ST.GALLEN SYMPOSIUM

#NEWGENERATIONALCONTRACT

THOUGHT PROVOKING IDEAS OF THE GLOBAL ESSAY COMPETITION 2023

Beyond Just a Moral Imperative: The Legacy of Open-source Research in Artificial Intelligence

Gaurav Kamath is one of the top 25 contributors to this year's Global Essay Competition Award. He studies at McGill University and attended the 52nd St. Gallen Symposium as a Leader of Tomorrow.

Introduction

In the not-too-distant future, we will likely look back at artificial intelligence (AI) as the crowning scientific achievement of our generation. Advances in natural language processing—the field of AI that deals with human language—have culminated in the release and widespread popularity of ChatGPT, along with all the controversy it has left in its wake (Weale, 2023). Advances in image generation technology, combined with that very same progress in natural language processing, have brought us the ability to generate images—striking in both how fantastical and life-like they are—from a few keyboard inputs (OpenAI). And advances in the field of machine learning more broadly have algorithms feeding us recommendations we enjoy far more than we'd like to admit. All of these recent advances in AI, however, have only been made possible by the choice of some to keep their work open to the larger scientific community. Though it would be hyperbolic to call open-source research, code, and data the single best legacy we have inherited from generations before us, they are the foundation of what will inevitably turn out to be the best or worse legacy of our generation.

What an AI System Does and Doesn't Do

To understand the significance of open-source research, code, and datasets toward the current state of AI, we need to understand what it actually involves. Let's take the example of today's conversational AI systems, such as

ChatGPT—they not only seem to have a hold on basic grammar, but also produce unique, coherent dialogue that is specific to their ongoing conversations with users. How do they do that? Importantly, what they do not do is simply execute a fixed list of commands—a 'recipe'—provided to them by a human. While we could conceivably frame grammar as a fixed set of rules written as a program for a computer to interpret—rules like "objects go after verbs", and "the word 'eats' is a verb"—a list of such specific rules for conversational relevancy would be near-infinite. You would have to spell out endless combinations of conversational topics and unique responses, even assuming you programmed a means by which to distinguish between different conversational topics. The trick, instead, is to not even come up with a fixed set of instructions for a specific problem, but rather only a fixed set of instructions on how to learn to solve the problem. The roots of contemporary AI research lay in trying to determine what problems could be solved by following fixed 'recipes' (Wooldridge, 2021); current research instead puts the focus on how artificial systems can, to use the same metaphor, learn those recipes themselves.

Here's how it works: most modern AI systems are no more than a model of some phenomenon, meant to make predictions about that phenomenon. A meteorological model might make predictions about whether it will rain tomorrow; a language model (as most language AI systems are better called) makes predictions about what words fit well in a sentence. For example, a language model predicts whether 'ran' or 'running' is a better continuation to a phrase like 'a dog is...'. If such a model can make good predictions, it can also produce decent-sounding sequences of language, such as 'a dog is running', instead of 'a dog is ran'. Generalize the same idea to a higher level of abstraction, and one reaches a problem like conversational relevancy: if a model can robustly make good predictions about whether a given sentence is appropriate given the previous parts of a conversation, it can also produce natural-sounding responses to an ongoing conversation.

Which begs the question: how does one get such models to make good predictions in the first place? The answer is simple: statistics. Contemporary AI systems are models that use massive amounts of data to induce certain statistical patterns, and then use these statistical patterns to make predictions. A language model, for example, induces the statistical pattern that 'a dog is running' is far more common a phrase than 'a dog is ran', and thereby makes the correct prediction that 'running' is a better fit than 'ran' in that context. The exact means by which such statistical patterns are captured by a model varies based on the type of model in question—state of the art language AI systems are all neural network models, which capture information through millions (and often billions) of numerical parameters—but in all cases, the model induces some representation of statistical patterns from the data it is exposed to. Statistics—and more importantly, big data—allows us to avoid having to hardcode any specific protocol into an AI system, besides the protocol by which it induces statistical patterns.

Open-Source Research and Data: the Life-Blood of Research in AI

All of this makes data, and datasets, indispensable for AI research. Indeed, it is entirely likely that even if modern machine learning methods had been invented a

century ago, we would nevertheless have to wait until today for our modern AI systems. The reason is that these machine learning methods require large amounts of data to induce accurate statistical patterns for complex phenomena. We therefore have the internet to thank as a endless source and repository of such data. Text, photos, videos—in 2023, it is estimated that a total of 120 Zettabytes (that's 120 trillion GB) of data will be produced (Taylor, 2022). And these volumes of data are central to the high performance of current AI systems. The language model underlying ChatGPT, for example, was exposed to 300 billion words during its 'training' (the period in which a model attempts to induce patterns); most of this textual data came from openly accessible text on the internet—Wikipedia, online articles, and uploaded books (Brown et al., 2020). Similarly, most AI systems built for image recognition rely on massive datasets such as ImageNet—an open-access dataset containing over 14 million images for over 20,000 concepts (Deng et al., 2009). Had all this data not been on a common, accessible platform, it is fair to say we would have never made the breakthroughs in AI we see today.

But it isn't just the data that must be open-access—transparency and openness in research and code have been equally important for the recent progress in the field. Language-focused AI presents one of the most striking examples of this. The rapid improvement in the performance of language AI systems in the past few years can largely be traced to the invention of a particular type of model, dubbed the transformer model, in 2017 (Vaswani et al., 2017). The researchers behind this model, however, published all of its details, as well as the code they used to implement it; within just a year, other researchers in the field were able to take inspiration and build off of their initial efforts, developing new types of transformer-based models that improved on the original. One such transformer-inspired model was GPT—a predecessor to the model that underlies ChatGPT (Radford et al., 2018). Such efforts at open-sourcing have been tremendously helped by the rise of platforms like Hugging Face, which allow companies, researchers and students alike to share, download and tweak a whole range of AI models for free (Hugging Face). This sort of open-sourcing greatly boosts the speed of scientific progress: researchers can quickly build on each other's work, without having to reimplement any code from scratch.

Deviations from the Spirit of Scientific Openness

Unfortunately, we currently see deviations from the very approach that brought AI research so much success. OpenAI, the creators of ChatGPT, have (in a cruel sense of irony) not open-sourced the model that underlies ChatGPT; people can use the conversational AI system, but not work with its underlying model to try and build on it. One of the most advanced image recognition systems, a model named CoCa (Yu et al., 2022), is trained on a dataset that is strictly internal to Google. And in a move that could have had serious health outcomes in the long run, Google Health published research claiming an AI system of theirs could efficiently conduct breast cancer screening (outperforming radiologists under certain conditions), without releasing the code nor model used for the study, hindering any attempts at validation (Haibe-Kains et al., 2020; McKinney et al., 2020). What is particularly concerning about such cases is that they are often the state-of-the-art systems in the subfield. By failing to open-

source this research, such companies not only hurt efforts towards verifying the scientific claims they make, but also slow down the progress of the field in general. It is highly unlikely, for example, that we would have an AI system like ChatGPT by now if not for the decision of the original researchers of the transformer model to open-source their work.

Realizing Long-Term Benefits in the Present

So what is the solution? While it is tempting to demand that companies like Google and OpenAI be forced to open-source all their research, this ignores the fact that in such cases, the potential for proprietary research outputs is the very thing that attracts significant research investment. Just as it is true that an AI system of the quality of ChatGPT would not have been possible without the open-sourcing of prior research, it is also true that it probably would not have been built—at least not so soon—without the promise of a proprietary product. What I instead propose, therefore, is that public and civil institutions incentivize open-sourced research, code, and datasets, by helping realise their long term benefits in the present, both in monetary and scientific terms.

On the purely scientific side, I propose that AI research journals and conferences reserve a section for openness in their evaluation of research submissions. While not banning or preventing research contributions that fail to open-source their code or data, additionally evaluating contributions on openness and accessibility would penalize contributions that do poorly on this metric, while rewarding and increasing the visibility of research contributions that are easy to replicate or build on. This could help build scientific prestige around open-source research, and potentially act as a soft incentive to the same end.

On the monetary side, I propose that relevant public institutions provide concrete, immediate financial incentives to individuals and organizations in AI research that commit to open research practices. For students and academics, for example, this would take the form of research grants provided on the condition that they publish their data and code. For AI research organizations, this would take a more diverse range of forms, including start-up loans and tax incentives for organizations that engage in such practices. These could help make up for the financial disincentives research organizations face in open-sourcing their work, and act as a more concrete incentive.

Conclusion

The rapid progress we see in AI today has only been made possible by a legacy of openness in scientific research in the field. Although that spirit persists for the most part, it is important that we ensure it survives, especially in the face of indications that some in the industry are turning away from it. Crucially, open-sourcing research in AI should not just be viewed as a moral imperative—it has also had immense scientific (besides monetary) value for the field, and will continue to do so. It is vital that our public and scientific institutions reflect this.

References

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in neural information processing systems, 33, 1877-1901.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). Ieee.

Haibe-Kains, B., Adam, G. A., Hosny, A., Khodakarami, F., Waldron, L., ... Aerts, H. J. W. L. (2020). Transparency and reproducibility in artificial intelligence. Nature, 586(7829), E14–E16. doi:10.1038/s41586-020-2766-y

Hugging Face. (n.d.). https://huggingface.co/

McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., ... & Shetty, S. (2020). International evaluation of an AI system for breast cancer screening. Nature, 577(7788), 89- 94.

OpenAI. (n.d.). DALL-E 2. https://openai.com/dall-e-2/

Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training.

Taylor, Petroc. (2022, September 8). Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025. Statista. https://www.statista.com/statistics/871513/worldwide-data-created/

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30.

Weale, Sally. (2023, January 13). Lecturers urged to review assessments in UK amid concerns over new AI tool. The Guardian. Lecturers urged to review assessments in UK amid concerns over new AI tool | Artificial intelligence (AI) | The Guardian

Wooldridge, M. (2021). A brief history of artificial intelligence: what it is, where we are, and where we are going. Flatiron Books.

Yu, J., Wang, Z., Vasudevan, V., Yeung, L., Seyedhosseini, M., & Wu, Y. (2022). Coca: Contrastive captioners are image-text foundation models. arXiv preprint arXiv:2205.01917.