Essay

Author: Reza Ghasemi

January 16, 2022

Abstract

Deep learning models have been able to not only match human performance but to surpass it in many tasks. Unfortunately, their great results are not extendable to novel situations. They lack many features that humans possess to tackle new problems, such as analogy, knowledge transfer, causal reasoning, etc. To achieve artificial general intelligence (AGI), we should implement innate human capabilities to enable machines to solve problems they have not seen before.

1 Introduction

Before delving into the obstacles preventing us from reaching artificial general intelligence (AGI), I believe it would be useful to take a look at the definition of intelligence, to have a general sense of what is intelligence, and to evaluate whether deep learning models can be considered intelligent or not.

Intelligence has been defined using different terms over the years. Even in the same period, various cultures had non-identical notions of intelligence. Intelligence could have meant: critical-thinking, reasoning, having a great memory, etc. Nonetheless, creating a single definition would be quite challenging, as intelligence is a multifaceted and complex phenomenon, and it cannot be reduced to the limits of working memory, pattern recognition, or games like chess or Go. Intelligence is not a single capability, but consists of several capabilities.

Neural networks do exceptionally well in terms of end-to-end mapping. The problem lies in finding a solution to novel unseen problems. Humans can transfer knowledge and apply their previous knowledge to new tasks, whereas deep learning models lack such features. If the environment is slightly changed, they would not perform well. The main challenge in AI is to build machines capable of learning and solving new problems.

2 Building Intelligent Machines

Initially, human intelligence was the inspiration for artificial intelligence, but currently there are substantial differences between the two. We have two approaches for computational models: the symbolic approach and the emergentist approach. In the symbolic approach (GOFAI), the rules are hand-coded explicitly and a vast amount of expert knowledge is required.

However, in the emergentist approach, representations are learned by receiving plenty of data. Being data hungry is noted as one of the main disadvantages of deep learning models [1]. Until recently, deep learning was not considered feasible due to limited data available and anemic computing resources. Publication of a few influential papers revolutionized the field, and deep learning became the standard.

Because of the relative success of artificial intelligence, people have overpromised over the years. In 1965, Herbert Simon famously claimed the following:

> “Machines will be capable, within twenty years, of doing any work a man can do” [2].

Geoffrey Hinton, a pioneer in AI, in 2016, declared the following:

> “We should stop training radiologists now, it’s just completely obvious within five years deep learning is going to do better than radiologists” [3].

Such high expectations with no delivery could be dangerous as they could cause another AI winter.

3 Shortcomings of AI

In modeling intelligence, humans cannot be objective. Since we do not have any other framework, we compare animals and machines to ourselves. To consider an animal intelligent, we look for common factors. For example, it was once thought that only humans possessed ”numerosity perception,” but it now appears that many animals share this trait with humans [4].

Observing deep learning models, it is evident that AI has not only been able to match human level performance but to outperform humans in many tasks like speech recognition, image recognition, and language translation. Recently, a solution to a grand challenge in biology (“protein folding”), was found by DeepMind [5]. Some research has also been conducted to use deep learning models to predict future events [6][7].

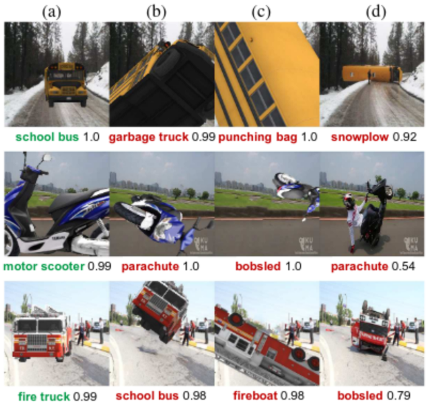

Figure 1: Inception V3 Model fails to recognize objects in different positions. Image credit: Alcorn, Michael A., et al. "Strike (with) a Pose: Neural Networks Are Easily Fooled by Strange Poses of Familiar Objects." arXiv preprint arXiv:1811.11553 (2018).

However, it appears that deep learning models are unable to extend their problem-solving abilities beyond what they have been trained on. Human-like intelligence is not bounded to pattern recognition, and humans are far more efficient concerning scene understanding or concept learning. Moreover, humans can generalize. For example, a kid seeing a tractor for the first time can understand that it is an object similar to a car. We also understand the logic behind a sentence, not just the pattern. This could be due to some "start-up software" we have, which acts as building blocks for abilities we acquire later in life. Such a function could boost learning tremendously, and implementing a similar feature in machines may get us closer to strong AI [8].

If we collected data from a hospital and used it to train a model, the same model could not be used for another hospital nearby. Humans, unlike deep learning models, can do such things. After observing a picture of a train, we can detect it. The angle or color will not affect our response. In deep learning models, if the color of the object is changed or the position is altered in any way, the model struggles.

From what we have discussed so far, it is evident that deep learning models are very shallow and cannot transfer knowledge to new scenarios. If the model is confronted with a new scenario with a slight configuration change, it might fail. Even minor perturbations can cause misclassifications. A study showed that Inception V3 model failed to correctly recognize objects if the position of the object was changed (Figure 1). The model must be retrained and reconfigured to correctly recognize the object.

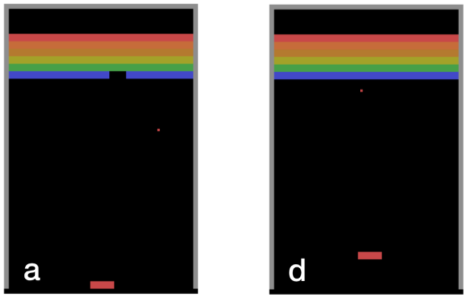

Figure 2: (a) The model performs well in the standard scenario (b) The position of the paddle is perturbed and the machine can no longer play. Image credit: Kansky, Ken, et al. ”Schema Networks: Zero-shot Transfer with a Generative Causal Model of Intuitive Physics.” arXiv preprint arXiv:1706.04317 (2017)

Another study, trained a model to play Breakout game, and it performed ten times better than humans. Then, the configuration was slightly perturbed. Here, the paddle was shifted a few pixels (Figure 2). The model could not play the game anymore. The model did not learn the concept of the paddle or the ball; it only learned some configurations, and when they were altered, the model became paralyzed.

Being data-hungry is another issue. Humans are capable of learning from a limited data set. They can learn richer representations and transfer them to new domains, whereas in “supervised learning” plenty of human-labeled data is needed. In the Frostbite game, a professional gamer was trained for only two hours on each ATARI game, whereas DQN had to be trained for 924 hours (Figure 3).

The term “learning” is debatable, given they cannot extend their knowledge to provide predictions for unseen sets. This becomes a major issue when we do not have access to abundant labeled data. Also, in specialized fields, there may not be a public or open data set. This means human annotators are required, and in some cases, this could raise an ethical question as annotators are poorly compensated. On Mechanical Turk website, annotators may earn less than 2 dollars per hour, which is less than the minimum wage defined by the U.S. Department of Labor, which is 7.25 per hour [9].

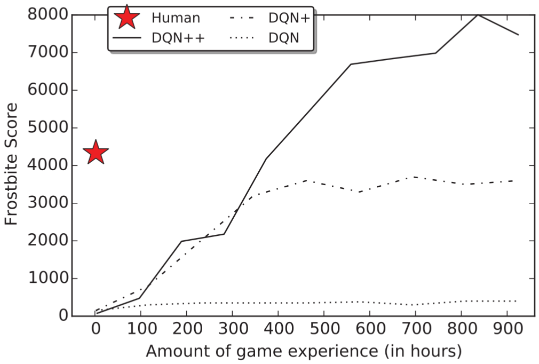

Figure 3: A comparison between a professional gamer and DQN in learning Frostbite games. Image credit: Lake et al. (2017). Building machines that learn and think like people. Behavioral and Brain Sciences, 40. https://doi.org/10.1017/s0140525x16001837

Additionally, explainable artificial intelligence (XAI) is also considered another point of interest since, in various cases, we may want to know how the model reached a certain decision. A patient may inquire as to why she or he would undergo surgery, or a customer may inquire as to why they are unable to obtain a loan. A 2018 UC Berkeley study showed a discrepancy in charging rates between black and Latino applicants compared with white borrowers [10]. Explainable AI would prevent similar scenarios in the future.

Another major concern is the absence of common sense. Common sense is knowledge that we do not state directly, but we expect another human being to know in order to interact and interpret the world. John McCarthy was the first person to propose common sense reasoning as a key ability in 1959 [11]. This ability is effortless for humans, but AI researchers have been struggling with it for a long time. Consider this sentence:

"A man went to a restaurant. He ordered a steak. He left a big tip."

Humans can easily infer that the man ate a steak, while machines struggle to understand what he ate. Common sense would reduce the quantity of data which is needed and could assist machines in discovering solutions for new situations.

4 The New Alchemy

Some believe that artificial intelligence has become present-day “alchemy” [12]. Researchers have little to no understanding of how the algorithms function. There are no criteria for choosing an architecture. They appear to be black boxes not only to the engineers implementing them, but equally to the machines. However, LeCun disagrees with this analogy, saying such comparisons are “insulting” and declares “engineering is messy”.

In his defense, one could argue that we use many systems daily that we don’t fully comprehend. But, we know entire fields are dedicated to such systems, and there are experts who can explain why certain functions work. Yet, even AI researchers have trouble explaining why the model works. The architecture is chosen through trial and error and intuition rather than understanding. This, in itself, should be a signal that a change is necessary.

The main factor determining whether AI will be abandoned like alchemy or remain with us is how we plan to approach it in the upcoming years.

5 Path Forward

One understudied area of AI is analogy. Using analogy, humans can solve new problems they have not seen before. Melanie Mitchell believes that analogy could be the missing key to reaching artificial general intelligence [13]. Such a mechanism would allow a machine to map prior experiences to new situations. Mitchell also thinks that to have a human-like analogy, it may be necessary to create a body. This would enable machines to perceive visual problems in three dimensions and to understand spatial relationships between objects [14].

For effective communication between machines and humans, machines require a sense of self. A physical body would allow a machine to understand other people. Thus, we may have to create a body for artificial intelligence. In this fashion, they could grasp the spatial relationship between objects and it could also be of benefit to “common sense” as it introduces a new passage for perceiving input from the environment [15].

One alternate suggestion to reach AGI would be to instill causal reasoning in machines. Judea Pearl believes it is not sufficient to simply learn the correlation between malaria and fever, but the machine must understand that malaria was its cause [16].

He believes modern techniques put far too much emphasis on correlation and curve fitting. A machine should be able to reason, given a certain action, what would have been the outcome.

If we want machines to conduct scientific experiments in the future, causal reasoning is needed. It could also contribute to better automated caption generation. Without causal understanding, machines will generate unreliable captions as they do not perceive the relationship between objects.

Recently, fathers of artificial intelligence have acknowledged the criticisms. Nonetheless, they do not view integrating symbolic artificial intelligence with deep learning models (hybrid AI) as a valid solution [17].

In a paper published by LeCun, Hinton, and Bengio, different methods that could aid deep learning models were discussed. For instance with "transformers", a model could learn without needing labeled data. Another promising technique worth mentioning is "contrastive learning" where instead of predicting exact values of a pixel, a vector representation of missing regions is found.

Another solution is "system 2 deep learning" and it could help with causal inference and transfer learning challenges. Note that this method is still in its early stages and more work needs to be done. Capsule networks are another area of research, where instead of focusing on detecting features, objects and their physical properties and relationships are detected. For instance, using capsule networks, it would be possible to add "intuitive physics" to deep learning models.

6 Conclusion

The progress of AI has led to numerous technological advancements. But to reach human-like intelligence, pattern recognition alone is not sufficient. New methods must be discovered to solve the current challenges.

For the time being, a valid solution might be to integrate cognitive abilities of humans with neural networks. Nevertheless, further research should be done in this area. Recent advancements such as transformers and contrastive learning appear very promising. In the future, we may resort not just to one, but a combination of techniques. Additionally, if we wish to integrate cognitive abilities with artificial intelligence, more collaboration between AI researchers and cognitive scientists must take place.

References

[1] Marcus, G. (2018). Deep Learning: A Critical Appraisal arXiv:1801.00631 cs.AI.

[2] Simon, H. A. (1977). The New Science of Management Decision. Prentice Hall.

[3] Geoff Hinton: On Radiology. (2016, November 24). [Video]. YouTube. https://youtube.com/watch?v=2HMPRXstSvQ

[4] (2021, August 12). Animals Can Count and Use Zero. How Far Does Their Number Sense Go? Quanta Magazine. https://www.quantamagazine.org/animals-can-count-and-use-zero-how-far-does-their-number-sense-go-20210809/

[5] Jumper, J., Evans, R., Pritzel, A. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021). https://doi.org/10.1038/s41586-021-03819-2

[6] Ratner, P. (2021, September 30). Secretive agency uses AI, human 'forecasters' to predict the future. Big Think. https://bigthink.com/the-present/secretive-agency-uses-ai-human-forecasters-to-predict-future/

[7] Leetaru, K. (2011). Culturomics 2.0: Forecasting large-scale human behavior using global news media tone in time and space. First Monday, 16(9). https://doi.org/10.5210/fm.v16i9.3663

[8] Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2016). Building machines that learn and think like people. Behavioral and Brain Sciences, 40. https://doi.org/10.1017/s0140525x16001837

[9] Minimum Wage. (n.d.). U.S. Department of Labor. https://www.dol.gov/general/topic/wages/minimumwage

[10] Bartlett, R., Morse, A., Stanton, R., & Wallace, N. (2019). Consumer-lending discrimination in the FinTech Era. Journal of Financial Economics, 143(1), 30–56. https://doi.org/10.1016/j.jfineco.2021.05.047

[11] McCarthy, J. (1960). Programs with common sense.

[12] Ali Rahimi - NIPS 2017 Test-of-Time Award presentation. (2017, December 6). [Video]. YouTube. https://youtube.com/watch?v=ORHFOnaEzPc

[13] Mitchell, M. (2021). Abstraction and Analogy-Making in Artificial Intelligence arXiv, cs.AI.

[14] The Computer Scientist Training AI to Think With Analogies. (2021, August 4). Quanta Magazine. https://www.quantamagazine.org/melanie-mitchell-trains-ai-to-think-with-analogies-20210714/

[15] Knight, W. (2020, April 2). Alexa needs a robot body to escape the confines of today's AI. MIT Technology Review. https://www.technologyreview.com/2019/03/26/136354/alexa-needs-a-robot-body-to-escape-the-confines-of-todays-ai/

[16] To Build Truly Intelligent Machines, Teach Them Cause and Effect. (2018, June 26). Quanta Magazine. https://www.quantamagazine.org/to-build-truly-intelligent-machines-teach-them-cause-and-effect-20180515/

[17] Bengio, Y., Lecun, Y., Hinton, G. (2021). Deep learning for AI. Communications of the ACM, 64(7), 58–65. https://doi.org/10.1145/3448250