Intelligence from the perspective of inference. The problem of abduction for general artificial intelligence

Mihail DUCA
Doctoral School of Humanities
Moldova State University
ducii@mail.ru

Abstract

In this article, we want to analyze some of the limits of artificial intelligence. In order to have a significant development of this type of technology, one of the problems that requires a fundamental approach is the problem of inference. We will explore three types of inferences: deduction, induction, and abduction. Deduction was the basis for developing classical Artificial Intelligence (expert or symbolic system). Induction is used by modern Artificial Intelligence (neural networks), which requires a massive amount of data and very high computing powers. General Artificial Intelligence (currently in research) needs a theory of abduction that experts do not yet know how to program into a computational system.

Keywords: *inference, deduction, induction, abduction, reasoning, knowledge, Artificial Intelligence.*

Introduction

Today's technology has reached an advanced level that astounds us in numerous ways. In particular, Artificial Intelligence (AI), a broad term encompassing various types and research domains, has yielded remarkable results, fueling the perpetuation of myths surrounding sentient machines. This era is witnessing the revolutionary impact of AI, not only shaping our actions but also reshaping our understanding of ourselves. The predictions of artists, media, and renowned researchers regarding the near and distant future seem incredibly fantastical. While current achievements in weak AI, such as Deep Blue, Watson, AlphaGo, and ChatGPT, provide insights into the potential of strong or general AI, they also mask the existing barriers and limitations faced by AI today.

In this article, we aim to delve into the analysis of AI's limitations, which can temper the exuberant optimism of those promoting a highly technologized future. Doing so will situate ourselves in a discussion about the present reality. Unlike half a century ago, when AI systems were far less impressive than they are today, and limitations were seen as temporary and solvable, the omnipresence of AI in our lives now allows us to witness real-time testing of its inherent gaps. Consequently, it becomes evident that contrary to initial beliefs, the limitations of AI stem from fundamental aspects of knowledge. One crucial issue that necessitates a fundamental approach to advancing AI technology is the problem of inference. Eric J. Larson accurately argues that we currently lack promising ideas to efficiently program the specific types of inferences necessary for AI to perform tasks at a level comparable to human general intelligence. Larson concludes, “The problem of inference is central to the debate about artificial intelligence because it directly relates to intelligence, both humans and machines” (Larson, 2022, p. 12). While we acknowledge the existence of other equally significant challenges, this article will specifically focus on the problem of inference.

Inference towards the best explanation (abduction)

Firstly, let us clarify the meaning of the concept of inference. According to a more general definition provided by the Encyclopedic Dictionary of Philosophy, inference is the mental process through which one draws conclusions from one idea to another (Clement, 1999, p. 247). In simpler terms, inference is a type of reasoning that can be either deductive, where the conclusion is logically necessary (e.g., syllogism), or inductive, where the conclusion is based on probability. While some sources consider inference and deduction to be synonymous, it is evident that deduction is merely a specific type of inference. Moreover, deduction and induction are not the sole forms of inference. “The Dictionary of Philosophy of Knowledge” (volume II), following a different general definition of inference, provides an explanation of a distinct type called inference towards the best explanation, also known as abduction. This form of inference serves as a legitimate non-deductive reasoning that presents an alternative to both deduction and enumerative induction (Dancy, 1999, pp. 36-40). Abduction can be represented in the following structure:

O - occurs.
If E had occurred, we would have expected O.
Therefore, it is highly plausible that
E - occurred.

Let’s consider a simple example to illustrate this form of argument. Imagine an experienced hunter who comes across wolf tracks in the snow. Based on his expertise, he may infer that a wolf passed through that area because he knows from experience that wolves are likely to leave such distinctive tracks, unlike those of bears or foxes. These types of abductive inferences are common in our daily lives, and we make them effortlessly without much intellectual effort. In fact, they dominate many of our reasoning processes. However, it's important to note that while abduction is a valid form of reasoning, it can still lead to incorrect conclusions, similar to deduction or induction. This is because there are often numerous alternative explanations for a given event, such as the wolf tracks in the snow.

To demonstrate this, let's imagine another scenario where a mischievous hunter decides to play a prank on their friend. He creates a template of wolf tracks and imprints them on the snow. In this case, the inference that a wolf made those tracks would be incorrect. Although the hypothesis of the prankster hunter could be true, the experienced hunter would find it unlikely because wolf tracks are smaller than human tracks, and it would be highly unlikely to find only wolf tracks without any accompanying human footprints. This peculiar situation challenges us to question how such a scenario is possible. However, if we consider the possibility of bear tracks or tracks from a larger animal, and the mischievous hunter wore the templates while walking through the snow without leaving human footprints, then the hypothesis of the prankster becomes more plausible.

Thus, a valid abductive inference requires us to consider a wide range of factors that depend on our existing knowledge. The more we know about a particular subject, such as internal combustion engines, the better equipped we are to observe details and make accurate inferences. For instance, a skilled mechanic can use his hearing alone to infer the likelihood of certain engine malfunctions. On the other hand, someone without knowledge of engines would be unable to make the same inferences. While not everyone possesses expertise in hunting or mechanics, every person possesses a substantial amount of general knowledge that enables them to make various inferences within their specific knowledge domains.

For example, based on our experience, we can easily distinguish between a cat and a dog. When we see a cat, we instinctively recognize it without much contemplation. However, based on additional observations, we may also infer that this particular cat is our neighbor's favorite because we have frequently seen it on their balcony. Furthermore, if we notice the cat outside, limping, and considering that we haven't seen it wandering outside before, we might deduce that it likely fell from the balcony. Children can even perform these relatively simple reasoning processes, yet they can present challenges for AI programs. In certain cases, how these difficulties can be overcome is not evident. For instance, image recognition programs rely on massive datasets and deep learning algorithms to identify objects like cats in images. Although these programs have achieved impressive performance, they can still be easily misled by manipulating a small number of pixels, resulting in erroneous classifications.

At this moment, we will not go into the specific technical details of how such programs make errors. Similarly, we will not explore the notion that minor pixel modifications do not pose any challenges for human vision in correctly classifying an object in an image.

It is essential to highlight that humans possess not only simple vision (where the concept of a cat is not necessary to see one) and associative vision (connecting visual input with mental faculties to conclude it's a cat and not a dog) but also deliberative vision. Through deliberative vision, humans make abductive inferences, for instance, speculating about the real-life circumstances of the identified objects. Consequently, despite the impressive performance of machines in image recognition, they have no understanding of the actual existence and context of the recognized objects.

One might argue that even humans lack knowledge about the real-life aspects of a person they encounter for the first time. While this is true, humans can still make plausible hypotheses about that person based on observations (e.g., clothing, manner of speech). Initially, these hypotheses can be either true or false. However, our inferences become increasingly precise as we gather more knowledge about the person. This process is an inherent part of our daily lives. We continually form inferences, some of which are trivial, and it is worth noting that many of these inferences cannot be easily converted into algorithms.

The Role of Abduction in Science

Aside from the mentioned usual inferences, there are numerous examples of inferences in the history of science. One notable event in modern science is the Copernican Revolution, which brought about a paradigm shift by replacing the geocentric model with the heliocentric model (Koestler, 1995). This historical fact may not seem particularly significant today, as even young children learn early on that the Earth revolves around the Sun. However, in the early 16th century, when Copernicus was promoting his heliocentric hypothesis, such an idea was considered scandalous and ignited intense discussions. It is intriguing to understand how Copernicus arrived at formulating this hypothesis when, using terms employed in computer programming, the majority of the available data seemed to prove the exact opposite and, consequently, support the Ptolemaic system.

First and foremost, Copernicus needed considerable courage to reinterpret a vast amount of data accumulated over centuries. He did not wholly discard everything that was known at the time. On the contrary, by observing the complexity of the Ptolemaic system and studying certain ancient Greek philosophers (such as Philolaus and Seleucus of Seleucia) who were not part of the mainstream philosophers of that era, Copernicus made an inference towards the simplest explanation. This inference led him to propose heliocentrism as a model capable of replacing geocentrism. It assisted him in explaining certain astronomical phenomena more simply and logically, even though the Copernican system was not yet fully developed or proven. Within the new model, certain phenomena remained uncertain for the time being. What we want to underscore through the narrative of the development of the Copernican system is that scientists consistently put forward various hypotheses that, at that moment, are not proven but are subsequently subjected to multiple tests to be either confirmed or refuted. This process is crucial for the further advancement of science.

Once again, it is clear that we currently lack the knowledge to design machines capable of exhibiting intelligent behavior similar to humans. This becomes even more complex when we consider that many scientific hypotheses do not arise automatically from the vast amount of accumulated knowledge, despite the importance of big data for deep learning algorithms. The sheer quantity of data does not guarantee the emergence of new information, and the abundance of information available today does not inherently lead to the production of knowledge. Remarkably, there are instances where significant breakthroughs occur in the minds of scientists, sudden flashes of inspiration. In his book "We, the Particle and the World" (Chapter 6 – "The Vision of Reality and the Reality of Vision" (Nicolescu, 2002, pp. 114-136)), the distinguished Romanian physicist Basarab Nicolescu presents numerous examples of such enlightening moments that have unexpectedly propelled the progress of science. Nicolescu emphasizes the complexity of the scientific creative process, stating that while technical and partial scientific results are often achieved through the rigorous development of specific formalisms, the fiery imagination plays a predominant role in the grand pursuit of scientific invention, surpassing the unyielding calculations of scientific logic. This erroneous conflation of human intelligence with computational capabilities not only oversimplifies the concepts of intelligence and humanity but also undermines the very essence of scientific inquiry.

In 1931, the renowned Austrian logician Kurt Gödel published the proofs of two groundbreaking theorems in mathematical logic, commonly referred to as the incompleteness theorems, which exposed the limitations of formal mathematical systems. Gödel's first theorem demonstrated that within a minimal formal system encompassing arithmetic, there exist propositions that are true but cannot be proven within the confines of that system. The philosophical implications of this theorem, particularly in the context of artificial intelligence, highlight the impossibility of achieving complete formalization of human thought. Gödel's work revealed a striking insight: provability and truth are distinct concepts. Consequently, it becomes evident that the algorithms employed in AI cannot produce all possible truths merely by following formally correct rules. Gödel's theorems shed light on inherent limitations associated with formal systems and, by extension, present formidable barriers to achieving Artificial General Intelligence comparable to human intelligence. Furthermore, as with Copernicus, the fact that these groundbreaking ideas originated in the minds of brilliant individuals attests to our unique human capacity to perceive phenomena that technology alone cannot grasp.

The limitations of deduction in the development of intelligent machines

What exactly is deduction? According to the "Dictionary of Philosophy and Logic," deduction refers to a "valid reasoning process in which it is impossible, without contradicting oneself, to affirm the premises and deny the conclusion" (Flew, 1999, p. 88). In essence, deduction tells us that if the premises are true, the conclusion must also be true. This type of reasoning has been extensively utilized in classical AI, which, although it has not made significant strides in achieving general intelligence, has played a crucial role in the advancement of AI. Deductive-based artificial intelligence programs have been able to generate automated proofs for various mathematical theorems. Additionally, deduction has been employed in error-checking, logical consistency verification, and even computer manufacturing. Deduction has proven more beneficial in these domains than modern systems relying on statistics and machine learning. Therefore, while we cannot disregard the use of deduction in AI, it is important to recognize its limitations when striving for general intelligence.

One of the primary issues to consider is that certain conclusions can be formally correct but, if based on false premises, they end up asserting nonsensical statements. Let's consider a well-known example where both the premises and the conclusion are true:

If it is raining, the streets are wet.
It is currently raining.
Therefore, the streets are wet.

However, suppose we modify one or both premises in such a way that they become false in relation to reality. In that case, we obtain a formally valid argument with an evidently false conclusion:

If it is raining, pigs can fly.
It is raining.
Therefore, pigs can fly.

This example highlights that strictly adhering to logical rules does not necessarily guarantee the attainment of truth. The human mind can easily discern the falsehood of the first premise because our collective experience informs us that there is no connection between rain and pigs flying. Moreover, it is blatantly evident that pigs cannot fly. Despite the argument being valid in its structure, it is, in reality, an absurdity.

Another important aspect to consider relates to the concept of relevance. If the premise of flying pigs is considered absurd, then a statement claiming that if it's raining outside, the boss goes to work becomes irrelevant. While there may be instances where this premise holds true, such as when it's raining and the boss does indeed go to work, there is no direct causal connection between rain and going to work. In fact, in many cases, rain might serve as a reason for not going to work. Numerous examples can be provided to illustrate deductive logical constructions that are valid but lack relevance.

Let's envision an AI system that incorporates a database and deduction-based rules. Applying such a model might not be inherently problematic, but it is crucial to recognize that the program lacks understanding and cannot differentiate between what is relevant and what is absurd. Thus, machines relying solely on deduction are oblivious to these errors and necessitate human experts to ensure their proper functioning.

The above example also highlights another challenge of deductive reasoning, namely, the problem of knowledge. As Larson suggests, "deduction never adds any additional knowledge" (Larson, 2002, p. 112). The knowledge already exists within the stated premises, and deduction simply reaffirms a conclusion. However, the issue of knowledge, intertwined with common sense, warrants separate attention (often referred to as the problem of the bottomless bucket). Knowledge plays a pivotal role in designing general intelligence. It is important to emphasize that relying solely on deductive inferences is far from sufficient for significant advancements in AI.

Can induction be of any assistance?

The emergence of the internet and, subsequently, social media platforms has led to an astonishing phenomenon: an explosion of data accessible to researchers in the field of AI. Prior to this development, the AI field was divided into two camps with different approaches. Some sought to teach computers to emulate intelligent behavior through the encoding of specific logical rules (deduction) in what is known as classical AI, employing expert systems or symbolic approaches. On the other hand, proponents of modern AI aimed to recreate the human brain itself, focusing on neural networks. Despite neural networks being introduced as early as the 1960s, they did not achieve the same level of success as classical AI, leading to what is referred to as an "AI winter" in the 1980s. The history of AI has witnessed multiple such winters, characterized by decreased interest and funding for AI projects.

However, with the vast amount of data available on the internet and the advancements in computing power, neural networks have gained prominence in AI research. The massive volume of data is crucial for training neural networks, enabling them to recognize patterns effectively. Present-day researchers have access to extensive databases with a wealth of examples, and the current computing capabilities allow for rapid analysis. Consequently, deep learning-based Artificial Intelligence has emerged and achieved remarkable performance. Nevertheless, it is essential to emphasize that we are still far from achieving a technology of general intelligence.

While classical AI relies on deduction, modern AI is propelled by induction. Induction is a method of reasoning whereby general laws or principles are inferred from observed particular cases. Unlike deduction, which does not generate new knowledge, induction employs an enumeration mechanism and can acquire fresh insights from experiential data. Induction serves a vital purpose, not only by organizing the world through categorizable hypotheses but also by facilitating predictions. For instance, when observing 100 crows in nature, all of which are black, we cannot make a certain deductive inference that all crows in nature are black. However, we can make a probabilistic inference that most crows are black and, therefore, it is likely that all crows are black. In addition to drawing conclusions, induction allows us to predict that the next crow we encounter will probably be black. The level of certainty increases significantly with a larger sample size, such as observing a hundred thousand crows. However, an unexpected occurrence, such as encountering a white crow, undermines our inference about crows, as it is a rare event in nature. The ever-changing nature of our world implies that much of the knowledge gained through inductive reasoning is provisional. Even our current understanding of intelligence may evolve in the future.

However, the advancement of modern science would not have been possible without the utilization of induction. It is crucial to remember that scientific knowledge cannot be solely derived from induction alone, as relying solely on observed instances is not foolproof. As we have witnessed, induction can still lead to false conclusions even with true premises. Many authors use the "Russell's turkey" analogy to illustrate this point. Imagine a turkey being fed at the same time every day, leading it to assume that this pattern will always hold true. However, during a festive occasion, the turkey is slaughtered, shattering its inductive inference. In addition to the inherent fragility of induction (as new data may deviate from past observations), our preconceived notions based on past experiences can sometimes hinder us from perceiving new possibilities, as we tend to be strongly anchored to traditional patterns. Therefore, in order to achieve general artificial intelligence, just as with deductive inferences, inductive inferences are necessary but not sufficient.

Conclusions: Deduction, Induction, and Abduction

Through a brief analysis of these three types of inference, we can conclude that all of them are indispensable for general intelligence. While classical AI relies on deduction and modern AI on induction, a comprehensive theory of abduction is required for general Artificial Intelligence (which is currently in the developmental phase). However, experts are still grappling with how to incorporate it into computational systems. Since these are distinct types of inference, none of them can be transformed into another. We cannot expect deduction or induction alone to encompass abduction. Nonetheless, abduction is the core inference mechanism behind general intelligence. As we have demonstrated, humans consistently employ this type of inference in their everyday lives. The essence of human existence and what we refer to as common sense necessitate a nuanced understanding of the real world. Even the comprehension of seemingly straightforward matters is essential for individuals and society as a whole. The complexity of the surrounding reality renders formal systems insufficient for its comprehensive description. It is nearly impossible to amass a sufficient amount of elementary knowledge in a machine to enable it to perform the myriad of real-life tasks that humans excel at. Our intelligence relies on all possible senses, including intuition, to make inferences about our environment. Ray Kurzweil posits that by the end of the third decade of this century, we will have Artificial Intelligence capable of attaining human-level general intelligence and subsequently surpassing it with superintelligence. Despite the predictions made by visionaries and futurists like Kurzweil, we are currently uncertain about how to reach such a level, as solving the problem of inference is one among numerous challenges that need to be addressed.

References:

1. Clément, E., Demonque, C., Hansen-Løve, L., & Kahn, P. (1999). *Filosofia de la A la Z* [Philosophy from A to Z] (M. Cojocea-Mărculescu & A. Cojocea, Trans.). București: All Educational.
2. Dancy, J., & Sosa, E. (1999). *Dicționar de Filosofia Cunoașterii.* [A Companion to Epistemology] Vol. II (G. Ștefanov, C. Matei, A. Zagura & D. Bărbulescu, Trans.). București: Trei.
3. Flew, A. (1999). *Dicționar de filozofie și logică* [A dictionary of philosophy] (2 ed). București: Humanitas.
4. Koestler, A. (1999). *Lunaticii. Evoluția concepției despre Univers de la Pitagora la Newton* [The Sleepwalkers: A History of Man's Changing Vision of the Universe]. București: Humanitas.
5. Larson, E. J. (2022). *Mitul inteligenței artificiale: de ce computerele nu pot gândi la fel ca noi* [The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do]. Iași: Polirom.
6. Nicolescu, B. (2002). *Noi, Particula și Lumea* [Us, the particle and the world]. Iași: Polirom.