Deep Learning Considered Harmful?

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Michael Andrzej Klaczynski

Spring 2020

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Travis Elliott, Department of Engineering and Society

Over the past decade, the tool known as deep learning has risen to prominence, being hailed by tech media outlets as a driver for innovation. In and outside the world of tech, deep learning is used by researchers, businesses and governments to analyze data, which they then use to make informed decisions.

Deep learning is unquestionably useful for a variety of tasks. But there are drawbacks to deep learning that often go ignored, leading to misuse. This paper examines the social factors behind these cases of misuse through the lens of the sociology of scientific knowledge (SSK), in particular the strong programme, a school of thought that attempts to explain the success and failure of scientific theories through the application of sociology. Deep learning is not just a technology, but a tool for science, and it is in this respect that it is so often misused.

In my research, I have found two types of social factors that lead to this: there are external factors, where outside forces mislead people on the use of deep learning. Then there are internal factors, where people use deep learning out of a lack of better options.

A History of Disappointments

Since the dawn of computers, people have sought to use them to mimic human thought. The first automated theorem provers were created in the mid-1950's. The "Logic Theorist" program, in 1956, was capable of proving most basic mathematical theorems. Because of the ease at which basic logic could be simulated, it was thought that true simulation human reasoning would only be a few decades away.

It was not.

Several movements rose and fell over the later decades of the 20th century, each with its own bubble of enthusiasm for new technologies. Symbol-systems, connectionism, expert systems. As they failed to live up to the massive expectations set for themselves, they all fell into

what is termed “AI Winter” when enthusiasm falters and funds dry up. While many great strides were made in the advancement of AI, they were always less than expected.

In particular, two problems plagued every attempt at AI: *common sense* and *combinatorial explosion*. It’s hard to feed a computer program every basic fact about the world that humans seem to understand so easily. And, even when dealing with simplified systems of facts, the search for solutions to problems increased exponentially with the number of variables. Essentially, the new high-tech AI programs couldn’t work outside of toy examples (McCarthy, 1974).

The Dawn of Deep Learning

Though the idea of artificial neural networks has been around since Turing’s day, they didn’t see much successes in the 20th century due to the amount of computing power required to run them, as well as the amount of data required to train them. Deep learning, which describes not only neural networks but the processes used to train them, didn’t see a boom in usage until the early 2010’s, when significant advances were made in image recognition.

“Deep learning” started trending on Google in 2013, reaching a peak in 2017. During this time, technologies that used to seem like gimmicks rapidly improved: speech-to-text, facial recognition, and handwriting software all became usable features in the average smartphone.

What Deep Learning Can Do (essentially)

A neural net is a large complex function whose behavior controlled by a large number of parameters. Deep learning asks the net a question, then adjusts the parameters based on the correctness of the answer. It does this a large number of times. Eventually, this results in a net capable of answering correctly the type of questions it has already been asked.

To paraphrase deeplearning.ai founder Andrew Ng, deep learning can automate anything that a human can do with less than a second of thought. It solves the problem of combinatorial explosion by identifying the relevant features of a large set of data, and working with those. It also partially solves the common sense problem, as it learns through experience, rather than manually assigned knowledge.

As wonderful as that sounds, there are some limitations.

What Deep Learning Can't Do

First, it's important to establish that deep learning can't do the impossible. This may seem obvious, but sometimes it *appears* to do so. For instance, a Super-Resolution GAN network is designed to apparently increase the resolution of a down-sampled image. Truly recovering the lost pixels, however, is in reality an impossible task. That information has been destroyed, and there is no guarantee that the remaining pixels can be used to determine what was originally there. The net can only make guesses based on its previous experience.

Furthermore, deep learning can not tell why a neural net answers the way it does. While an experienced data scientist might be able to guess based on which questions it gets wrong, the underlying functions are often too complex to manually examine. While there is much research currently focused on "interpreable AI," trying to explain the inner workings of neural networks, there still isn't a good solution to this problem. Neural nets are "black boxes." Questions go in, answers come out, but you don't know what's going on inside. This can cause a number of problems.

The first problem, which we see a lot with deep learning applications, is fairness. Whenever a system is put in place to make decisions that impact peoples' lives, inevitably someone will feel like they are treated unfairly. If the decisions are made by a neural net, there is

no way of proving if they were or weren't. This problem has come up in a number of recent lawsuits. Machine learning algorithms that determine prison sentence based on criminal history have been called racist, and algorithms which determine someone's loan status have been called sexist. In both cases, race and sex were not features included in the data sets. With increasing attention and regulation, AI lawsuits will likely become more common in the near future (Lewis, 2020)

The second problem is more abstract. Deep learning can often seem like a shortcut to solving problems. It is. It's a great shortcut. But in some cases, the joy is in the journey. When a problem is solved by deep learning , no new knowledge is learned by humans, other than "a neural network of this size and shape can solve this problem." For engineering purposes this is fine, but it's somewhat scientifically troubling. If no new knowledge is discovered, what scientific progress can be made?

Will these problems prove to be the death of deep learning in a few years? Is it just another over-hyped toy that is doomed to fail? Are we headed for another AI winter?

I think not.

Why Deep Learning is Different

Deep learning works, very effectively. Unlike the many AI technologies of the past, deep learning can actually be used to the direct benefit of businesses, governments and individuals. Despite the limitations, it is very unlikely to fall out of use anytime in the near future. It is simply too useful.

Misuse Cases

So long as deep learning exists and is popular, there will be misuse. This paper focuses on three ways in which deep learning can be misused:

1. The task is impossible with the data given.
2. The task is possible, but the user is incompetent.
3. The task is possible and the user is competent, but explainability is important.

In the first case, perhaps the user tries to use deep learning to predict someone's lifespan by looking at their palms. Some basic correlation might be found, resulting in a faint heuristic, but overall the data is clearly insufficient.

In the second case, the user has taken on a task beyond their capability. This can happen quite often for reasons we will discuss later. In this case the resulting model may be an oversimplification of the problem, getting most questions right but still missing an unfortunate number.

In the third case, the user has decided to sacrifice understanding for expediency, creating an effective solution, but ignoring how they got there. Usually, this does not count as misuse. This can be a perfectly valid decision. But if explainability is important, deep learning is not the answer.

So what factors lead to these cases? They are all technologically disadvantageous, so the answer must be social.

The Circle of Hype

It is a well-documented phenomenon that media hype can occur in a way that is disproportional to real events. For example, a single report of a violent crime can trigger a chain reaction, drawing attention towards the same sort of story, resulting in a self-sustaining wave of news stories over a "crime wave" that doesn't actually exist (Vasterman, 2005). News media does not simply report news, they decide what is news, which is influenced in turn by the news.

The same sort of wave could happen over innovation. One story about an AI beating a champion at Go could focus more media attention on AI, as tech journalists seek out stories related to the ongoing narrative about advances in AI.

To businesses, this smells of free publicity. Give the journalists what they want (Austin, 2004). If journalists are looking for stories about AI, somehow make your business have to do with AI (even if it's not). Tech companies always want to look like they're on the forefront of technology, and non-tech companies also want to seem high-tech. AI technologies like deep learning fortunately happen to be easily applicable to a lot of businesses, so it's not usually a complete lie. Usually. There have been some cases of businesses using "pseudo AI," where companies pass on their "AI" tasks to crowd-sourcing organizations such as Mechanical Turk, substituting real humans for AI that doesn't exist (Solon, 2018).

Explosions in hype therefore don't need a real catalyst. Researchers and companies can exaggerate their findings, the role of AI in those findings, or the novelty of their methods, in order to gain more attention and funding for the projects. The news media begins a hype wave, which then inspires more businesses to use AI technology, creating a self-sustaining cycle of hype.

And then there is the even greater hype that is generated by the entertainment media, which draws inspiration from the news media. The field of AI has always been haunted by the high expectations set by science fiction. On one hand, it draws more people into the field; on the other, it creates many misconceptions about current capabilities. Indeed, because science fiction is by definition detached from reality, it can make whatever promises it wants about the future of innovation. While these claims may be less influential than the more reality-based ones, they still have a reasonable effect on enthusiasm.

Because of science fiction, AI is “cool”. Engineers and businesses want to work with AI, and researchers want to meet the expectations of science fiction. IBM may have missed the deadline of 2001 to make the HAL 9000, but by 2011 they made a computer capable enough at natural language processing to win at *Jeopardy!*. Art influences life. Then, of course, the news media reports on that, which in turn inspires more news and art.

This can account for all three misuses of deep learning. Businesses want to advertise their AI skills, whether or not they have them. Engineers want to use deep learning, to get a raise or to get prestige, but lacking formal training in data science, they often get it wrong. Even researchers may need something flashy to propose to get grant money. Therefore, much spectacle is generated with little substance. And the media is there to eat it all up.

The Pressure for Progress

But hype is not the only reason one might misuse deep learning. There are also more specific social pressures that lead to misuse. Deep learning is genuinely a useful tool, and people want to see if they can use it. But, as the saying goes, when all you have is a hammer, everything looks like a nail.

Take for example the response of the data science community to the 2020 COVID-19 pandemic. Every other news article about AI was some group claiming to be using deep learning to fight the outbreak, from detecting the virus based on x-rays, to detecting fevers with infrared cameras, to predicting the spread. The front page of kaggle.com, a site dedicated to machine learning challenges, had links to several epidemic-related data sets. There was an air of optimism and purpose that was ever being promoted by a thirsty news media.

The hype mixed with the desperation to help caused some to ignore the significant lack of data actually available. The x-ray virus detectors praised by several news outlets turned out to

have far too little training data to actually function, let alone at the 96% accuracy rates they were claiming (Engler, 2020). Actual testing was too sparse to do much reliable epidemiology. The most successful example of predicting local outbreaks did so by scanning local news outlets for increases in concern. One might see why this was not particularly useful, as if the local news is covering it, the outbreak has already occurred (Heaven, 2020).

Because deep learning is so inexpensive to use, people desperate to solve a problem quickly, or people who want to appear to be solving problems, turn to it when other options run out. This very often leads to misuse. Bad science is worse than no science, as it draws attention away from legitimate efforts.

Even in less extreme scenarios, pressure for results can lead to bad practices. If a scientist is given a choice between using a black-box deep learning model, which will produce solid results relatively quickly, but provide little insight into the problem, or years of unimpressive incremental progress that will increase understanding, what will they choose? Especially if getting funding is an issue? At the moment, there are few social disincentives to using deep learning. In most cases, deep learning seems to improve life for the user and everyone around them, for the immediate future. Problems which may arise from not understanding deep learning models may only occur further down the road.

Going Forward

One wonders if, in the future, more disincentives will develop. The discrimination lawsuits that seem to pop up every other month may in the future dissuade companies from jumping into the AI business so quickly. As for scientific research, perhaps scientists will come to frown on lazy deep learning practices if they start to see progress stagnate due to a lack of new

knowledge being generated. Perhaps a code of best practices will be developed. Or, maybe progress will be made interpretable AI, alleviating this problem partially.

The hype cycle will likely continue to proliferate its self, at least until until the next big technology comes along. Buzzwords tend to accumulate stigma over time as their overuse eventually makes them meaningless (Capoor, 2017), and perhaps “deep learning” and “AI” will be replaced by similar phrases in due time.

Deep learning itself is not inherently harmful, but it *can* be if we allow social pressures to rule over when we use it. Deep learning and AI are exciting, but should always be taken with a healthy dose of skepticism. Hype is the enemy of realistic expectations. It takes attention away from legitimate and useful endeavors and puts focus on the projects least likely to work. The constant desire for results often overrides the desire for knowledge. The scientists and engineers wielding these tools must consider carefully both whether they can, and whether they should, in each situation.

Bibliography

Austin, P., & Austin, B. (2004). Getting free publicity: The secrets of successful press relations. How To Books Ltd.

Capoor, B. (2017). A General Theory of" Buzzwords": Synergistic Meta-Linguistic Paradigm Shifts. Inquiries Journal, 9(02).

Engler, A. (2020, April 26). Artificial Intelligence Won't Save Us From Coronavirus.

Heaven, W. D. (2020, April 10). AI could help with the next pandemic-but not with this one.

Lewis, N. (2020, February 28). AI-Related Lawsuits Are Coming.

McCarthy, J. (1974). Artificial intelligence: a paper symposium: Professor Sir James Lighthill, FRS. Artificial Intelligence: A General Survey. In: Science Research Council, 1973.

Solon, O. (2018, July 6). The rise of 'pseudo-AI': how tech firms quietly use humans to do bots' work. Retrieved from https://www.theguardian.com/technology/2018/jul/06/artificial-intelligence-ai-humans-bots-tech-companies

Vasterman, P. (2005). Media-Hype Self-Reinforcing News Waves, Journalistic Standards and the Construction of Social Problems.

Vasterman, P. (2018). From Media Hype to Twitter Storm. Amsterdam University Press.