Implementing Job Scheduling for Distributed Training

(Technical Paper)

The Importance of Interpretability, Accountability, and Ethics in Deep Learning

(STS Paper)

A Research Paper Submitted to the

Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements of the Degree
Bachelor of Science, School of Engineering

Brian Yu
Spring 2021

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

Signature ______________________ Date 12 May 2021

Brian Yu

Approved ______________________ Date 12 May 2021

Professor Richard Jacques, Department of Engineering and Society

I. Introduction

The field of deep learning is one of the most active research areas in modern computing. Deep neural networks are complex machine learning algorithms trained on very large datasets. These models are often trained on millions of data points, with the hope that they will be able to generalize to data points they have never seen before. The efficacy of deep learning models increases with the amount of data that the models are trained on and the complexity of the model. With larger datasets, deep learning models become more robust to outliers and overfitting, allowing them to generalize better to data that they have never seen before. More complex models with a higher number of parameters can learn more complex probability distributions of data and thus also generalize better. The tradeoff is that increasing model complexity and dataset size can drastically increase the amount of time and computing resources it takes to train deep learning models. Modern, industry-standard deep learning models can take weeks and even months of non-stop computation on clusters of hundreds of computers to fully train.

The field of distributed systems is another active research area in computing. Distributed systems allow computer systems to scale beyond the computing capacity of a single computer by linking many computers together over a network. Distributed systems are what make modern large-scale computer systems possible and allow companies like Google, Amazon, and Facebook to handle millions of concurrent users. Given the high computational demands of deep learning models and the prevalence of distributed systems, one common approach to increasing the training speed of deep learning models is to leverage distributed systems in a process known as distributed training. Oftentimes, one computer cluster is used to train multiple deep learning models concurrently. Given a cluster of computers and an arbitrary number of models, it is difficult to place all of these models on the cluster so that they fully utilize cluster resources and finish training quickly. This is the main focus of my technical project. My project involves implementing state-of-the-art deep learning scheduling algorithms on both simulated and real computer clusters. These implementations provide a baseline that researchers in my lab can use to compare their own scheduling algorithms.

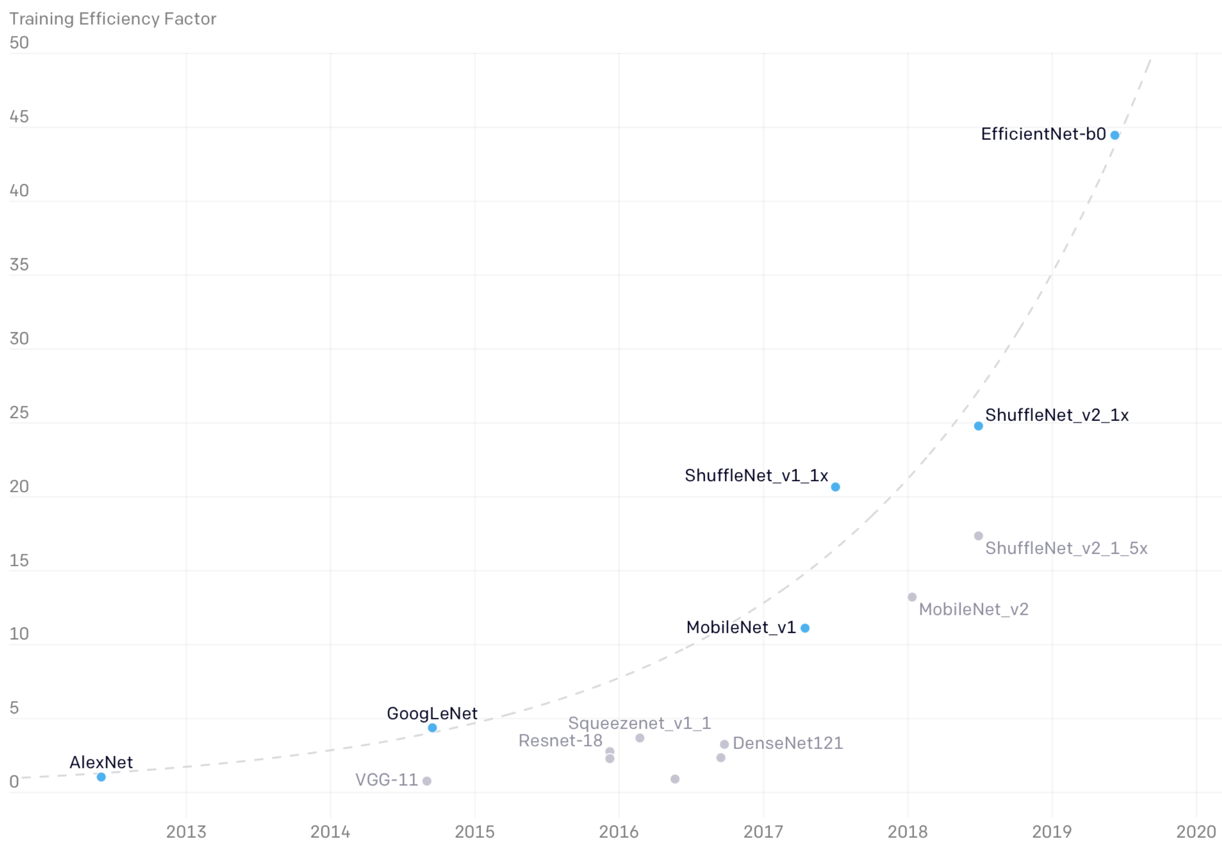
Deep learning models are being used in novel and powerful ways. These models power the latest generation of autonomous vehicles, drive content recommendation algorithms that control what we watch and read online, and are used to analyze massive amounts of data in order to detect fraud and crime (Liu et al., 2017). However, the mechanisms by which deep learning models work are often considered to be black boxes that are difficult for humans to understand (Guidotti et al., 2018). Given the use of deep learning models on increasingly important problems and their increasing pervasiveness in our lives, we need deep learning models that are interpretable, accountable, and ethical. In this research paper, I will elaborate on the increasing power of deep learning models and draw from several STS frameworks to argue for the importance of interpretability, accountability, and ethics in the field of deep learning.

II. The Increasing Influence of Deep Learning

Deep learning models have increasing power in our society as we become more reliant on them. They already heavily influence the content that we see and the products that we buy online, via search and recommendation systems used by Amazon, Google, YouTube, Twitter, and Facebook. These data-driven recommendation systems aim to maximize user engagement based on prior data but may create a “pernicious feedback loop” that “homogenizes user behavior without increasing utility” (Chaney et al., 2018). These algorithms are designed to maximize user engagement over everything else. Oftentimes, this means surfacing content that either agrees with a user's predisposed biases or is also liked by many other users, even at the cost of correctness. This leads to filter-bubbles and echo-chambers that sometimes allow misleading information to thrive and spread (DiFranzo & Gloria-Garcia, 2017).

Figure 1

Training Efficiency Improvements of Deep Learning Models Over Time



Note. Figure reproduced from Hernandez and Brown (2020).

We are also finding uses for deep learning in the physical world outside of the internet. Deep learning systems are transforming a wide array of fields. In medicine, deep learning allows us to rapidly discover new drugs (Ramsundar et al., 2015). In science, deep learning is accelerating new scientific discoveries. Recently, researchers have created AlphaFold, a deep learning system that has solved the problem of protein folding, a longstanding challenge in the field of biology (Senior et al., 2020). In transportation and robotics, deep learning is the main driver behind advances in computer vision and planning, allowing for advances in self-driving cars and other autonomous robots (Sünderhauf, et al., 2018). Furthermore, the efficiency of deep learning model training is improving at an exponential rate, doubling every 16 months (Figure 1). This rate of improvement outpaces Moore's law and will allow for even larger deep learning models that can solve problems that were once thought computationally intractable (Hernandez & Brown, 2020). As we apply deep learning to an expanding set of physical use-cases, we will also experience an increase in physical accidents attributed to these autonomous systems. Notable examples of such accidents include the 2018 killing of a pedestrian by an Uber autonomous vehicle (Wakabayashi, 2018) and numerous fatal accidents attributed to Tesla's self-driving functionality (Krishner, 2020).

Despite open problems in AI safety, deep learning models are constantly improving and are only becoming more pervasive in our daily lives. These models can automate processes that once necessitated expensive human labor, leading to cost-saving, increased productivity, and capital accumulation for the creators and implementers of deep learning models (Acemoglu & Restrepo, 2018). Deep learning models are most effective when trained on large datasets using complex model architectures. Furthermore, there is a well-documented reproducibility crisis in the field of deep learning research (Barber, 2019). Thus, deep learning is uniquely suited for institutions that can collect massive datasets, afford large computer clusters, and hire researchers from top universities and companies who know the trade secrets of the field. The deep learning models that these institutions deploy will have outsized impacts on society. Leading deep learning companies realize this, which is why Google, one such company, has gone so far as to create an Ethical AI team. However, Google has recently drawn criticism for its firing of two prominent AI ethicists (Metz, 2021).

III. Examining Deep Learning Through the Lens of User Studies

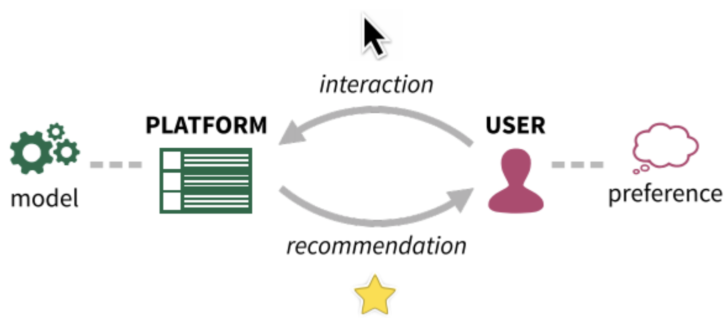
The field of user studies is concerned with the interactions between designers, technology, and users. One important concept is that of *scripts*, which are assumptions that designers make about users and the way that they use technology (Akrich, 1992). These assumptions create a framework of action that is prescribed for users. Sometimes these assumptions are incorrect and the technology is ill-designed for users. In other cases, technology can mold users and society in powerful ways.

In 1990, Steve Woolgar chronicled the production of a new computer system while acting as a project manager (Woolgar, 1990). Woolgar wrote that the “machine text is organized in such a way that ‘its purpose’ is available as a reading to the user.” In other words, the machine is designed so that users can figure out how to use it as the designers intended. In the process of creating a computer, the designers define their target users and establish suggestions and bounds of the users’ actions. This is known as configuring the user.

The main challenge that faced the designers in both Woolgar’s and Akrich’s writing is that of knowing one’s users. Designers go to great lengths to learn more about users lest their products fail to be adopted. It is interesting to study machine learning-driven recommendation algorithms under this framework because data is constantly being collected about users and used to improve the recommendations that they give. In turn, users are configured by those recommendations and their new behavior is fed back into the algorithm, forming a feedback loop (Figure 2). Chaney et al. (2018) discovered that recommendation systems that did not consider this feedback loop offered recommendations that increased the homogeneity of user behavior and decreased the utility of recommendations. Furthermore, the feedback loop amplified the impact of the recommendation system on the distribution of item consumption. This means that in online marketplaces such as Amazon, the recommendation system has a powerful sway over which merchants' items get recommended and thus purchased more often by users.

Figure 2

Recommendation-Interaction Feedback Loop



Note. Figure reproduced from Chaney et al. (2018)

In 2016, the European Parliament passed the General Data Protection Regulation (GDPR), a comprehensive set of rules that regulates the collection, storage, and use of personal information. One regulation included in the GDPR restricts algorithms that make predictions based on user-level predictors, creating a "right to explanation" (Goodman & Flaxman, 2017). In other words, recommendation systems and other algorithms that make decisions based on user data must be interpretable to humans -- a big challenge for standard deep learning algorithms. Furthermore, these algorithmic decisions must be contestable. That is, users should be able to contest these decisions and the algorithm should be able to modify future decisions accordingly. While these new restrictions are problematic for current algorithms, Goodman and Flaxman argue that these problems are beneficial for deep learning because they will lead to models that are more transparent and fair to users.

We use services like Google, Facebook, Amazon, and YouTube daily. Thus, as users of those services, we configure the algorithms running behind the scenes with our user behavior and are subsequently configured by the algorithms when they use that data to produce recommendations. As the European GDPR recognizes, we have both the right to know how deep learning models are using our data to make decisions about us and the right to hold these models accountable and contest their decisions. In order for these rights to be upheld, we need interpretable and ethical deep learning models that can be held accountable for decisions that may harm us.

IV. The Politics of Deep Learning

In 1980, Langdon Winner wrote that technological artifacts can be political, even if they were not designed to be. Winner later analyzed the case of a mechanical tomato harvester designed at Berkeley in the 1940s (Winner, 1986). While it produced a better yield than manual picking, it caused thousands of workers to lose their jobs. This outcome was not intended by the researchers, but reinforced the power of large growers and diminished that of the workers. Other technologies are inherently political. Some require a particular social structure in order to work and some are strongly aligned with a particular social system.

Given the potential of deep learning algorithms to automate work done by humans, it is obvious that these algorithms can be political, similar to the tomato harvester developed at Berkeley. Content recommendation algorithms may also exhibit political biases, even if unintended by their creators. Moreover, algorithms are being used to make increasingly important decisions about us, like creditworthiness, the probability of failure at school, and the likelihood of criminal recidivism (Corbett-Davies et al., 2017). These decisions have broad societal and political implications and MacCarthy (2019) argues that “algorithms are intrinsically ethical in character” and that “it is often impossible to choose between competing algorithms without making ethical judgments". For instance, when creating a model that predicts criminal recidivism, should we choose an algorithm that prioritizes equal accuracy for different races (and thus more frequently predicts African Americans to re-offend than White Americans), or one that minimizes racial bias (Larson & Angwin, 2016)? It is important to be able to understand the biases and inner workings of deep learning models that make decisions of political or ethical importance, which is yet another reason for the value of interpretability and accountability in deep learning models.

Deep learning models also have more obvious political consequences than the ethical character of their algorithms. Recommender systems that utilize them contribute to a phenomenon known as filter bubbles. Since the goal of recommender systems is to maximize user engagement, users are more likely to engage with content and other users that share their beliefs, thus reinforcing those beliefs (El-Bermawy, 2016). Furthermore, recommender systems may sometimes recommend misinformation that aligns with users’ beliefs, causing fake news to go viral. Silverman (2016) found that in the run-up to the 2016 presidential election, the most viral fake news stories generated more engagement than mainstream news on Facebook. Combine this with the fact that Facebook is the top source for political news among Millennials (Mitchell et al., 2015), and it is apparent that deep learning models can wield influence over our political systems.

Despite the potential negative political consequences of deep learning, researchers are also investigating ways in which it can help strengthen our political system and media. Popat et al. (2018) proposed an end-to-end deep learning model that can perform an evidence-aware credibility assessment of arbitrary textual claims without any human intervention. This system can detect fake news with an accuracy of 67% to 83%. While far from perfect, this shows great potential for future systems that can detect and alert users about fake news at scale.

Institutions that create and deploy deep learning models at scale must recognize the ethical and political implications these models might entail and work to minimize any harmful consequences. Leading AI companies such as Google, Microsoft, and Facebook have created internal AI ethics review boards (Vincent, 2019), but Whittaker et al. (2018) argue that "we have not seen strong oversight and accountability to backstop these ethical commitments". As mentioned previously, the GDPR sets an example of how to strongly enforce accountability and transparency in deep learning algorithms. As the number of important applications for deep learning grows, so too must regulation that enforces accountability and ethics in the way that they are used.

V. Deep Learning and Subpolitics

The body of knowledge of science and technology is constantly growing and is showing no signs of slowing down or becoming any less important to society. As such, experts have increasing power to determine technological standards and regulations (De Vries, 2007). This may be problematic because these experts often are not elected public officials and may make these decisions behind closed doors. Thus, they set standards that deeply affect the public without involving the public in the decision-making process.

This rings true in the field of deep learning. In the absence of government regulation, technology companies are setting de facto industry standards in the use of deep learning algorithms. These companies may be over-optimistic about the abilities of their technology, leading to outcomes that would have been prevented by more stringent regulations, such as the fatal crash of a self-driving Tesla Model S in 2016 (Stilgoe, 2017). As discussed previously, many leading institutions have created private AI ethics boards in an effort to combat this problem. Most notably, Amazon, Apple, DeepMind, Google, Facebook, IBM, and Microsoft have collaborated to form The Partnership on AI (The Partnership on AI, 2021), a cooperative nonprofit focused on ethical AI. However, it is unclear what impact these private ethics boards will actually have.

Greene et al. (2019) writes that the vision statements of these private AI ethics boards frame the problem of AI ethics as "a project of expert oversight, wherein primarily technical, and secondarily legal, experts come together to articulate concerns and implement primarily technical, and secondarily legal solutions". Moreover, The Partnership of AI defines the public as a body to be educated and surveyed while the experts (scientists, engineers, and businesspersons) are the ones who will educate and survey. Greene et al. write that these ethics boards give the impression that, "Experts make AI happen, Other Stakeholders have AI happen to them". Since these ethics boards are sponsored by industry, their ethical discussion is often limited to topics of ethical design and implementation, and not whether certain AI systems should even be built in the first place. These boards also assume that only poor ethics and bad design lead to harmful outcomes, ignoring other equally possible outcomes, such as "large corporations dominating political processes with no democratic accountability" (Greene et al., 2019).

How can we make sure that the public is involved in deep learning ethics and ensure that the governance of AI ethics is democratic? According to de Laat (2018), we need to create a transparent oversight body that regulates the use of deep learning in critical applications, a kind of “FDA for algorithms”. While total transparency would be ideal from the public’s perspective, it would cause a loss of privacy when datasets become public and cause companies to lose any competitive edge they may gain from their deep learning models. Thus, some sort of algorithmic safety agency would ensure that companies are still held accountable for their deep learning models by the public, even if indirectly. de Laat also argues that deep learning models should become more understandable in order to aid in this oversight. This sentiment echoes the one expressed in the European General Data Protection Regulation. The public deserves the right to an explanation of how the deep learning models that impact their lives function and the right to contest the decisions made by them.

VI. Conclusion

Deep learning models are already ubiquitous and will only become more pervasive in the future. They play important roles in many parts of our digital lives, influencing the content we see, the items we purchase, and the people we interact with online. In finance, they help determine our creditworthiness and catch fraudsters. These deep learning models are also migrating to the physical world, transforming the fields of biology, science, and transportation. As the capabilities of these algorithms grow, so too will the impacts that they have on society. If we are not careful about monitoring and regulating the power of artificial intelligence, we could face grave and harmful consequences. In this paper, we have used Akrich and Woolgar’s user studies, Winner’s theory of the politics of artifacts, and De Vries’ conception of subpolitics to examine the ethical and political implications of deep learning and explore ways to ensure that deep learning develops ethically and responsibly.

With user studies, we learned about how our online behavior is being compiled into datasets on which deep learning algorithms that impact our behavior, such as recommender systems are trained, creating a feedback loop. Thus, we, as users, can configure deep learning models, and deep learning models in return configure us. If deep learning experts are not careful when creating recommendation systems, this feedback loop can cause increased homogeneity among users and decrease the utility that users get from online services (Chaney et al., 2018).

In studying the politics of deep learning, we saw how deep learning models are being used to make decisions with weighty ethical and political implications. The consequences of mistakes made by these models are grave. Models like these can mistakenly predict that an inmate will re-offend, denying the inmate a chance at parole or choose to recommend misinformation to millions of people, potentially altering the course of an election (MacCarthy, 2019).

Using the sociotechnical framework of subpolitics, we investigated how companies are trying to self-govern their use of deep learning by creating private AI ethics boards. These boards are often opaque to the public and have incentives that align more closely with the companies that sponsor them than the public for which they claim to benefit (Greene et al., 2019). The high-profile value statements from these private AI ethics boards are shaping the moral background upon which ethics conversations are taking place and are imposing the ethical frame espoused by the big AI technology companies that sponsor them.

After studying deep learning through the lens of these three socio-technical frameworks, there are several clear paths forward. First, we must draft and pass regulations that give members of the public the right of explanation of how deep learning models make decisions about them and the right to contest these decisions. Second, we must create oversight bodies that regulate the use of deep learning algorithms in commercial settings. Third, we must research ways in which we can make deep learning models more interpretable, allowing them to be understood and regulated more effectively. By pursuing these pathways, we can strive towards ethical deep learning models that benefit society and minimize their potential for harm.

References

Acemoglu, D., & Restrepo, P. (2018). *Artificial intelligence, automation and work* (No. w24196; p. w24196). National Bureau of Economic Research. https://doi.org/10.3386/w24196

Akrich, M. (1992). The de-scription of technical objects. In *Shaping technology/building society: Studies in sociotechnical change*. MIT Press.

Barber, G. (2019). Artificial intelligence confronts a "reproducibility" crisis. *Wired*. https://www.wired.com/story/artificial-intelligence-confronts-reproducibility-crisis/

Chaney, A. J. B., Stewart, B. M., & Engelhardt, B. E. (2018). How algorithmic confounding in recommendation systems increases homogeneity and decreases utility. *Proceedings of the 12th ACM Conference on Recommender Systems*, 224–232. https://doi.org/10.1145/3240323.3240370

Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., & Huq, A. (2017). Algorithmic decision making and the cost of fairness. *ArXiv:1701.08230 [Cs, Stat]*. https://doi.org/10.1145/3097983.309809

de Laat, P. B. (2018). Algorithmic decision-making based on machine learning from big data: Can transparency restore accountability? *Philosophy & Technology*, *31*(4), 525–541. https://doi.org/10.1007/s13347-017-0293-z

de Vries, G. (2007). What is political in sub-politics? : How Aristotle might help STS. *Social Studies of Science*, *37*(5), 781–809. https://doi.org/10.1177/0306312706070749

DiFranzo, D., & Gloria-Garcia, K. (2017). Filter bubbles and fake news. *XRDS: Crossroads, The ACM Magazine for Students*, *23*(3), 32–35. https://doi.org/10.1145/3055153

El-Bermawy, M. (2016, November 18). Your filter bubble is destroying democracy. *Wired*. https://www.wired.com/2016/11/filter-bubble-destroying-democracy/

Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation." *AI Magazine*, *38*(3), 50–57. https://doi.org/10.1609/aimag.v38i3.2741

Greene, D., Hoffmann, A. L., & Stark, L. (2019). *Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning*. Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2019.258

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Pedreschi, D., & Giannotti, F. (2018). A survey of methods for explaining black box models. *ArXiv:1802.01933 [Cs]*. http://arxiv.org/abs/1802.01933

Hernandez, D., & Brown, T. (2020, May 5). *AI and efficiency*. OpenAI. https://openai.com/blog/ai-and-efficiency/

Krishner, T. (2020, January 3). 3 crashes, 3 deaths raise questions about Tesla's Autopilot. *AP NEWS*. https://apnews.com/article/ca5e62255bb87bf1b151f9bf075aaadf

Larson, J., & Angwin, J. (2016, December 30). *Bias in criminal risk scores is mathematically inevitable, researchers say*. ProPublica; ProPublica. https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say

Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing*, *234*, 11–26. https://doi.org/10.1016/j.neucom.2016.12.038

MacCarthy, M. (2019, March 15). *The ethical character of algorithms—And what it means for fairness, the character of decision-making, and the future of news*. The Ethical Machine. https://ai.shorensteincenter.org/ideas/2019/1/14/the-ethical-character-of-algorithmsand-what-it-means-for-fairness-the-character-of-decision-making-and-the-future-of-news-yak6m

Metz, R. (2021, February 19). *Google is trying to end the controversy over its Ethical AI team. It's not going well*. CNN. https://www.cnn.com/2021/02/19/tech/google-ai-ethics-investigation/index.html

Mitchell, A., Gottfried, J., & Matsa, K. E. (2015, June 1). Facebook top source for political news among millennials. *Pew Research Center's Journalism Project*. https://www.journalism.org/2015/06/01/facebook-top-source-for-political-news-among-millennials/

Popat, K., Mukherjee, S., Yates, A., & Weikum, G. (2018). Declare: Debunking fake news and false claims using evidence-aware deep learning. *ArXiv:1809.06416 [Cs]*. http://arxiv.org/abs/1809.06416

Ramsundar, B., Kearnes, S., Riley, P., Webster, D., Konerding, D., & Pande, V. (2015). Massively multitask networks for drug discovery. *ArXiv:1502.02072 [Cs, Stat]*. http://arxiv.org/abs/1502.02072

Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Žídek, A., Nelson, A. W. R., Bridgland, A., Penedones, H., Petersen, S., Simonyan, K., Crossan, S., Kohli, P., Jones, D. T., Silver, D., Kavukcuoglu, K., & Hassabis, D. (2020). Improved protein structure prediction using potentials from deep learning. *Nature*, *577*(7792), 706–710. https://doi.org/10.1038/s41586-019-1923-7

Silverman, C. (2016, November 16). *This analysis shows how viral fake election news stories outperformed real news on Facebook*. BuzzFeed News. https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook

Stilgoe, J. (2018). Machine learning, social learning and the governance of self-driving cars. *Social Studies of Science*, *48*(1), 25–56. https://doi.org/10.1177/0306312717741687

Sünderhauf, N., Brock, O., Scheirer, W., Hadsell, R., Fox, D., Leitner, J., Upcroft, B., Abbeel, P., Burgard, W., Milford, M., & Corke, P. (2018). The limits and potentials of deep learning for robotics. *The International Journal of Robotics Research*, *37*(4–5), 405–420. https://doi.org/10.1177/0278364918770733

The partnership on AI. (2021). The Partnership on AI. https://www.partnershiponai.org/

Vincent, J. (2019, April 3). *The problem with AI ethics*. The Verge. https://www.theverge.com/2019/4/3/18293410/ai-artificial-intelligence-ethics-boards-charters-problem-big-tech

Wakabayashi, D. (2018, March 19). Self-driving Uber car kills pedestrian in Arizona, where robots roam. *The New York Times*. https://www.nytimes.com/2018/03/19/technology/uber-driverless-fatality.html

Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., West, S., Richardson, R., Schultz, J., & Schwartz, O. (2018). *AI Now Report 2018*. AI Now. https://ainowinstitute.org/AI_Now_2018_Report.pdf

Winner, L. (1980). Do artifacts have politics? *Daedalus*, *109*(1), 121–136.

Winner, L. (1986). *The whale and the reactor: A search for limits in an age of high technology* (Nachdr.). Univ. of Chicago Press.

Woolgar, S. (1990). Configuring the user: The case of usability trials. *The Sociological Review*, *38*(1), 58–99. https://doi.org/10.1111/j.1467-954X.1990.tb03349.x