Demystifying Artificial Intelligence

Gheorghe Tecuci

Professor of Computer Science and Director of the Learning Agents Center, George Mason University
Former Chair of Artificial Intelligence, Center for Strategic Leadership, US Army War College
Member of the Romanian Academy
tecuci@gmu.edu, http://lac.gmu.edu

What is Artificial Intelligence

Artificial Intelligence (AI) is the Science and Engineering domain concerned with the theory and practice of developing systems that exhibit the characteristics we associate with intelligence in human behavior, such as perception, natural language processing, problem-solving and planning, learning and adaptation, and acting on the environment. Its main scientific goal is understanding the principles that enable intelligent behavior in humans, animals, and artificial agents. This scientific goal directly supports several engineering goals, such as developing intelligent agents, formalizing knowledge, and mechanizing reasoning in all areas of human endeavor, making working with computers as easy as working with people, and developing human-machine systems that exploit the complementariness of human and automated reasoning.

There has been a lot of hype about AI, with claims that AI agents will become more intelligent than humans and even display humanity. Recently, more than 27,000 people, including several tech executives and very reputable researchers, such as Elon Musk, Steve Wozniak, and Stuart Russell, have signed an open letter calling for a pause on training the most powerful AI systems for at least six months because of "profound risks to society and humanity," and several leaders from the Association for the Advancement of Artificial Intelligence signed a letter calling for collaboration to address the promise and risks of AI (Durden, 2023).

This paper will present a scientific approach to AI (Tecuci and Schum, 2024a, b), showing what it can and cannot do. We will argue that all these fears are unjustified, that AI fundamentally differs from human intelligence, that it is syntactic, and that human intelligence is semantic. An AI program can behave intelligently but, as a trained animal, does not have a "semantic understanding" of the commands received and is not "intelligent."

Lauchbury (2017), former Director of DARPA's Information Innovation Office, identified three waves in the development of AI technology.

The First Wave of AI: Handcrafted Knowledge

This wave started in 1980 with the development of the expert systems industry. An *expert system* is an intelligent computer program that uses knowledge and inference procedures to solve problems that are difficult enough to require significant human expertise for their solution. The knowledge necessary to perform at such a level, plus the inference procedures used, can be thought of as a model of the expertise of the best practitioners in that field (Feigenbaum, 1982, p.1). A characteristic of first-wave systems is that the knowledge of human experts is encoded into an ontology of concepts and a set of rules, enabling expert reasoning over a narrowly defined domain.

Consider an AI agent that plays chess, trying to select the best opening move (see Figure 1). There are 18 possible moves, each changing the game board into a new position where it is the turn of the adversary to move. Thus, the AI agent now considers all 18 possible moves the adversary can make, then all its possible responses, and so on. This continues until states are reached, which represent end positions in the game (i.e., win, lose, or draw from the point of view of the AI agent). Then, starting from the bottom up, the AI agent determines each intermediate node's value (win, draw, or lose) based on how the game will end from that node if both players make their best moves. After all this projection is made, the AI agent is ready to make its opening move. Next, the adversary makes its move. After that the AI agent has again to decide where to move, and so on.

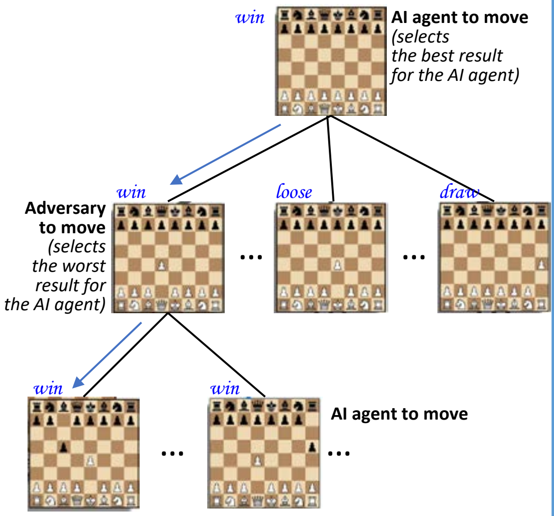

Figure 1. Top of the game tree in chess.

The problem with this exhaustive algorithm is that the search space is huge for any non-trivial game. In the case of the checker, for instance, it has been estimated that a complete game tree has around 10^{40} nonterminal nodes. If one assumes that these nodes are generated at a rate of 3 billion per second, the generation of the whole tree would still require around 10^{21} centuries! (Samuel, 2011, p.211). The search space for chess is much larger but significantly smaller than the search space for Go, which itself is much smaller than that of military operations, which involve more players, more possible moves, uncertainty about the state of the world (such as the actual dispositions of the opponent's units), and the use of deception by both forces.

The computational complexity of chess explains why only in 1997 was an automated agent (Deep Blue of IBM) able to defeat Gary Kasparov, the reigning world champion, although the above algorithm was known for 40 years. Deep Blue runs on a powerful parallel computer, generating up to 30 billion positions per move to explore about 14 moves in advance. It contains a database of about 4000 open positions, 700,000 grandmaster games, and many end-game solutions, coupled with a heuristic evaluation function based on about 8000 features.

Another example of a first-wave system is TurboTax, where the knowledge of tax lawyers and accountants has been encoded to help us do our taxes.

The main limitation of the first-wave systems is the *knowledge acquisition bottleneck*: It is very difficult to represent all the needed knowledge because of these systems' lack of powerful learning capabilities that makes this process manual and the poor treatment of uncertainty that results in sub-optimal representations.

The Second Wave of AI Technology: Statistical Learning

This wave started around 2000 with the development of deep neural networks, where engineers create statistical models and train them to learn complex concepts. A neural network consists of interconnected perceptrons. A perceptron is a simple computational unit inspired by the human neuron (see Figure 2). The neuron has a tree-like structure, with a corona consisting of the cell body and nucleus, branches (dendrites), a trunk (axon), and roots (dendrites). The dendrites connect with the dendrites of other neurons to form a very complex web of interconnected neurons.

The brain works on electricity. The axon is like a wire with insolation (called myelin sheaths). The neuron sends bolts of lightning (electrical impulses) in the axon that travel along the axon to the dendrites. The dendrite from one neuron ends and a dendrite from another neuron begins. This connection is called a synapse. There is a gap. The transmission is not electrical but chemical. The synapse causes the release of chemicals to the other neuron, which gets a signal. Each neuron receives signals from other neurons. If the sum of electricity exceeds a threshold, then the neuron fires. The synapse can be strong, medium, or weak. If the synapse is weak, when a signal comes in, it creates a weak signal in the next neuron. But if the synapse is strong, it creates a strong signal. What makes the connection strong or weak is your experience. This is where memory and learning occur. If this neuron makes that neuron fire, then their connection becomes stronger. This is what drives learning. From a statistical point of view, If the neurons fire together, it means they are correlated. When you see one firing, you would expect the other to fire.

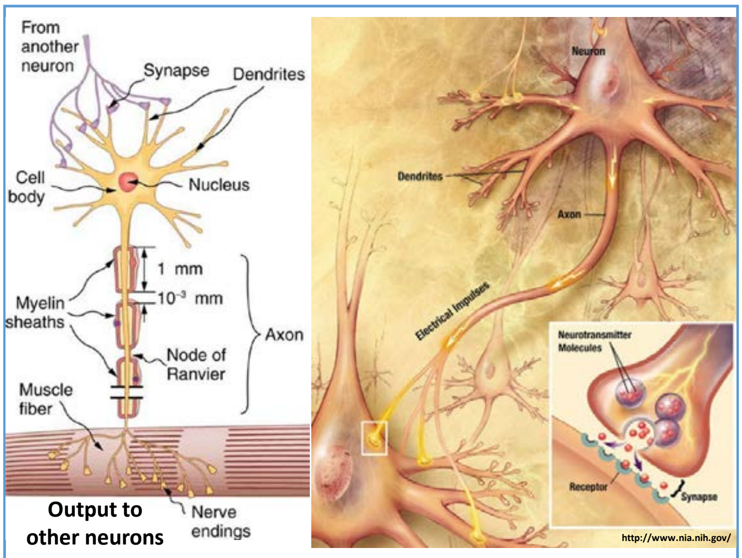

Figure 2. The human neuron.

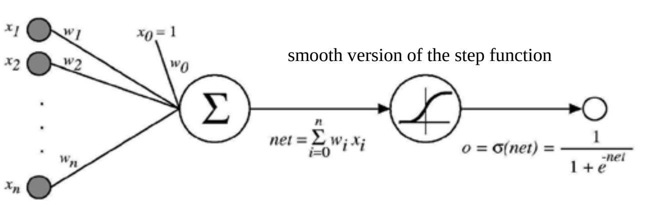

Figure 3. The perceptron.

The perceptron is a very crude approximation of a human neuron. It computes the weighted sum of its inputs and outputs 1 (true) if this sum is positive, and -1 (false) otherwise (see Figure 3).

Figure 4 shows a single-level neural network with four perceptrons as a hidden layer. This network can be trained to drive a car by tuning the weight of the connections to produce a steering wheel rotation that corresponds to the input image (Mitchell, 1997, p.84).

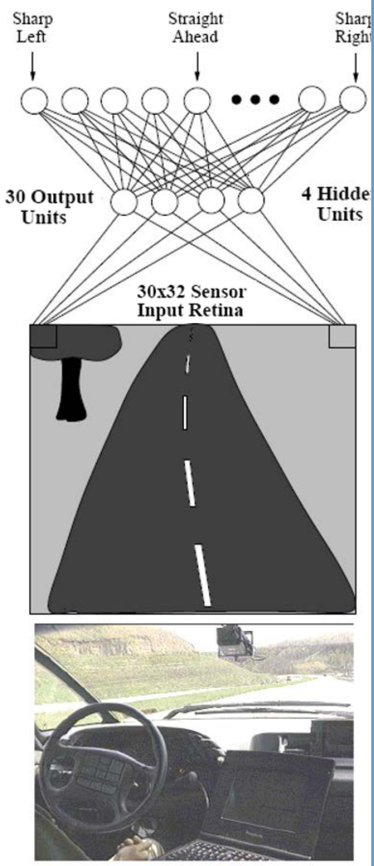

Figure 4. A neural network for driving a car.

A deep neural network has a huge number of layers with many perceptrons (Mitchell, 2019). They can learn to distinguish one human face from another (see Figure 5) or a vowel sound from another and play complex games like Go better than any human.

Large Language Model (LLM) systems, such as ChatGPT (Radford et al., 2019), represent and integrate what was posted on the Internet and can answer any question that Google can answer. Their natural language generation abilities allow them to compose answers and author stories and letters for different age groups and with different levels of detail. They do this by "reading" a large amount of existing text and learning how words appear in context with other words. Then, they use what was learned to predict the next most likely word that might appear in response to a user request and each subsequent word afterward. This is like auto-complete capabilities on search engines, smartphones, and email programs.

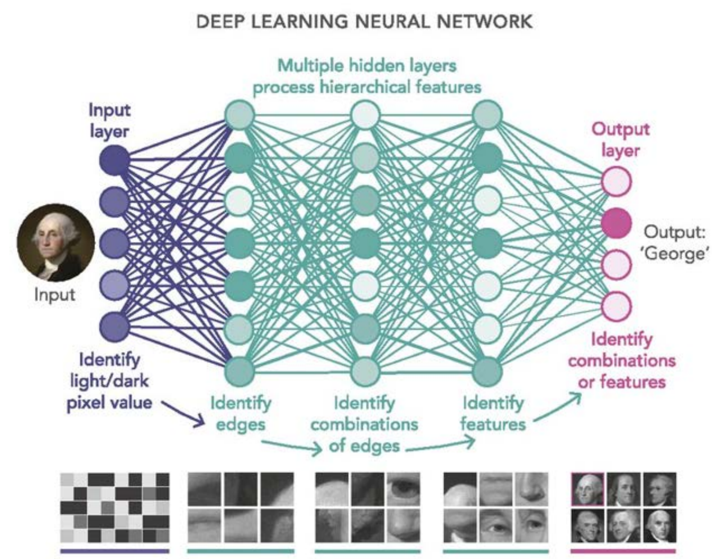

Figure 5. Deep neural network.

ChatGPT and other LLMs-based systems use sophisticated learning algorithms and deep neural networks to learn and generate answers. Their results are so impressive that Geoffrey Hinton, one of the inventors of deep learning, claims that computers can "understand" and even surpass human intelligence. However, he is mistaken because a computer only performs syntactic symbol manipulation, where syntax concerns the order and arrangement of words and phrases in sentences.

A very simple and convincing demonstration of this statement is Philosopher John Searl's Chinese room argument (Searle, 1980), illustrated in Figure 6:

John is inside a room where there is a book containing a huge collection of if-then rules:
IF you receive the symbol X, then return the symbol Y.

Through a door opening, John receives from outside the room the symbol X, representing a question in Chinese, and, following one of the rules, returns the symbol Y, representing the answer in Chinese. For the outside observer, John seems to understand Chinese. But John does not know any Chinese.

Figure 6. The Chinese Room Argument.

True "understanding" requires semantic processing, which is concerned with the meaning of a word, phrase, sentence, or text and can only be performed by humans. Human intelligence is thus qualitatively different from artificial intelligence.

Unlike the first-wave systems, the second-wave systems have very poor reasoning capabilities. They suffer from the *data engineering bottleneck*, requiring an enormous amount of data.

The Third Wave of AI Technology: Contextual Adaptation

We are now entering the explainable AI wave with hybrid, explanatory, interactive, and human-centric AI systems. An example of a third-wave AI system is the Scientist's Apprentice

(Tecuci, 2004), which synergistically integrates humans' imaginative reasoning and computers' knowledge-based critical reasoning. The human scientist will teach the AI agent how to perform scientific discovery, and the agent will perform the learned operations but will also help the scientist to be more creative. Like an expert system, the AI agent is a knowledge-based system. It also integrates an LLM system, giving it powerful natural language processing capabilities.

Typically, a knowledge engineer develops a knowledge-based system by formally encoding the domain's expertise into the agent's knowledge base through a knowledge acquisition process known to be *difficult*, *time-consuming*, and *error-prone* (Tecuci et al., 2016). In contrast, the scientist demonstrates to the agent how to solve a typical problem; the agent learns rules by generalizing individual reasoning steps and uses these rules to solve similar problems. The analyst reviews and corrects the analysis performed by the agent, and the agent refines or revises the rules and learns additional ones. As a result, the agent incrementally learns the problem-solving expertise of the scientist.

The overall reasoning of the agent follows *the discovery scientific method* (Tecuci and Schum, 2024), illustrated in Figure 7 and presented in the following.

First the scientist will use *abductive (imaginative) reasoning* (that shows that something is *possibly* true) to generate possible answers to the question or hypotheses that would explain the observation.

Each hypothesis will guide the discovery of relevant evidence by employing *deductive reasoning* (that shows that something is necessarily true). They will develop arguments that decompose the hypothesis into simpler and simpler hypotheses until the simplest ones point directly to this evidence.

Finally, they employ inductive reasoning (that shows that something is probably true) to test the hypothesis.

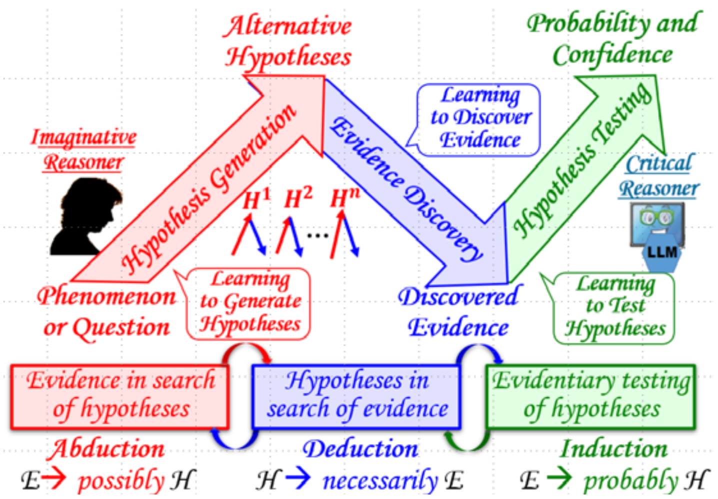

Figure 7. The discovery scientific method.

If the tests of our initial H do not consistently support it, we might buy support for H by revising it. This is entirely legitimate since hypotheses are conceptual entities that can be revised. What we cannot do, of course, is to revise the evidence to make it fit our conceptual hypotheses. *H* might be true if we add condition X to it. So we revise H to H^1 = H ∩ X. We might have to revise H^1, based on our testing, by adding condition Y to it so that H^2 = H^1 ∩ Y = H ∩ X ∩ Y, and so on.

The scientist will use its LLM component to develop an *ontology* for the application domain, such as the one in Figure 8 for *engineering design* (Arciszewski, 2016).

Figure 8 shows how, from Example1 (*Floodwall protects the residential house from water*) and its explanation (*The structure of floodwall is such that water from which it protects is also used to strengthen the protection, through the pressure it exerts on the base part of the L-shaped* *floodwall.)* the agent learned a general rule through ontology-based generalization.

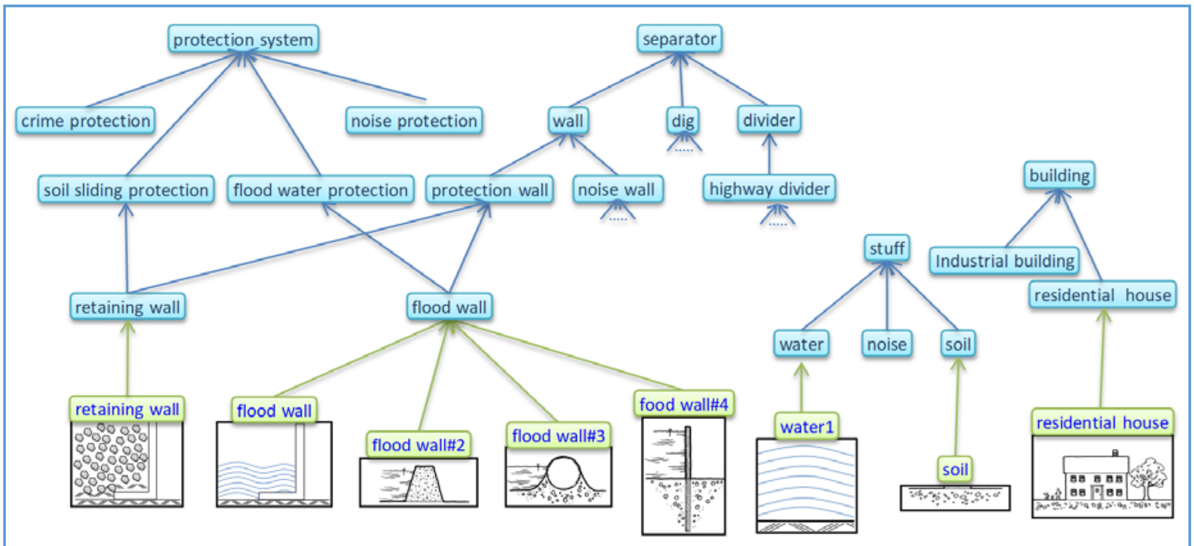

Figure 8. Engineering design ontology.

The argument pattern was obtained by replacing the instances from the argument (i.e., *floodwall, residential house, and water*) with variables (?O1, ?O2, ?O3). The rule has an applicability condition that indicates the possible values of these variables for which the argument is likely to be correct. Notice, however, that instead of a single applicability condition, the agent learned a lower and an upper bound for this condition using two complementary strategies:

- The strategy of a *cautious learner* who wants to minimize the chances of making mistakes when applying the learned rule (lower bound). This strategy increases the confidence in reasoning but may fail to apply the rule in situations where it is applicable.
- The strategy of an *aggressive learner* who wants to maximize the opportunities of employing the learned rule (upper bound). This strategy increases the number of situations where the rule can be applied, although the reasoning may not be correct

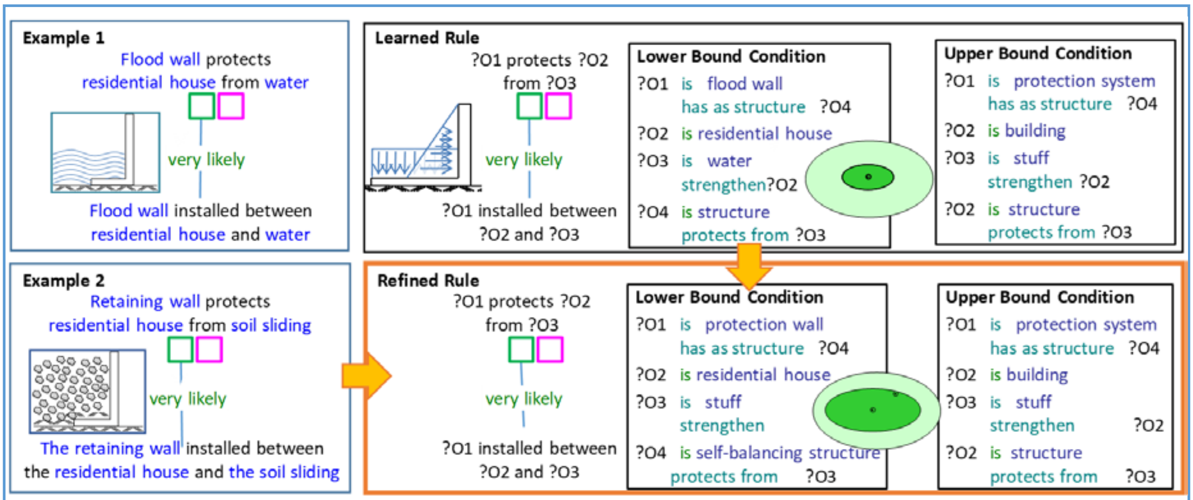

Figure 9. Rule learning and refinement.

in some situations.

The rule was automatically applied to generate Example 2, that was accepted as correct by the scientist: Retaining wall protects residential house from soil sliding.

The result is the generalized rule:

> *Self-balancing system, where the structure is such that the stuff from which it protects is also used to strengthen the protection through the pressure it exerts on the base part of the L-shaped self-balancing system.*

The rule bounds may be further refined based on additional examples.

Democracy, Critical Thinking, and Artificial Intelligence

Socrates criticized (Athenian) democracy for allowing selfish individuals to gain power and wealth by using speech-making tricks and flattery to gain the support of citizens.

Plato also emphasized the risks of bringing dictators, tyrants, and demagogues to power, that democracies have leaders without proper skills or morals, and that it is quite unlikely that the best equipped to rule will come to power.

Finally, Eminescu warned that politicians' mistakes are crimes because millions of innocent people suffer. As a result, the development of an entire country is hindered, and its future is hindered for decades to come.

All these fears have been confirmed today when politicians, political parties, and media, have lost all credibility. As a result, the Internet, Google, YouTube, Facebook, Instagram, LinkedIn, personal Podcasts, and other individual and social media enterprises have become primary, yet unreliable, news sources. Therefore, it became necessary for each of us to distinguish better between factual and fabricated information, not only in transnational events, such as the war between Russia and Ukraine or the barbaric attack of Israel by Hamas, but in every aspect of our lives, including politics, elections, work, and personal aspects. Thus, we need to become better critical thinkers.

Critical thinking is a complex process that was developed over the past 2500 years through the work of some of the greatest minds, including Aristotle, Galileo Galilei, John Locke, Isaac Newton, William Whewell, Charles Peirce, John Wigmore, and David Schum, who have tried to understand the world through a process of *discovery and testing of hypotheses based on evidence*. In essence, critical thinking refers to the ability to analyze information objectively and make a reasoned judgment (Tecuci, 2004)

Artificial intelligence can help us become better critical thinkers, for example, by learning to answer questions through the process outlined in Figure 7.

Conclusions

As with any new and powerful technology, such as nuclear power, Artificial Intelligence comes with risks and opportunities. It is up to us to manage the risks and take advantage of the enormous opportunity offered. Besides the technological advances made possible (e.g., only AI can determine fake videos or images), it may help us become better critical thinkers, this being the best way of preserving democracy, which, with all its imperfections, is still the best system of government (Tecuci, 2024).

None of the artificial intelligence systems mentioned above (including ChatGPT) have attributes unique to human intelligence, particularly imaginative reasoning, consciousness, semantic understanding, creativity, intuition, and wisdom.

The scary, futuristic presentations of AI by the media (and now even by Geoffry Hinton, Elon Musk, and Steve Wozniak) have no basis in reality. There is no competition between humans and AI robots on the horizon, and probably is not even possible (AI 100, 2016).

References

AI 100 (2016). *One Hundred Year Study on Artificial Intelligence*, Report of the 2015 Study Panel, Stanford University, September 2016. https://ai100.stanford.edu/2016-report

Arciszewski, T. (2016). *Inventive Engineering: Knowledge and Skills for Creative Engineers*, CRC

Durden T. (2023). Musk, Wozniak Call for Pause in Developing 'More Powerful' AI than GPT-4, *Zero Hedge*, March 29, 2023.

Feigenbaum, E. A. (1982). Knowledge Engineering for the 1980s, *Research Report,* Stanford University, Stanford, California, U.S.A

Lauchbury J. (2017). A DARPA Perspective on Artificial Intelligence, YouTube, https://www.youtube.com/watch?v=-O01G3tSYpU

M. Mitchell Waldrop (2019). PNAS 2019;116:4:1074-1077

Mitchell, T.M. (1977). Version Spaces: A Candidate Elimination Approach to Rule Learning, *Proceedings of the 5th IJCAI,* Cambridge, MA. https://www.ijcai.org/Proceedings/77-1/Papers/048.pdf

Mitchell, T.M.(1997). *Machine Learning*, McGaw Hill.

Radford et al., (2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

Samuel AL. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development* 1959, 3:210-229.

Searle J.R. (1980), Minds, brains, and programs. *Behavioral and Brain Sciences* 3 (3): 417-457

Tecuci, G. (1988). *Disciple: A Theory, Methodology, and System for Learning Expert Knowledge*, Thèse de Docteur en Science, University of Paris-Sud. http://lac.gmu.edu/publications/1988/TecuciG_PhD_Thesis.pdf

Tecuci, G. (1998). *Building Intelligent Agents: An Apprenticeship Multistrategy Learning Theory, Methodology, Tool and Case Studies*. London, England: Academic Press. http://lac.gmu.edu/publications/1998/TecuciG_Building_Intelligent_Agents/default.htm

Tecuci, G., Marcu, D., Boicu, M., Schum, D.A. (2016). Knowledge Engineering: Building Cognitive Assistants for Evidence-based Reasoning, Cambridge University Press.

Tecuci, G., (2023) How Intelligent is Artificial Intelligence, *Evidence Dialogues* https://evidencedialogues.wordpress.com/
http://lac.gmu.edu/publications/2023/How%20Intelligent%20is%20AI.pdf).

Tecuci, G., (2024) *Gentle Introduction to Critical Thinking,* CRITIS Press.

Tecuci G., Schum D.A. (2024a). *Science of Evidence*, Learning Agents Center, 6 pages.

Tecuci G., Schum D.A. (2024b). *Science of Evidence*, CRITIS Press.