proceedings

MDPI

Proceeding Paper

Advancing Human Understanding with Deep Learning Go AI Engines †

Attila Egri-Nagy [1,*] and Antti Törmänen [2]

[1] Department of Mathematics and Natural Sciences, Akita International University, Yuwa, Akita City 010-1292, Japan
[2] Nihon Ki-In—Japan Go Association, 7-2 Gobancho, Tokyo 102-0076, Japan; antti@tormanen.net
* Correspondence: egri-nagy@aiu.ac.jp
† Presented at the 13th International Workshop on Natural Computing (IWNC), IS4SI Summit 2021, Online, 12–19 September 2021.

Abstract: Humans mainly learned from other human beings for thousands of years. Recent advancements in artificial intelligence (AI) seem to have changed this setup. Due to deep learning, we now have access to automatically generated high-quality statistical knowledge beyond human expert intuition in many fields. However, the representation is not human-friendly: an opaque mass of pure associations instead of narrative, causal explanations. Here, we investigate the epistemological problem of using AI data for human understanding and suggest an active approach based on the scientific method. Following tradition in AI, we focus on a game. Go is a well-defined problem domain that is complex enough that our approach may provide solutions in other fields of knowledge, too, where AI technology outperforms humans.

Keywords: deep learning; artificial intelligence; epistemology; game of Go

Citation: Egri-Nagy, A.; Törmänen, A. Advancing Human Understanding with Deep Learning Go AI Engines. *Proceedings* **2022**, *81*, 22. https://doi.org/10.3390/proceedings2022081022

Academic Editors: Marcin J. Schroeder, Masami Hagiya, Yasuhiro Suzuki and Gordana Dodig-Crnkovic

Published: 10 March 2022

1. Introduction

In ancient myths, humankind dreamt of the possibility of obtaining knowledge from a more cognizant, divine source of information. However, so far, we only had ourselves, and thus we developed science to learn from experience and from each other. In our technological society, scientific investigations are the most important source of knowledge (i.e., information with causal power [1]). Though they are philosophically debated, in science we have established practices for obtaining and verifying knowledge.

What if an oracle suddenly appeared? What if we had instant answers without carrying out long experiments? How would that change our scientific practices? This is precisely what happened recently in several fields due to the advances in AI technologies, and now we have to re-examine and possibly extend some of our epistemological methods.

1.1. The Black-Box Nature of Deep Learning AIs

The progress in deep reinforcement learning, the *artificial neural networks* [2] combined with *reinforcement learning* [3] and supported by the growth of computational power yielded several breakthroughs in speech recognition, object detection and recognition, drug discovery, genomics, and in numerous other domains. However, the AIs operate in a black-box manner: we get a correct answer with high probability, but with no explanation. This causes several problems, including the one we address here: the AIs do not directly contribute to *human* knowledge and understanding. We assume that improving human skills could lead to more meaningful lives, but this worthy application of AIs is hampered by their opacity. Therefore, we ask:

> How can we learn from AIs?, i.e., How can we understand a complex problem domain better with access to high-quality statistical patterns?

Deep learning AIs model human intuitive knowledge: knowing what without knowing how. Therefore, we are interested in verbalizing implicit knowledge represented as a vast pool of pure associations. *Our tentative answer is that humans, as AI users, should do more cognitive work and apply a scientific method,* which goes against the commonsense idea of AIs doing the thinking instead of us.

To answer the above question we need to investigate how existing epistemological approaches can operate in the presence of non-human knowledge (e.g., the output of the AlphaGo Zero [4] and AlphaZero [5] algorithms that use no human-played games as inputs for training) and to do an empirical study of what people instinctively do in this situation and provide improved methods for deepening human understanding. In general, this task is enormous. Here we focus on a single application domain: the game of Go. This choice follows the tradition of research in artificial intelligence: before tackling real-world applications, one works with a game, which is a little world in itself.

The Go world had a rapid and unexpected transition from human supremacy to superhuman AIs. In 2016, AlphaGo concluded the mission of besting human players in abstract board games. Then came the real revolutionary change: the technology went open-source and became widely available for everyone. Crucial challenges remain: Go AIs show good moves and their principal variations (the predicted follow-up sequences) are expressed non-verbally as board, tree, and graph diagrams. Lacking an explanation why a move is good, the situation is similar to having a teacher who does not speak, therefore restricting communication to pointing and general gestures. However, humans cannot simply memorize all the good moves, as there are far too many, and they depend on the surrounding context. *Improving in Go involves learning higher-level explainable concepts and patterns that we apply in different positions.*

1.2. AIs and Humans

The current trends in the field of artificial intelligence focus on the explainability of AI decisions. The deep-learning paradigm is missing the critical ingredient of understanding cause and effect. The *causal AI* approach aims to combine deep learning with causal reasoning [6]. In the case of Go, a causal AI would supply an explanation bundled with the best move recommendation.

Currently, much of the public discussion is about AIs versus humans, contrasting their performances, neglecting *'human plus AI'* combinations, the collaborative approach. One of the identified risks of a possible general AI was potential obsolescence, loss of human competence [7]. However, people and computers thinking together can have surprising power [8] and this combination was suggested for chess [9]. We advocate the less pessimistic view by increasing human competence (Figure 1).

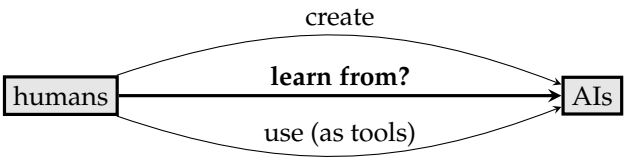

Figure 1. We suggest to add a new relationship to the human–AI interaction repertoire. We view AI tools as a way to realize accelerated education.

1.3. AIs and the Go Community

The iconic move 37 in the first game of the historical Lee Sedol vs. AlphaGo match showed a new kind of machine creativity. Deep learning neural networks opened up a novel source of knowledge. We think that the game is not over yet for humans [10], and consider this development as an opportunity. Pairing human skills (e.g., high-level reasoning, thought experiments, counterfactuals, explanations, narratives) with an unlimited supply of high-quality statistical experiments is indeed a new situation. The widespread availability of AI technology for Go players can be viewed as a grand-scale collective intelligence experiment. *What do people do with these new tools? How efficiently do they use or misuse them*

(e.g., cheating)? How do they talk about their methods? These questions are worth investigating as the 'experiment' progresses. Ideas in this paper are also motivated by observations of people using the newly available software tools.

2. Possible Solutions for the Black-Box Problem

There are two different approaches for using AIs as tools for learning about the game of Go. We can try to open the black box of the neural network or embrace its opaque nature by using only the input–output pairs.

- Internal analysis of the neural networks—intelligible intelligence: In deep learning AI systems, unlike in the brain, we have complete access to the whole neural network, down to single neurons. We can uncover the abstract hierarchical representation of Go knowledge inside the network by using feature visualization used for image recognition networks [11]. However, we know that neural networks may or may not have comprehensible representations. The space of possible Go-playing neural networks may have a vanishingly small fraction in human-accessible formats. Despite the difficulties, this work is underway now for chess [12].
- Improve our learning methods: Learning the game at the professional level proceeds from intuitive understanding to explicit verbalizations. In the case of Go, strategic plans are explanations for what is happening on the board. Therefore, the methods of scientific knowledge creation do apply here. Go AIs are inexhaustible sources of experiments providing high-quality statistical data. Growing Go knowledge can be faster by formulating plans when choosing moves rather than just looking up the best move recommended by the AIs.

The black-box problem is possibly a temporary one. If there was a fundamental reason why we cannot extract meaningful information out of neural networks, then understanding our brain would also be hopeless. However, here we take the more direct, immediately applicable second option.

3. The Requirements of the Teaching–Learning Situation

When is learning possible? Who can we learn from and under what circumstances? After a commonsense analysis of the teaching–learning situation, we identify the following three conditions.

1. The teacher should know more—otherwise, what can we learn?
2. The teacher should be able to explain.
3. The student should be able to comprehend.

Next, we can see whether a Go-playing AI engine can satisfy these conditions. We consider a single human player using a computer, trying to improve her understanding.

1. The AI knows more, as it has superhuman playing strength. This statement might be challenged by a standard argument that neural networks do not 'know' anything since they are just big tables of numbers, and the AIs' thinking is just matrix multiplication. Be it a simple mechanism, its winning record neutralizes ontological complaints.
2. The AI does not explain. It only gives probabilities and score estimates.
3. How can a student then achieve understanding? It is up to the student to actively create situations where learning is possible.

A main advantage that human teachers have over an AI right now is that they can adjust the teaching content to the student. A good teacher can more or less guess what the student is thinking, can pinpoint what their understanding lacks, and then can explain concepts by working from what the student knows and understands. This problem may be easier to solve than the lack of explanation: a well-engineered deep learning AI could be taught to spot deficiencies in human understanding and use general teaching patterns.

4. Acting as a Scientist

To give our tentative solution to the problem of missing explanations, we will lean on the more than a century-old discussion of the philosophy of science [13]. We can adopt a natural form of *scientific realism*, since the game–world analogy does not transfer ontological issues (perfect knowledge exists in the complete game tree). We start from the fact that the human mind has a better cognitive architecture. We have explanations, reasoning, the ability to form hypotheses and test them. In contrast, Go AIs (AlphaGo and subsequent open-source implementations) are still purely associative structures, mapping board positions to estimated results and good moves. From this perspective, it is somewhat ironic that a narrow AI advanced beyond the only general intelligence currently in existence, the human intelligence.

We concentrate on the activity of creating explanations, motivated by [1], which in turn has its origin in Popperian epistemology. Explanations answer *Why?* questions. We draw a parallel between a scientific investigation and active game analysis.

Science We create explanations and test them by observations, experiments.

Go We create plans based on judgments of the current board position. They are verified or falsified by the opponent, or by the analyzing AI.

A plan is like an explanation: it gives meaning to individual moves. Often, an expert player will first reason out the effect she wants to achieve. Then she calls upon her intuition to find good moves and finally spends time checking those. This method resembles falsification [14]. One has to check variations from the perspective of the opponent, whose task is to find ways to show that the plan is not working. All scientific theorems are conjectural—says the Popperian approach. However, in the case of Go, we know that the ultimate truth exists in the form of the complete traversal of the game tree, though it may not be realizable [10].

An active game analysis can be likened to a series of experiments. We suggest that an AI-based game review should start with forming hypotheses: *Why did I lose the game? Was my board judgment or plan mistaken? Would that other move I considered have fared better?* Then, and only then, one should look at the output of the AI engines, since understanding happens in the context of a plan. But why do plans have this central role for human players?

5. Game Review as Storytelling

It is increasingly clear that stories are not some entertaining decorations of life, but the ability to understand stories is in the fundamental structure of the human mind. Trying to understand what makes a good story, and what the authors should write, led to interesting discoveries in psychology and cognitive science [15–17]. One shocking finding is that our past is a similarly constructed story as our planned future.

The difference between a story and a sequence of events is the same as between a game analysis and a game record. A good analysis has a narrative for how the subsequent moves are connected together. It has the shape of a story with a protagonist which can be a single stone, a particular shape, or the player herself. A game analysis also talks about events on the board that did not happen. It includes reasons for certain moves not appearing on the board by using counterfactuals [6,18].

We can ask what makes a particular game interesting. For humans, it is not the optimal move in each turn, but the battle of ideas. How the strategies of the players pan out, and how they influence each other. It is the excitement of 'What happens next?'. Thus, another way to look at plans is to consider them as stories. We conceptualize games as stories since that is what humans naturally do. Narratives seem to be a genuinely higher level of description of games, and therefore it will be interesting to see whether they can be somehow extracted from the raw neural network data.

6. Discussion

We argued that we can treat deep learning AIs as black boxes (as they appear now) and still improve our understanding of the problem domain. Instead of waiting for science to uncover the concepts stored in the neural networks, we can continue building our personal representations. We propose that the hypothetico-deductive model can help in gaining knowledge from the raw deep learning neural network output.

More and more people will work with AIs in the near future. AI tools are finding ways into the top achievements of the human intellect, most recently mathematical discovery [19]. Thus, it is crucial to have positive examples of AI–human collaboration. The ultimate goal is accelerated education, where people can learn faster and understand concepts deeper in any field of knowledge.

Author Contributions: All authors contribute equality to the article. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Deutsch, D. *The Beginning of Infinit: Explanations That Transform the World*; Penguin Publishing Group: New York, NY, USA, 2011.
2. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
3. Sutton, R.; Barto, A. *Reinforcement Learning: An Introduction*, 2nd ed.; Adaptive Computation and Machine Learning Series; MIT Press: Cambridge, MA, USA, 2018.
4. Silver, D.; Schrittwieser, J.; Simonyan, K.; Antonoglou, I.; Huang, A.; Guez, A.; Hubert, T.; Baker, L.; Lai, M.; Bolton, A.; et al. Mastering the game of Go without human knowledge. *Nature* **2017**, *550*, 354–359. [CrossRef] [PubMed]
5. Silver, D.; Hubert, T.; Schrittwieser, J.; Antonoglou, I.; Lai, M.; Guez, A.; Lanctot, M.; Sifre, L.; Kumaran, D.; Graepel, T.; et al. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science* **2018**, *362*, 1140–1144. [CrossRef] [PubMed]
6. Pearl, J.; Mackenzie, D. *The Book of Why: The New Science of Cause and Effect*; Penguin Books: London, UK, 2018.
7. Tegmark, M. *Life 3.0: Being Human in the Age of Artificial Intelligence*; Knopf Doubleday Publishing Group: New York, NY, USA, 2017.
8. Malone, T. *Superminds: The Surprising Power of People and Computers Thinking Together*; Little, Brown: Boston, MA, USA, 2018.
9. Kasparov, G. *Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins*; Millennium Series; Hodder & Stoughton: London, UK, 2017.
10. Egri-Nagy, A.; Törmänen, A. The Game Is Not over Yet—Go in the Post-AlphaGo Era. *Philosophies* **2020**, *5*, 37. [CrossRef]
11. Olah, C.; Mordvintsev, A.; Schubert, L. Feature Visualization. *Distill* **2017**, *2*, e7. [CrossRef]
12. McGrath, T.; Kapishnikov, A.; Tomašev, N.; Pearce, A.; Hassabis, D.; Kim, B.; Paquet, U.; Kramnik, V. Acquisition of Chess Knowledge in AlphaZero. *arXiv* **2021**, arXiv:cs.AI/2111.09259
13. Godfrey-Smith, P. *Theory and Reality: An Introduction to the Philosophy of Science*; Science and Its Conceptual Foundations Series; University of Chicago Press: Chicago, IL, USA, 2009.
14. Popper, K.; Popper, K. *The Logic of Scientific Discovery*; ISSR Library, Routledge: London, UK, 2002.
15. Cron, L. *Wired for Story: The Writer's Guide to Using Brain Science to Hook Readers from the Very First Sentence*; Clarkson Potter/Ten Speed: New York, NY, USA, 2012.
16. Gottschall, J. *The Storytelling Animal: How Stories Make Us Human*; Houghton Mifflin Harcourt: Boston, MA, USA, 2013.
17. Storr, W. *The Science of Storytelling: Why Stories Make Us Human and How to Tell Them Better*; ABRAMS: New York, NY, USA, 2020.
18. Marletto, C. *The Science of Can and Can't: A Physicist's Journey Through the Land of Counterfactuals*; Penguin Books, Limited: London, UK, 2021.
19. Davies, A.; Veličković, P.; Buesing, L.; Blackwell, S.; Zheng, D.; Tomašev, N.; Tanburn, R.; Battaglia, P.; Blundell, C.; Juhász, A.; et al. Advancing mathematics by guiding human intuition with AI. *Nature* **2021**, *600*, 70–74. [CrossRef] [PubMed]