minsky. Artificial Intelligence and Language Comprehension.

[29] Carbonell, J. R. (1970). AI in CAI: An artificial-intelligence approach to computer-assisted instruction. IEEE transactions on man-machine systems, 11(4), 190-202.

[30] Shortliffe, E. H., Axline, S. G., Buchanan, B. G., Merigan, T. C., & Cohen, S. N. (1973). An artificial intelligence program to advise physicians regarding antimicrobial therapy. Computers and Biomedical Research, 6(6), 544-560.

[31] Dreyfus H. L. (1972) What computers can't do. Harper & Row, New York, NY Dreyfus HL

[32] Weizenbaum, J. (1966). ELIZA—a computer program for the study of natural language communication between man and machine. Communications of the ACM, 9(1), 36-45.

[33] Weizenbaum J (1976) Computer power and human reason. Freeman & Company, San Francisco

[34] Penrose R (1989) The emperor's new mind. Concerning computers, minds, and the laws of physics. Oxford University Press, Oxford

[35] Penrose R (1994) Shadows of the mind. A search for the missing science of consciousness. Oxford University Press, Oxford

[36] Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE access, 6, 52138-52160.

[37] Weizenbaum J (1976) Computer power and human reason. Freeman & Company, San Francisco

[38] Chandrasegaran, S. K., Ramani, K., Sriram, R. D., Horváth, I., Bernard, A., Harik, R. F., & Gao, W. (2013). The evolution, challenges, and future of knowledge representation in product design systems. Computer-aided design, 45(2), 204-228.

[39] Johnson, K. W., Torres Soto, J., Glicksberg, B. S., Shameer, K., Miotto, R., Ali, M., ... & Dudley, J. T. (2018). Artificial intelligence in cardiology. Journal of the American College of Cardiology, 71(23), 2668-2679.

[40] Mehr, H., Ash, H., & Fellow, D. (2017). Artificial intelligence for citizen services and government. Ash Cent. Democr. Gov. Innov. Harvard Kennedy Sch., no. August, 1-12.

[41] Hiremath, S., Yang, G., & Mankodiya, K. (2014, November). Wearable Internet of Things: Concept, architectural components and promises for person-centered healthcare. In 2014 4th International Conference on Wireless Mobile Communication and Healthcare-Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH) (p. 304-307). IEEE.

[42] Flew, T., Spurgeon, C., Daniel, A., & Swift, A. (2012). The promise of computational journalism. Journalism Practice, 6(2), 157-171.

[43] Nowak, A., Lukowicz, P., & Horodecki, P. (2018). Assessing Artificial Intelligence for Humanity: Will AI be the Our Biggest Ever Advance? or the Biggest Threat [Opinion]. IEEE Technology and Society Magazine, 37(4), 26-34.

[44] Samsonovich, A. V. (2020). Socially emotional brain-inspired cognitive architecture framework for artificial intelligence. Cognitive Systems Research, 60, 57-76.

[45] Gadanidis, G. (2017). Artificial intelligence, computational thinking, and mathematics education. The International Journal of Information and Learning Technology.

[46] Yampolskiy, Roman V., and M. S. Spellchecker. "Artificial intelligence safety and cybersecurity: A timeline of AI failures." arXiv preprint arXiv:1610.07997 (2016).

[47] D. B. Neill, "Using Artificial Intelligence to Improve Hospital Inpatient Care," in IEEE Intelligent Systems, vol. 28, no. 2, p. 92-95, March-April 2013, doi: 10.1109/MIS.2013.51.

[48] Sharif, M., Bhagavatula, S., Bauer, L., & Reiter, M. K. (2016, October). Accessorize to a crime: Real and stealthy attacks on state-of-the-art face recognition. In Proceedings of the 2016 acm sigsac conference on computer and communications security (p. 1528-1540).

[49]. Zhu, Y., Mottaghi, R., Kolve, E., Lim, J. J., Gupta, A., Fei-Fei, L., & Farhadi, A. (2017, May). Target-driven visual navigation in indoor scenes using deep reinforcement learning. In *2017 IEEE international conference on robotics and automation (ICRA)* (pp. 3357-3364). IEEE

[50] Zhu, J., Kaplan, R., Johnson, J., & Fei-Fei, L. (2018). Hidden: Hiding data with deep networks. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 657-672).

[51] Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. Boston: Houghton Mifflin Harcourt.

[52] https://www.theguardian.com/technology/2011/feb/17/ibm-computer-watson-wins-jeopardy

[53] Newborn, M. (1997). *Kasparov versus Deep Blue: Computer chess comes of age*. New York: Springer.

[54] Strickland E (2019) How Ibm Watson overpromised and underdelivered on AI Health Care. IEEE Spectrum, 2 April 2019. https://sprectrum.ieee.org/ biomedical/diagnostics/how-ibm-watson-overpromised-and-underdeliveredon-ai-health-care. Accessed 5 Jan 2020

[55] https://en.wikipedia.org/wiki/AlphaGo_versus_Lee_Sedol

[56] Pearl J, Mackenzie D (2018) The book of why. The new science of cause and effect. Basic Books, New York. p. 37

[57] Turing A (1950) Computing machinery and intelligence. Mind LIX 236:433–60

[58] Laplace PS (1951) Philosophical essay on probabilities (1814). Dover Publications, New York

[59] Jones N (2014) The learning machines. Nature 505:146–48

[60] Langley, P. (2019, July). Explainable, normative, and justified agency. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, No. 01, pp. 9775-9779).
[61] Polanyi M (2009) The tacit dimension (1966). The University of Chicago Press, Chicago
[62] Dreyfus, H. L. (c1986.). Mind over machine: The power of human intuition and expertise in the era of the computer / (S. E. Dreyfus & T. Athanasiou, Eds.). Free Press, p. 27.
[63]Gangadharbatla, H. (2021). The Role of AI Attribution Knowledge in the Evaluation of Artwork. Empirical Studies of the Arts, 0276237421994697.