AI: Forging tomorrow's double-edged sword

Subin Kim

March 29, 2023

1 Tower of Babel in 21st century

The renowned 'Tower of Babel' is a mythology about humanity challenging God with a massive tower. This tower was the ultimate tool for the people of Babylon with a common language to showcase their unity and craft. Note that God, infuriated as he was, would use 'language' as the ultimate tool to disperse the arrogant humanity and remove the proficiency. It is said that the people could no longer communicate effectively and were forced to abandon the project [2].

This is interesting, considering that the strongest point of one of the leading AI, Chat-GPT, is its linguistic skills. From the point of view of AI, language is not the past anymore, which indicates that the barrier of language is not hindering interaction. The ultimate role of Chat-GPT is to provide communication channel between people and computers [3]. 'Globalization' is not a recent term, but AI is propelling the speed and efficiency of the world uniting together.

Could we say that we are building the 'Tower of Babel' through artificial intelligence once more? In a biblical sense, the tale of 'the Tower of Babel' conveys a warning about excessive pride against divine authority. If the former proposition is true, to which heaven are we trying to touch, and what would be the consequences? The rate of progress is so fast that it is unpredictable.

2 The opportunity of AI

The essence of AI is pattern recognition, trained from a large set of data. Given a very large set of data points, artificial intelligence would try to fit a model that can represent the innate structures of the world [1].

Thus, any domain that has a predictable pattern of behavior can benefit from AI. Indeed, AI has enabled us to do a lot of things in multiple domains in an optimized, autonomous way. A famous example is 'AlphaFold' developed from DeepMind, which is an AI designed to predict the 3D structure of proteins based on their amino acid sequences [4]. Healthcare, finance, marketing, manufacturing, and e-commerce are all industries in that AI is being used actively. Increasing automation and optimization leads to more productivity and less need for labor.

Then what about creativity? Creativity is no exception, despite the early predictions that art will be the last stronghold of humanity against AI. The essence of creativity is grasping a pattern, breaking it down more delicately, and then reassembling it in a new way, which is something that AI specifically does.

Games such as Chess and Go are well-known examples. They have a limited number of possible states is a combination of predictable patterns, though very large [5]. AI can manipulate creative moves that often throw the opponent off-guard. This also applies to music and art. 'Deep forger' in 2015 can recreate any input image in the style of classic paintings such as Van Gogh or Picasso. 'Magenta' of Google composed an 80-second-long piano number [6].

3 Bias of data distribution

This is the impressive capacity of statistics: we can 'inference' the future from prior concepts, based on the belief that there is a certain 'pattern' existing in the interweaving rollouts of incidences [1]. Pattern recognition skills of AI would not have had much impact on a world of chaos. The baseline idea is that the future is determined by the past through a series of causal relationships. This is in line of context with the infamous concept of 'Laplace's demon'. Laplace's Demon is a hypothetical, infinitely intelligent being that knows the precise location and momentum of every particle in the universe. It also has a complete understanding of the laws of physics, thus, being able to predict the exact future. As to say, it is the ultimate world model that has the perfect understanding of the transition probability and the current state [7].

However, AI cannot be deterministic, since it is based on the concept of probability. This may be the crack in the 'tower of Babel' we were so keen to build. Extracting results through numerous data by the pattern from their distribution ultimately means that data that doesn't fit the pattern will be filtered out. In other words, the truth of data points that do not align with the distribution or do not have sufficient examples to train is not likely to be included in the model. This narrows diversity and follows mainstream patterns, which causes a reduction in diversity indicating a reduction in the potential for creativity and uniqueness.

A prominent example would be the 'Youtube' algorithm. Youtube is indisputably one of the most influential video platforms globally. A famous meme is "Youtube Algorithm brought me here", implying that the majority of users are watching videos recommended by an algorithm of youtube. The algorithm prioritizes videos that generate longer watch times and higher engagement rates, such as likes, comments, and shares. This is based on the assumption that videos with higher engagement tend to be more valuable and interesting to users [8, 9]. Although recommendations are tailored to balance the exploration of popular content and the exploitation of new, potential media, people are increasingly consuming similar content. As a result, individuals may find it difficult to maintain their unique personalities as their tastes converge.

In a similar context, bias, and discrimination of data distribution is also possible. If the training data is biased, the resulting world model inevitably attains the bias. This can lead to further discrimination and social inequality. In 2018, MIT Media Lab conducted the 'gender shades project', analyzing the facial recognition algorithms of Microsoft, IBM, and Megvii in terms of classifying gender. They used over 1,270 images of individuals of various gender and skin combinations. As a result, the systems would give a better performance on light-shaded male individuals, and the error rate for dark-colored women was over 35% [10]. As a similar example, in 2016, the AI chatbot of Microsoft, 'Tay'

was stopped within 16 hours of its initial release. This was because Tay left sexually abusive and racist messages, in response to the deliberately offensive and discriminating behaviors of Twitter users [11].

Some would point out that gathering more data points and using a large training set would be able to reduce bias and reach more objective conclusions. However, the hyper-scaled language model GPT-3 couldn't avoid such bias as OpenAI acknowledged that occupations demonstrating higher levels of education such as legislator, banker, or professor were heavily referring to men [3]. Multi-modal image-from-text generation model DALL-E2 was not an exception and would also tend to generate images of women when given the term 'flight attendant'. Increasing data would likely have a large coverage of concepts, but ironically, could also reinforce the bias [12].

4 Polarization of power

Knowledge and power have eternally been intertwined and the privileged class used the gap in knowledge to maintain power [13]. Thousands of years ago, the measure of power was based on geological terms. Spain and Portugal empires that were able to dominate the sea through navigation techniques colonized the world. After the industrial revolution and the incorporation of capitalism, the United States was able to dominate the economy by making the dollar a key currency.

Now, the core of power in this IT empire lies in data and the control of information flow. Data is replacing machines and factories as the source of power. However, what's unique about data is that it can be concentrated in one place, unlike the prior sources of power. All data can be gathered in one place and one multinational company can control the flow of data worldwide. This can be seen as a new era of AI imperialism.

The concept of concentration of power via AI is already present. Already, numerous startups are developing applications integrated with the GPT-4 plugin model commercially. OpenAI raised the price of GPT-4 to $0.12 per token, which is four times more expensive than GPT-3.5. In addition, inequality for non-English speaking countries is eminent as Korean, for instance, is 6 times more expensive when using GPT-4 model than English. Moreover, when the number of users for each of these apps is added up, the authority of GPT-4 becomes tremendous worldwide. Considering that GPT-4 is not free of bias in data distribution, this should be seriously thought through.

In addition, the ability to generate compact sentences with the elite vocabulary of Chat-GPT exceeded or was similar to that of people, raising the question of plagiarism. Essays or even research papers written with Chat-GPT were unrecognizable from that human-written papers. This can result in apathy towards requiring knowledge, as the majority of people tend to follow the easier, faster path. In short term, this may be convenient, but in the long term, students from early ages are deprived of the opportunity to train their writing and reasoning abilities.

5 Privacy and ethics

Consistent with the fact that collecting a massive amount of data is the key to training successful AI models, data privacy is emerging as a serious issue. In 2021, Amazon's AI voice assistant Alexa came under suspicion as a case of invading the privacy of the users. Amazon employees transcribed the user's commands to Alexa to enhance software through feedback. Despite the claim of Amazon that the voices of users are not identifiable as a record, it cannot be overlooked that the user's commands could contain highly sensitive information such as address or personal contact. Such incidents happen because the appearance and development of AI rapidly progressed and did not give space for legal registrations to catch up. There is a need for discussing regulations over the blind spots of ethical issues involved in the matters of AI.

6 Embracing AI

The world is changing faster than the speed that people can adjust to, from both the perspective of society and individuals. If one cannot keep up with learning new skills every ten years, they will not simply become unemployed but will become a socially 'useless class' who are unable to work. The craze for medical school in Korea, which can be inferred from the trend of university entrance exams, is one of the outcomes of the need to find a professional career that can stably apply 10, to 20 years from now.

Everyone knows what artificial intelligence is capable of. What we fear is our capacity ourselves, and worry what kind of Frankenstein made with layers of attention and neural networks will be able to come up from our boiling pot of intelligence. Depending on the person holding the handle, a knife can be either armor for murder or a tool for convenience. The choice is up to us.

References

[1] Bishop, Christopher M., and Nasser M. Nasrabadi. Pattern recognition and machine learning. Vol. 4. No. 4. New York: Springer, 2006.

[2] Hiebert, Theodore. "The Tower of Babel and the Origin of the World's Cultures." Journal of Biblical Literature 126.1 (2007): 29-58.

[3] Brown, Tom, et al. "Language models are few-shot learners." Advances in neural information processing systems 33 (2020): 1877-1901.

[4] Jumper, John, et al. "Highly accurate protein structure prediction with AlphaFold." Nature 596.7873 (2021): 583-589.

[5] Silver, David, et al. "Mastering the game of Go with deep neural networks and tree search." nature 529.7587 (2016): 484-489.

[6] Engel, Jesse, et al. "DDSP: Differentiable digital signal processing." arXiv preprint arXiv:2001.04643 (2020).

[7] Shermer, Michael. "Exorcising Laplace's demon: Chaos and antichaos, history, and metahistory." History and theory (1995): 59-83.

[8] Davidson, James, et al. "The YouTube video recommendation system." Proceedings of the fourth ACM conference on Recommender systems. 2010.

[9] Zhao, Zhe, et al. "Recommending what video to watch next: a multitask ranking system." Proceedings of the 13th ACM Conference on Recommender Systems. 2019.

[10] Buolamwini, Joy Adowaa. Gender shades: intersectional phenotypic and demographic evaluation of face datasets and gender classifiers. Diss. Massachusetts Institute of Technology, 2017.

[11] Neff, Gina. "Talking to bots: Symbiotic agency and the case of Tay." International Journal of Communication (2016).

[12] Ramesh, Aditya, et al. "Hierarchical text-conditional image generation with clip patents." arXiv preprint arXiv:2204.06125 (2022).

[13] Foucault, Michel, and Paul Rabinow. "Space, knowledge, and power." Material Culture (1982): 107-120.