Scientist is the new scribe

What the printing press can teach us about the future of AI

Antoni Maciąg
r11922182@ntu.edu.tw

Introduction

With every technology that is expected to be revolutionary comes a great deal of speculation on its potential effects on our civilization, put forward by amateurs and scientists alike. The scientific side of these speculations, which is futurology, is a notoriously unreliable discipline, though. In particular, the most relevant impact of a given technology often proves to be indirect and consist in its promoting other advancements rather than being applied itself. A good historical example of this is the printing press, whose immediate influence as a machine enabling cheap production of books was dwarfed by the influence it achieved indirectly by facilitating and contributing to the Scientific Revolution.

Below, I will draw some paralells between Artificial Intelligence and the printing press and justify why, in the future, the influence of the former may follow a similar pattern, i.e. enhance humanity's research capabilities, and thereby enable countless inventions in other, even unrelated fields and without direct usage of AI. I will discuss two key factors that AI may be expected to improve soon: knowledge extraction and research reliability, and make predictions about the more distant future of AI-driven research.

Knowledge extraction

Access to preexisting knowledge has always been vital to research, and historically, every breakthrough in the methods of storing and sharing information has caused profound scientific and societal changes. The most notable examples of this include writing, the Internet, and the printing press. While we cannot yet appreciate the significance of the Internet, due to its being a recent invention, it is safe to say that science as we know it today would not have been possible without the printing press. Its invention contributed greatly to the Scientific Revolution by introducing unprecedented information flow, reinvigorating the Renaissance and democratizing access to information - indeed it was the printing press that enabled the wide distribution of Copernicus' 'De revolutionibus orbium coelestium', sometimes dubbed the launching work of the Scientific Revolution.

Today, our hopes for a smilar breakthrough might lie in AI-aided comprehension of scientific litarature. Let us consider several tasks from the domain of Natural Language Processing, for example Question Answering (QA), which involves answering a question related to a passage of text. This technology will be immensely helpful to researchers once it is able to automatically answer questions about all existing literature on a given topic. For example, we could ask: 'have any other authors encountered this problem before? Is there any research contradicting my thesis? I need to understand a phenomenon from a scientific discipline I'm not familiar with - how does it work, in simple terms?' Alternatively, we could utilize the Summarization task and ask for a rundown of previous work on a given subject. Just as the impact of the printing press consisted in making book production, a task that could previously be done by humans anyway, significantly faster and cheaper, these technologies only need to outperform humans in speed in order to have an enormous impact.

However, in fact I mentioned QA precisely for the reason that it already scores better than humans in the standard evaluation metric for the task, called SQUAD [1] [2]. While its scoring only evaluates answers to questions related to short paragraphs, this can serve as a proof of concept, showing that an extension of this task to the totality of scientific literature is imaginable.

Another promising area, more specifically tailored to reading papers, is Document Understanding. Recent advancements make it possible to automatically capture even such detailed and error-sensitive data as that related to printed circuit board design [3]. In this case, information was extracted from visual data, and since visualizations are a prevalent tool for presenting scientific findings, we can expect the field of image processing to be as important to automatic knowledge extraction as NLP is.

The examples above suggest that an AI capable of answering every sensible question about all existing scientific literature might in fact be created in the foreseeable future.

Research reliability

Just as the printing press significantly reduced the amount of mistakes in published books, we may draw another loose analogy to AI and surmise that, like in other fields, automating human labor in research will increase the reliability of the results. Not long ago, the academia in the scientific field of psychology was shaken to the core by the replicability crisis [4], indicating that 21-century science still has vast room for improvement in this respect. As it happens, an AI-based tool for estimating replicability is already available [5], although its mechanism is rather simple and relies on human psychology - it detects the use of language indicating the author's own uncertainty about the soundness of their results. More reliable and rigorous tools could be created, based on similar solutions as mentioned in the previous section, but applied in a different way - rather than facilitating good research, making it impossible to publish bad research.

Such compulsory automated peer review could include a central system for finding preexisting papers in support of (or contradiction to) the one being published. If the results of this system were required to be published alongside the article, the author would be forced to address the contradictions, and cherry-picking results to fit a particular thesis would be considerably harder. That said, this part of the review would not impact papers containing completely novel findings, for which no support or contradiction could be found in the literature.

Another promising area lies in the flawed use of statistics, which scientists are notorious for [6]. Problems like linear models which should have never been constructed because the model's assumptions are not met; biased test groups; conveniently ignoring important observations by classifying them as outliers; and many other creative swindles or omissions could be stopped by automatic detection. Whole pseudoscientific movements like anti-vaccionism would have never had a semblance of academic soundness if such a system had been in existence while their launching works were being published.

For similar reasons as above, and also because of the early signs of creating systems for automatic research quality control mentioned above, such systems too might be created by this or the next generation of AI researchers.

AI-driven research

I should probably mention that in this section of the essay I will feel free to make some adventurous predictions, grounded more in my imagination than in reality. The only parallel I can find between them and the printing press is that the latter was, for some time after its invention (invention in Europe, that is [7]) called 'artificial writing'. To us, this may now sound - amusingly - as if printing was considered an inferior alternative to the work of the scribe, which, as we now know, was eventually entirely superseded by the new invention. Maybe in a century or two referring to artificial intelligence as such, as in - something that artificially mimicks

[human] intelligence, will seem equally amusing.

In science, this might manifest itself in AI-driven research (AIDR), rathen than AI-enhanced research (AIER) described in the previous sections. The latter can be expected to first develop and then slowly give way to the former. An intuitive early use of AIER could be to find correlations and links that humans would likely not notice and making preditions for further research. In fact, this has already been applied, and in mathematics, no less - to make conjectures that were later found to be true by human mathematicians [8]. Perhaps an even more astounding result has been obtained by a system capable of capturing latent knowledge from articles related to material design, which, having analyzed a number of publications, made its own preditions for a material of certain desirable properties, guided by knowledge **not directly available in any of those papers and not understandable yet to the researchers themselves** [9].

This system, in the logical next step towards AIDR, could analyze papers, conduct experiments by itself (if given control over appropriate machinery and materials), reason based on those experiments, and finally translate its findings into human language in the form of - if this is not too shocking to imagine - scientific articles.

This last part - translation of AI-obtained knowledge into human-understandable knowledge - is, in fact, worth exploring further. There are still no available methods for this process, and it is not for lack of knowledge that we would like to translate. Beside, for instance, a system for analyzing brain activity in rats, demonstrating an understanding of this activity exceeding human neurological knowledge [10], and the material design system mentioned earlier, a notable example of this is the AlphaFold system [11]. Again, the insights guiding its predictions for protein folding are not understandable to its creators, but they are there. This is surprising even from the philosophical standpoint - it is now unquestionable that a whole new form of knowledge has been created, encoded not in the structure and composition of a biological brain, as it was up until now, but in model parameters like the weights and biases of a neural network.

This has also given rise to a peculiar phenomenon in which we do not know the mechanisms behind protein folding, but have a tool at our disposal allowing us to predict the outcome of the process for any particular protein type (or almost - AlphaFold, to be strict, is not perfect). In this case, will we still insist on understanding these mechanisms? I suppose we will, and one can already predict two methods of coping with the problem of translating computer-obtained knowledge to human-understandable knowledge.

One if them is drastically increasing the emphasis on transparency in AI. We know that some model classes (e.g. decision trees) tend to be more understandable in their mechanisms than others (e.g. neural networks [12]), but for now this is mostly an added benefit rathen than deciding factor for choosing a model. We might be forced to develop models that are both powerful and transparent.

Alternatively - that is the other predictable solution - we might perfect our model distillation methods, allowing us to encode knowledge present in powerful but obfuscated models like AlphaFold into simpler, but more transparent models. Whichever path we take, it is not unimaginable that translations of AIDR-generated results into human-understandalbe form should become one of humanity's primary sources of knowledge.

Conclusion

The most promising area for future use of AI lies in enhancing human research capabilities, as systems developed for this purpose would have the potential to be applied simultaneously in all fields of science and technology and revolutionize all of them at once. This pattern of an invention making its biggest impact in an indirect and unintended way is not without parallel in history, and the example of the printing press can inspire us to make some predictions about the future of AI. This work explores such preditions, introduces the idea of AI-enhanced research, whose first signs are already observable today, and based on these signs, speculates about the idea being taken further to AI-driven research.

References

[1] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, Percy Liang. SQuAD: 100,000+ Questions for Machine Comprehension of Text, 2016

[2] Zhuosheng Zhang, Junjie Yang. Retrospective Reader for Machine Reading Comprehension, Hai Zhao, 2020

[3] Kuan-Chun Chen, Chou-Chen Lee, Mark Po-Hung Lin, Yan-Jhih Wang, Yi-Ting Chen. Massive Figure Extraction and Classification in Electronic Component Datasheets for Accelerating PCB Design Preparation, 2021

[4] Open Science Collaboration. Estimating the reproducibility of psychological science, 2015

[5] Yang Yang, Wu Youyou, and Brian Uzzi. Estimating the deep replicability of scientific findings using human and artificial intelligence, 2020

[6] John Gardenier David Resnik. The Misuse of Statistics: Concepts, Tools, and a Research Agenda, Accountability in Research: Policies and Quality Assurance, 9:2, 65-74, DOI: 10.1080/08989620212968, 2002

[7] Seung-Hwan Mun. Printing press without copyright: a historical analysis of printing and publishing in Song China, 2013

[8] Alex Davies, Petar Veličković, Lars Buesing, Sam Blackwell, Daniel Zheng, Nenad Tomašev, Richard Tanburn, Peter Battaglia, Charles Blundell, András Juhász, Marc Lackenby, Geordie Williamson, Demis Hassabis Pushmeet Kohli. Advancing mathematics by guiding human intuition with AI, 2021

[9] Vahe Tshitoyan, John Dagdelen, Leigh Weston, Alexander Dunn, Ziqin Rong, Olga Kononova, Kristin A. Persson, Gerbrand Ceder Anubhav Jain. Unsupervised word embeddings capture latent knowledge from materials science literature, 2021

[10] Markus Frey, Sander Tanni, Catherine Perrodin, Alice O'Leary, Matthias Nau, Jack Kelly, Andrea Banino, Daniel Bendor, Julie Lefort, Christian F Doeller, Caswell Barry. Interpreting wide-band neural activity using convolutional neural networks, 2021

[11] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli Demis Hassabis. Highly accurate protein structure prediction with AlphaFold, 2021

[12] Rahul Iyer, Yuezhang Li, Huao Li, Michael Lewis, Ramitha Sundar, Katia Sycara. Transparency and Explanation in Deep Reinforcement Learning Neural Networks, 2018