Why Artificial Intelligence Is Not a Threat to Humanity but a Huge Opportunity

Gheorghe Tecuci
Professor of Computer Science and Director of the Learning Agents Center, George Mason University, USA; Former Chair of Artificial Intelligence, Center for Strategic Leadership, US Army War College; Member of the Romanian Academy. Gheorghe.Tecuci@gmail.com

Abstract: There has been a lot of hype about artificial intelligence, with claims that artificial intelligence agents will become more intelligent than humans and even display humanity. We will show that this fear is unjustified, that artificial intelligence fundamentally differs from human intelligence, and that they are complementary, with artificial intelligence being better at some tasks but unable to perform others that may be performed by human intelligence. We will present a model of critical thinking that facilitates the synergistic integration of a human's imaginative reasoning with a machine's critical reasoning, able to solve problems that are limited only by our imagination. As with any new and powerful technology, artificial intelligence comes with risks and opportunities. Many more human jobs will be performed by machines, but these are all algorithmic jobs, leaving the truly creative ones to the people. Most importantly, artificial intelligence may help us become better critical thinkers, which is the best way of preserving democracy, which, with all its imperfections, is still the best system of government.

1 Introduction

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.

The main scientific goal of AI 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." 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).

I am one of its greatest supporters, and I think that AI is probably the greatest accomplishment of the human mind. I had the chance to interact with some of the founding fathers of AI, including Herbert Simon (who predicted the future of machines and the importance of data and won the Nobel Prize in Economics for his contributions to the theory of bounded rationality) and John McCarthy (who coined the term "artificial intelligence," created the LISP language, and invented of time-sharing interactive programming). I was one of the co-editors of the classical Machine Learning volumes, along with some of the best AI scientists, Tom Mitchell, Ryszard Michalski, Jaime Carbonell, and Yves Kodratoff. I have dedicated my professional life to developing a general theory of learning agents that can be taught similarly to how students are taught, through examples and explanations, and by supervising and correcting their problem-solving behavior. Because such agents learn to replicate the problem-solving behavior of their teachers, I have called them Disciple agents (see Figure 1).

Figure 1: With Herbert Simon and Ryszard Michalski in the Conference Room of the Artificial Intelligence Center, George Mason University, 1991 (top left). Presenting the Disciple approach to John McCarthy at the Conference of the American Association for Artificial Intelligence, 2002 (top right). The classical Machine Learning books were published in 1983, 1986, 1989, and 1994, respectively (middle). Disciple: A Theory, Methodology, and System for Learning Expert Knowledge, PhD thesis, 1988, and Instructible agent books, published in 1998, 2016, and 2024, respectively (bottom).

2 What an AI Agent Can and Cannot Do

There are indeed some very impressive accomplishments of AI, such as Deep Blue (the IBM chess program that defeated world champion Gary Kasparov), AlphaGo, which plays better Go than any human, IBM's Watson, who defeated the best human players at Jeopardy, and attribute to AI systems super-intelligence abilities. The latest one is ChatGPT (Radford et al., 2019), which represents and integrates what was posted on the Internet and can answer any question that Google can answer. It does 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.

Its superb natural language generation capabilities allow it to compose answers and author stories and letters for different age groups and with different levels of detail. It can compose music, essays, and poems, write and debug computer programs, play games, generate ideas for creative tasks, write personalized resumes and cover letters, etc.

These results are so impressive that Geoffrey Hinton, one of the inventors of deep learning, claims that computers can "understand" and will even surpass human intelligence. He points to the fact that they are based on neural networks that already contain more neurons than the human brain and can learn much faster than humans.

I will show that he is wrong, that computers are significantly less intelligent than humans. I will start comparing the human neuron with the artificial one (see Figure 2).

The human 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. 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, they are correlated. When you see one firing, you would expect the other to fire.

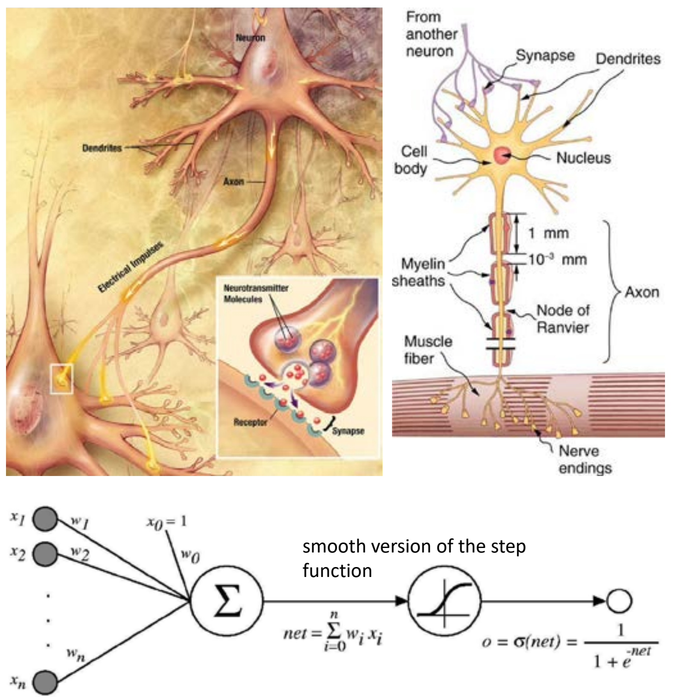

Figure 2: The human neuron (top) and the artificial one

Biological learning systems are built of very complex webs of interconnected neurons.

The artificial neuron is a *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.

Table 1 compares the characteristics of the two types of neurons.

Table 1: The human neuron versus the artificial one.

Human Neuron	Artificial Neuron
Switching time10^{-3} second	Switching time ≈ 10^{-9} second
Number of neurons≈10^{10}	Number of transistors/chip > 10^{10}
Connections/neuron≈10^{4-5}	Connections/ transistor ≈ 10
Power consumption brain: 20 watts	Power consumption equivalent computer ≈10^{6} watts
Scene recognition time ≈ .1 second	Scene recognition: much weaker

The artificial neuron is much faster, but its number of connections is much smaller. The artificial network is much larger, but it needs much more power and time to recognize a scene than the human brain.

There are three forms of reasoning: *deductive* (that shows that something is *necessarily* true), *inductive* (that shows that something is *probably* true), and *abductive or imaginative* (that shows that something is *possibly* true). *An AI agent can only perform deductive and inductive reasoning, but it cannot perform abductive (imaginative) reasoning.* As a result, an AI agent is fast, rigorous, precise, and objective but lacks intuition, imagination, and the ability to deal with new situations.

If you want to know what an AI agent cannot do, think of a task that requires imaginative reasoning.

3 Human Intelligence


John Searle

A computer only performs syntactic symbol manipulation, as convincingly demonstrated by Philosopher John Searle with the Chinese room argument (1980):

> *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.*


Mihai Drăgănescu 1929-2010

True "understanding" requires "semantic" processing that only humans can do.

Mihai Drăgănescu (1979, 1985) advanced the idea that the brain has both computing and non-computing ways of processing information. This was later demonstrated by Roger Penrose (1994). Drăgănescu distinguishes between two types of information: *structural*, which can be reduced to bits, and *phenomenological*, which has a manifestation in feelings, meanings, and qualia. In the philosophy of mind, qualia are defined as instances of subjective, conscious experience. Examples of qualia include the perceived sensation of pain from a headache, the taste of wine, and the redness of an evening sky. Both kinds of information may also act together, constituting a mixed type of information.


Roger Penrose

The brain has a mind and consciousness. According to Drăgănescu (2000), there are many levels of information processing in the brain:

- The highest level is the *psychological level*, which may be seen as a specific macroscopic level, which comprises behavior, intellectual activities, thinking, sentiments, will, and others.
- The *neuronal level* comprises the networks of neurons, modules of neurons, and the structural organization of the brain.

- The *molecular level* comprises the molecular activities inside the neurons and at the synapses between neurons.
- The *quantum level*, which was proposed by a number of physicists.
- The *experiential level* (phenomenological level) proved to be a reality of the brain and mind.

Humans are slow, sloppy, forgetful, implicit, and subjective, but they have a conscience, intuition, and imagination and can find creative solutions in new situations. Humans can perform all types of reasoning, including *abductive (imaginative) reasoning, which a machine cannot.*

Speaking on imagination, Einstein said,

> *When I examine myself and my methods of thought, I come close to the conclusion that the gift of fantasy [imagination] has meant more to me than my talent for absorbing absolute knowledge.*
> *The true sign of intelligence is not knowledge but imagination.*
> *Logic will get you from A to B. Imagination will take you everywhere.*

Albert Einstein [1879-1955]

As you can see, AI and human intelligence are highly complementary. Therefore, it makes sense to think of human-machine systems. For instance, what human or machine alone would be able to defeat Kasparov when assisted by Deep Blue?

4 Human-Machine Systems

Figure 3 is an overview of a human-machine model of critical thinking that facilitates the synergistic integration of a human's imagination and expertise with the computer's domain knowledge and critical reasoning. It is a general model grounded in the science of evidence (Tecuci and Schum, 2024) and the scientific method of hypothesis generation and testing.

Critical thinking is the ability to objectively analyze information and make reasoned judgments. It will start with an interesting phenomenon to be explained or a question to be answered (e.g., Who planted the bomb?). The explanations of the phenomenon or the answers to the question are the hypotheses to be analyzed, are through abductive (imaginative) reasoning that shows something is possibly true.

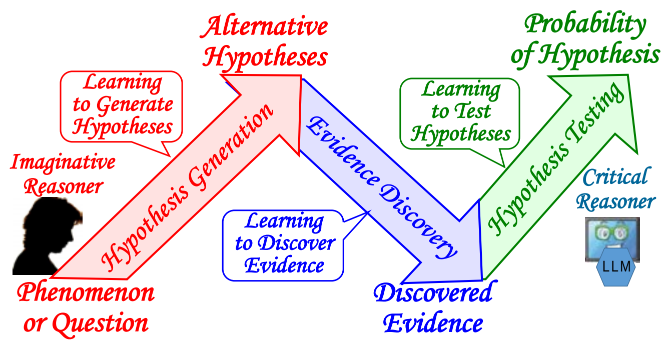

Figure 3: Human-machine model of critical thinking.

They develop a Wigmorean argumentation, a probabilistic inferential network illustrated in Figure 4. Hypothesis H is decomposed into simpler hypotheses by considering both favoring arguments (supporting the truthfulness of H), under the left (green) square and disfavoring arguments (supporting the falsehood of H under the right (pink) square, until the sub-hypotheses are simple enough to point to the evidence to be collected. The probabilities of the bottom hypotheses are assessed based on the credibility of the collected evidence, and the probabilities of the upper-level hypotheses are assessed based on the probabilities of their sub-hypotheses using min-max probability combination rules common to the Fuzzy and the Baconian probability views.

From each argument, the agent also learns a general rule by employing *apprenticeship multistrategy learning* (Tecuci et al., 2016, pp.252-328), which is a development of the *classical version space method* (Mitchell, 1977), and integrates learning from examples from explanation, and by analogy (see the top left of Figure 4). The argument was obtained by

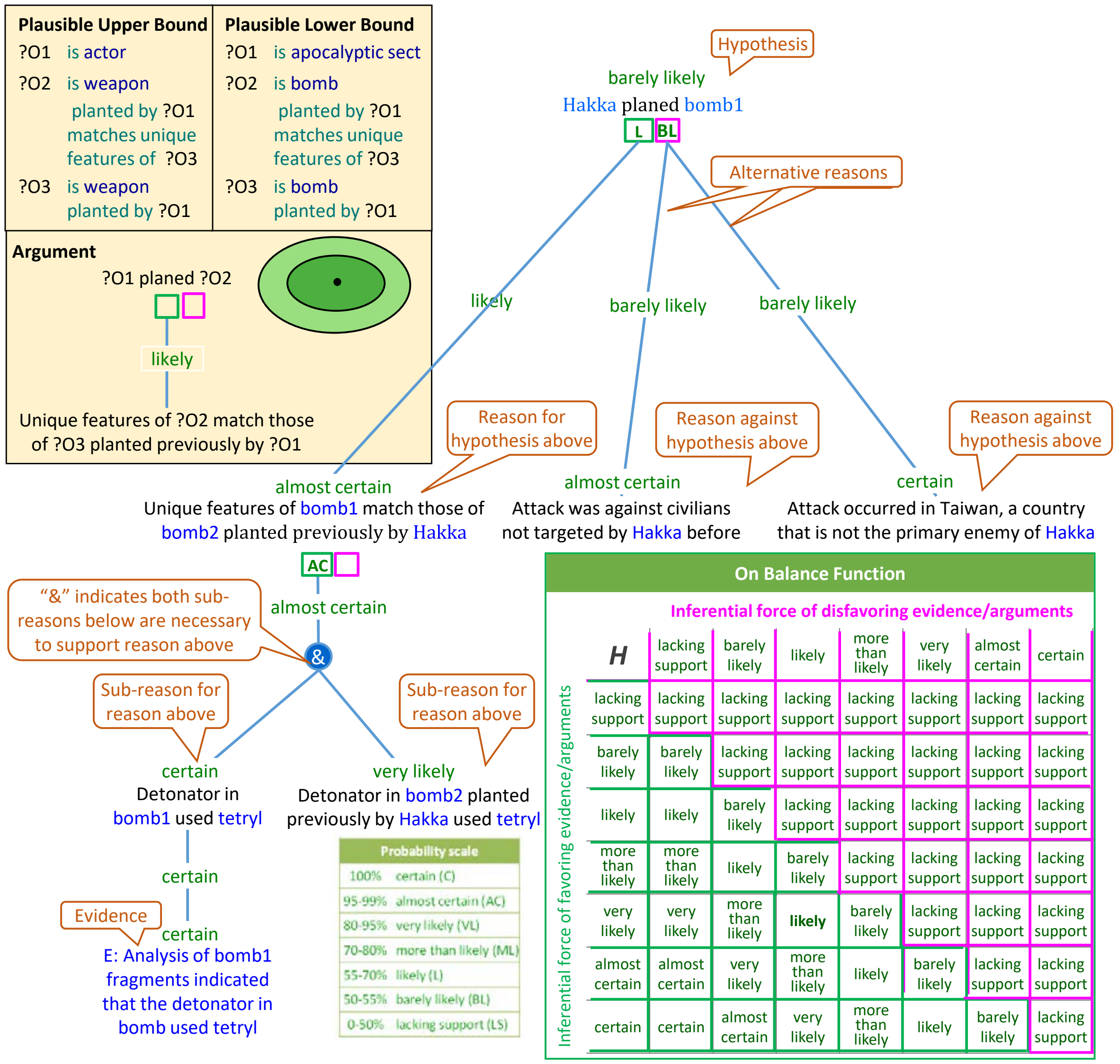

Figure 4. Wigmorean probabilistic infrence network (argumentation)

replacing the instances from the argument (i.e., *Hakka*, *bomb1, and bomb2*) 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 in some situations.

Such rules enable the agent to generate solutions to similar problems and, if correct, lead to the generalization of the rule.

The problems that can be solved by a human-agent system are only limited by human imagination. For example, Steven Rieber (2023) imagined an intelligence application where the system would automatically produce comments (feedback and recommendations) on a draft analytic report, highlighting additional relevant evidence and identifying strengths and weaknesses in the draft's reasoning. Analysts can use the comments to improve their reports. In contrast to current applications of structured analytic techniques, the system will automatically produce comments with no additional effort from analysts, who can use any comments they find valuable. These comments will be based on the automated application of effective structured analytic techniques. The comments will be analogous to those made by automated spelling and grammar checks, except that they will focus on improving argumentation instead of writing.

Steven Rieber

A proposed solution to this problem is given in (Tecuci, 2023).

5 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. Yes, many human jobs will be performed by machines, but these are all algorithmic jobs, leaving the truly creative jobs to people. 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).

6 References

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

Dawkins, R. (2008). *The Oxford Book of Modern Writing*, Oxford University Press.

Drăgănescu, M. (1979). *The Depths of the Material World*, Bucharest 1979; English edition 1997, http://www.racai.ro/books/doe

Drăgănescu, M. (1985). *Orthophysics*, Editura Sttintifica si Enciclopedica, Bucharest,

Drăgănescu, M. (2000). The Brain as an Information Processor, *NOESIS*, XXV, 2000, p.9-20.

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

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

Rieber, S. (2023). *REASON Overview*, IARPA, January 11, https://www.youtube.com/watch?v=FiHKoBi1No0

Penrose. R (1994). *Shadows of the Mind*, Oxford, Oxford University Press.

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

Tecuci, G., (2023). *Cognitive Agent for Rapid Explanation, Analysis, and Sourcing Online (Cogent REASON*), IARPA, January 11,https://www.youtube.com/watch?v=jYAFnKKVdpU, 2:05:43 to 2:10:42,

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

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

Tecuci G., Schum D.A. (2024). *Science of Evidence*, Critical Thinking Press.