Institute for National Security Studies

Report Part Title: What is Artificial Intelligence?

Report Title: Artificial Intelligence and National Security in Israel
Report Author(s): Liran Antebi
Published by: Institute for National Security Studies (2021)
Stable URL: https://www.jstor.org/stable/resrep30590.7

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

Institute for National Security Studies is collaborating with JSTOR to digitize, preserve and extend access to this content.

This content downloaded from 212.37.28.30 on Fri, 26 Jan 2024 11:12:27 +00:00
All use subject to https://about.jstor.org/terms

Part I: Artificial Intelligence and its Security Applications

By far, the greatest danger of artificial intelligence is that people conclude too early that they understand it.

Eliezer Yudkowsky, American AI researcher and writer

This content downloaded from 212.37.28.30 on Fri, 26 Jan 2024 11:12:27 +00:00
All use subject to https://about.jstor.org/terms

This content downloaded from 212.37.28.30 on Fri, 26 Jan 2024 11:12:27 +00:00
All use subject to https://about.jstor.org/terms

Chapter One: What is Artificial Intelligence?

The idea of AI first developed in 1945 when Vannevar Bush, one of the early founders, proposed a system to increase human knowledge and understanding. He was followed by Alan Turing, who in 1950 wrote an article on the capabilities of machines to simulate human beings and their ability to perform intelligent actions such as playing chess.[1] The term artificial intelligence (AI) evolved a few years later and is attributed to John McCarthy, a computer scientist and researcher in the field of cognitive sciences, who organized the first academic conference on the subject in 1956, and to Marvin Lee Minsky, who was trained as a mathematician and was involved in research, inventions, and many developments in the field. It was Minsky who coined the popular definition of AI, noting that "AI is the science of making machines do things that would require intelligence if done by men."[2]

At the beginning of the study of AI, the dominant paradigm was the "symbolic" one, which sought to duplicate high-level human thought. Over the years it was replaced by the "connectionist" paradigm, which endeavored to imitate the biological basis of human cognition through artificial neurons. These paradigms, however, failed to meet expectations beyond theoretical or laboratory demonstrations and led to the "winter of AI," when research and investments in AI were minimal for long periods of time.[3]

In the past decade, due to progress in computer science research, the development of hardware and software in computing and communication, as well as cloud computing and big data, AI has significantly progressed, including in subdomains such as machine learning and artificial neural networks (these concepts will be reviewed in detail later). Some studies have claimed that the progress in the areas of neural networks is so profound that it is almost considered synonymous with that of AI.[4]

This content downloaded from 212.37.28.30 on Fri, 26 Jan 2024 11:12:27 +00:00
All use subject to https://about.jstor.org/terms

Most common applications in AI belong to a subdomain called machine learning, which includes statistical algorithms that seek to imitate human cognitive tasks by analyzing large amounts of data and creating rules about them. The algorithm actually "trains" on existing information and creates a kind of statistical model of its own, in order to perform the same task in the future on new data that it has not previously encountered.[5]

AI belongs to the wider field of data science, and indeed, it needs a great deal of data to operate effectively, specifically big data, which is needed to generate significant insights with the help of learning algorithms. However, AI does not depend solely on big data, which is only one of the efficient means of generating value and knowledge from such an amount of data, which requires especially strong algorithms to analyze them.[6]

A considerable part of the work of the founders of AI was the theoretical basis for machine-learning algorithms, which are used in many contemporary systems and enable actions such as image identification and autonomous driving.[7] These systems belong to what is known as narrow AI or weak AI, although sometimes these can be advanced applications. This concept refers to algorithms that are designed to deal with a cluster of specific problems, such as games, image identification, or navigation.[8] This concept differs from general AI, which relates to a system capable of using human-level intelligence for a wide range of tasks.[9] As of this writing, general AI still does not exist, and opinions are divided on whether it will be created, at least within the next two decades. The AI that has been developed belongs mainly to deep learning applications. This technology indeed can be categorized as narrow AI, but it enables a more accurate form of computerized learning as well as a broader commercial use of AI applications.[10]

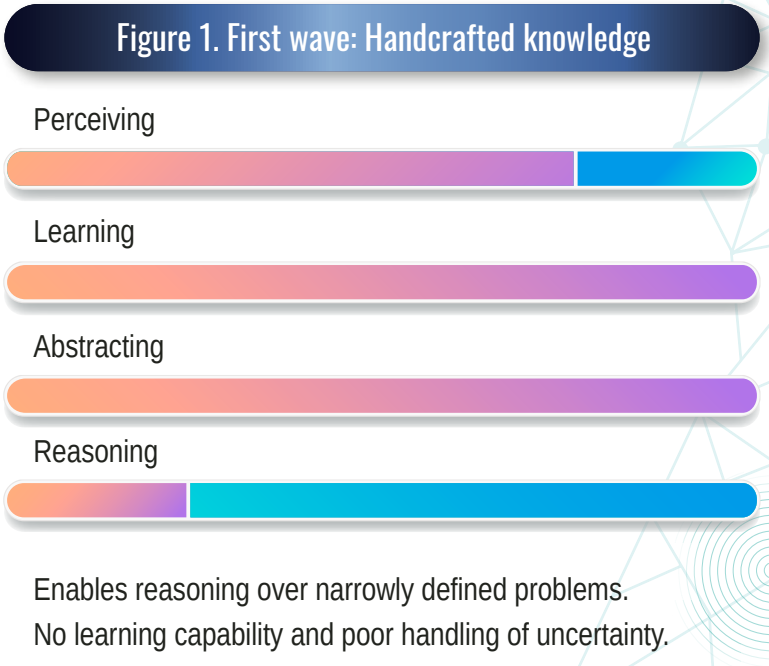
Historical Background: The First Three Waves of AI

The development of AI can be divided into three distinct waves, based on the development of AI's capabilities. The Defense Advanced Research Projects Agency (DARPA) of the US Department of Defense is one of the world's leading bodies in the development of AI for security purposes. DARPA defines AI as a "programmed ability to process information."[11] Alongside this simple definition, DARPA has divided AI into three waves, characterized by the Notional Intelligence Scale in which the following four capabilities are measured, similar to the dimensions of human intelligence:

This content downloaded from 212.37.28.30 on Fri, 26 Jan 2024 11:12:27 +00:00
All use subject to https://about.jstor.org/terms

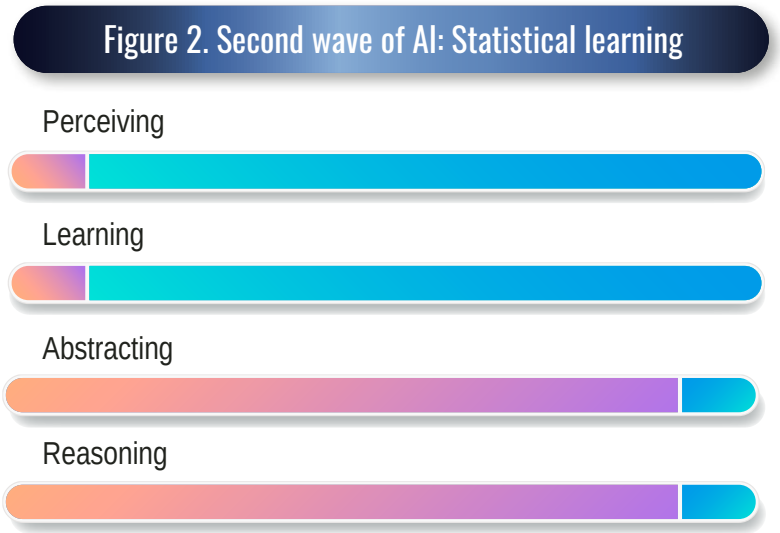
1. Perceiving: the ability to discern global events
2. Learning: the ability to learn things and adapt to various situations
3. Abstracting: the ability to take knowledge discovered at a certain level and to deduce from it or apply it to another level
4. Reasoning: the ability to explain logically, or to make logical decisions.

The first wave of AI was based on "handcrafted knowledge," in which experts collected existing knowledge on a particular subject and characterized it within the framework of rules that could apply to a computer, which in turn could learn their implications.[12] This generation of AI includes logistics software for planning operations such as shipments; software for calculating taxes; and software that could play chess games against people. Many computer programs and applications on smartphones or in software such as Microsoft Office are based on this wave of AI. According to DARPA, the products of the first wave have moderate sensory ability and can explain causality in very narrow aspects, but they lack learning abilities, and cannot cope with uncertainty. Nonetheless, DARPA claims that this wave had many achievements, such as in cyber defense, and it continues to be developed and is still relevant today.[13]



Source: Launchbury, "A DARPA Perspective on Artificial Intelligence."

This content downloaded from 212.37.28.30 on Fri, 26 Jan 2024 11:12:27 +00:00
All use subject to https://about.jstor.org/terms

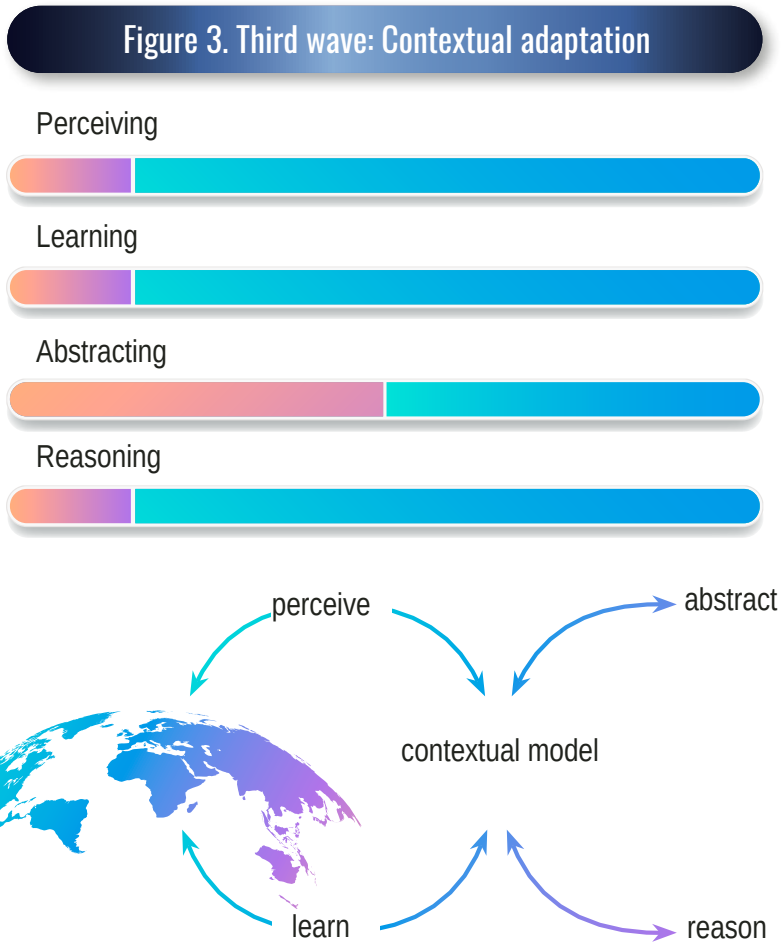


Source: Launchbury, "A DARPA Perspective on Artificial Intelligence."

The second wave is referred to as "statistical learning," characterized by categorization. In this wave, experts made use of more advanced capabilities facilitated by machine learning, in which algorithms for statistical learning rely on big data. In this wave, unlike the previous one, the experts taught the computers statistical models for various problems, instead of fixed rules and then trained the algorithms on many examples, until they reached the desired level of accuracy. The products of this wave enabled voice recognition or facial recognition on mobile phones and "bots" that provide customer service through internet chat correspondence.

This generation of AI includes systems for analysis or translation of text; personal assistant software in smart phones; and the ability to play challenging games such as the Chinese strategy game "Go." This wave of AI also includes autonomous driving. This generation of AI, however, does not have the ability to understand the rules or the causality behind the actions it performs, so it is subject to error or manipulation. According to DARPA, the second wave of AI could categorize things according to nuances and predictive ability but lacked contextual abilities and had minimal abilities for logical reasoning.

This content downloaded from 212.37.28.30 on Fri, 26 Jan 2024 11:12:27 +00:00
All use subject to https://about.jstor.org/terms



Source: Launchbury, "A DARPA Perspective on Artificial Intelligence."

The third wave, referred to as "contextual adaptation," is an explanatory one, which is currently being developed. The algorithms or systems from this wave will formulate models that explain certain topics. DARPA expects that systems built around contextual models will learn by themselves how different models should be structured. These abilities are significantly different from most of the algorithms that currently operate as a "black box" and create a challenge of explainability as to how they reached conclusions (a topic that will be expanded upon in a later section). Thus, this wave of AI will use information in an abstract manner and take it one step forward,

This content downloaded from 212.37.28.30 on Fri, 26 Jan 2024 11:12:27 +00:00
All use subject to https://about.jstor.org/terms

but currently the capabilities of these systems are still limited.[14] It is hoped that the products of this wave will be more "human" and will be able to communicate in natural language, will be able to teach and train themselves (like the Alpha-Go software that has trained itself in thousands of "Go" games against itself), and will be able to collect data from several different sources, and formulate well-explained conclusions.[15] According to DARPA, this wave should greatly improve the AI capabilities in sensory, learning, and reasoning fields, although the products will still only have medium-sized capabilities in the field of abstraction.

Technologies in the context of this wave include "smart assistants," whose capability to assist has advanced beyond the technologies of the second generation, such as Siri and Alexa.[16] Another example is Google Duplex, which can make appointments (such as making a reservation in a barbershop or restaurant) while managing a coherent vocal conversation with a human service representative. Besides the tasks that this software can do autonomously, it also knows how to identify and signal the user regarding tasks that it cannot perform on its own.[17]

While the three waves of AI are easily identified, most research dealing with AI in recent years, particularly in the field of national security, has addressed the fact that there is no one definition for the term AI. Formulating one accepted definition of AI is problematic for two main reasons: First, there are varied and diverse approaches to research in the area.[18] Second, there is a basic difficulty in defining or agreeing upon a definition of "intelligence," because of limitations that have not yet been breached in the study of neuroscience (and also in philosophy); therefore the ability to examine these concepts in relation to machines or to apply them to machines is limited. Despite this difficulty, this study will examine different definitions and suggest a definition for the remaining discussion in this document and the policy recommendations that follow.

AI—An Operational Definition

One of the known definitions of AI, which has already been mentioned, was formulated by Marvin Lee Minsky as "the science of making machines do things that would require intelligence if done by men."[19] The advantage of this definition is that it is broad enough to include different ideas, methods, and means. However, it lacks the use of the term "intelligent" in the

This content downloaded from 212.37.28.30 on Fri, 26 Jan 2024 11:12:27 +00:00
All use subject to https://about.jstor.org/terms

human context—a term that has not yet been defined and is characterized unambiguously by the scientific disciplines engaged in the subject. Moreover, when defining the discipline for national security and for making policy recommendations, a more dichotomous definition is necessary, which will help determine what it should include and what is irrelevant.

Darrell West and John Allen have claimed that "artificial intelligence (AI) is a wide-ranging tool that enables people to rethink how we integrate information, analyze data, and use the resulting insights to improve decision making." West and Allen believe that even though there is no uniform accepted definition, it is correct to refer to AI as "machines that respond to stimulation consistent with traditional responses from humans, given the human capacity for contemplation, judgment and intention."[20] According to West and Allen, "AI depends on data that can be analyzed in real time and brought to bear on concrete problems. Having data that are 'accessible for exploration' in the research community is a prerequisite for successful AI development."[21]

According to Shubhendu Shukla and Vijay Jaiswal, AI applications "make decisions which normally require human level of expertise" and help people anticipate problems or deal with issues as they arise.[22] Thus, AI applications act purposefully, intelligently, and adaptively.

After discussing some of the theoretical definitions, it is appropriate to examine how organizations involved in research and development or the regulation and legislation of AI practically define it. Despite DARPA's general definition of AI as a "programmed ability to process information," it needs to be clarified that not every computing system uses AI. AI algorithms are designed to make decisions and do so by using data entered in real time. When they are used on different systems, these are not passive machines capable of mechanical or preset reactions only, as in the era of automation (such as automatic doors or even automatic functions in the washing machine); rather these are machines with sensors, digital data, and even remote inputs, which can integrate the information from various sources, analyze it immediately, and act according to the insights based on the data. This enables a sophistication and speed in accepting the data that was not previously possible.[23]

As far as the US government is concerned, there is no official definition of AI, and various agencies may define it differently, according to their needs. However, a series of laws that regulates the US Department of Defense

This content downloaded from 212.37.28.30 on Fri, 26 Jan 2024 11:12:27 +00:00
All use subject to https://about.jstor.org/terms

budget (FY2019 National Defense Authorization Act) provides a definition of AI for the enactment of section 238, which engages in the research and development of the field:

- Any artificial system that performs tasks under varying and unpredictable circumstances without significant human oversight, or that can learn from experience and improve performance when exposed to data sets.
- An artificial system developed in computer software, physical hardware, or other context that solves tasks requiring human-like perception, cognition, planning, learning, communication, or physical action.
- An artificial system designed to think or act like a human, including cognitive architectures and neural networks.
- A set of techniques, including machine learning that is designed to approximate a cognitive task.
- An artificial system designed to act rationally, including an intelligent software agent or embodied robot that achieves goals using perception, planning, reasoning, learning, communicating, decision making, and acting.[24]

This definition is quite detailed and is indeed suitable for legislative purposes. It also helps, in comparison to other definitions, to decide which areas of programming and computing do not belong to the discipline of AI. Nevertheless, it is too long and technical. Given the purpose of this document—to make knowledge about AI accessible to decision makers and to recommend policy in the national security sector—this study needs a shorter and simpler definition such as DARPA's. DARPA's definition is more appropriate for the purposes of this research than Minsky's, for example, because it does not relate to the controversial issue of human intelligence, and, in fact, it allows for a variety of currently accepted applications or processing methods and even leaves an opening for future developments, without the burden of technical details that requires expertise to understand. Even if this definition is likely to include "inferior" capabilities of computing and processing, as explained above, some of the methods and perceptions of the first wave are still useful in various fields and applications and therefore valuable.

However, in cases where decision makers must narrow the definition in order to examine whether a development meets the definition of AI or not,

This content downloaded from 212.37.28.30 on Fri, 26 Jan 2024 11:12:27 +00:00
All use subject to https://about.jstor.org/terms

AI can be referred to as being able to create knowledge and insights that had not existed before, using information and relying upon machines and computers. Focused on the programmed ability to process information, this definition distinguishes between a significant part of AI applications and general computer applications and narrows the general definition in such a way that it still covers a large number of applications and a wide range of disciplines, while emphasizing the creation of new knowledge.

Therefore, the guiding definition of AI used here is **using information and computer systems to present behavior that appears intelligent, or to create knowledge and insights that never existed before**. This definition is broad enough to include various technologies and applications and different kinds of needs to realize these abilities. At the same time, this definition is narrow enough that it does not include all areas of computing, but only those in which properties of AI are expressed. This definition helped to formulate the following chapters.

This content downloaded from 212.37.28.30 on Fri, 26 Jan 2024 11:12:27 +00:00
All use subject to https://about.jstor.org/terms

This content downloaded from 212.37.28.30 on Fri, 26 Jan 2024 11:12:27 +00:00
All use subject to https://about.jstor.org/terms