THE IMPACT OF AI IN FUTURE START-UPS

Although the field of *Artificial Intelligence* (AI) was created in 1956 officially [1], its journey was troublesome with peaks and valleys and usually without large scale commercial deployment. This changed with Geoffrey Hinton's famous paper in 2006, "A Fast Learning Algorithm for Deep Belief Nets", where he showed how to train a neural network that is capable of recognizing hand-written-digits with above 98% precision[2]. This technique was branded as *Deep Learning* (DL). With this paper, AI rose to prominence once again, after two AI winters. Since the beginning of this era, we have come a long way in the field of DL and achieved many great things. DL applications have beaten human level in Large Scale Image Recognition Challenge [3], which can be seen in Figure 1.

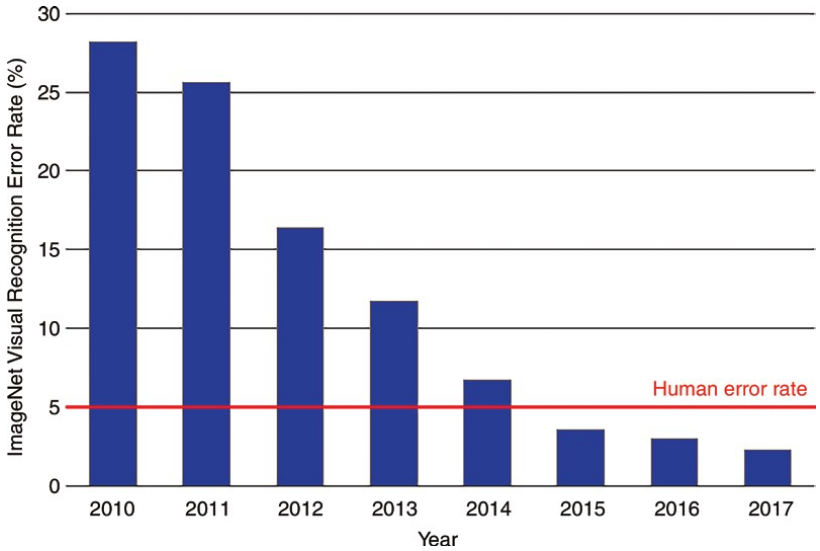

Figure 1: Error rates on the ImageNet Large-Scale Visual Recognition Challenge. Accuracy dramatically improved with the introduction of deep learning in 2012 and continued to improve thereafter. Humans perform with an error rate of approximately 5%. [4]

IBM Watson has beaten two champions in quiz answering show called *Jeopardy!*[5]. DeepMind's AlphaGO has beaten a world champion in the strategy game called GO[6]. The author finds this development especially interesting, because as he learned GO for the first time many years ago, there was prize for a computer that can play GO at 1 Dan level (advanced amateur level). And recently, DeepMind hit the news once again by solving the problem of protein folding with AlphaFOLD[7]. These are just some examples to illustrate the impact of AI in our modern world.

Most likely, the reader is already aware of the majority of all the advantages that AI can bring to humanity through start-ups. But we must also be wary of the potential of what we are aiming to create artificially. Today, our intelligence has enabled us to use nuclear fission to provide us with energy, but the very first thing that we used this technique for was utter destruction that was unmatched by anything we have done until that day. This is the potential of the object that we are aiming to create. It has potential for both creation and destruction. Although, the author of this paper doesn't share the concerns of Stephen Hawking or Elon Musk, he definitely thinks that we are threading on thin ice when it comes to AI due to its sheer capabilities and immense potential.

Our brains gather immense amounts of data. The human body sends 11 million bits per second[8] to the brain. By the same token, AI systems also require immense amounts of data to be trained on. The amount of data for large scale problems is in itself a problem. It has been a long discussion about which one is more important: having a better algorithm or having more data. Microsoft researchers Michele Banko and Eric Brill published a paper in 2001. In this paper, they showed very different *Machine Learning* (ML) algorithms, even relatively simple ones, can perform almost as good as any other algorithm once they were given enough data[9]. The results of different algorithms can be seen in the figure below. This notion is also supported by another popular paper called “Unreasonable Effectiveness of Data” by Peter Norvig et al[10].

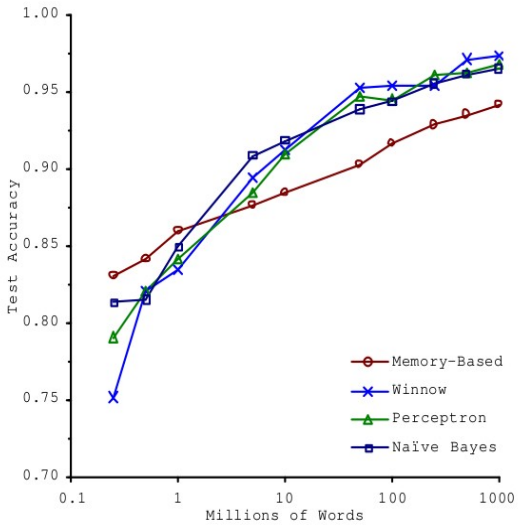

Figure 2: Learning Curves for Confusion Set Disambiguation[9]

However, finding and/or creating large, high quality datasets with annotations is a challenging task. Gathering, cleaning and preparing the data for an ML solution is usually hardest and most time-consuming task of an ML solution. This is probably going to be an issue for the start-ups to come. This brings us to the next topic, the quality of data. Even the best algorithms can produce horrible results if they are fed with bad data. This was exactly what happened with Microsoft’s Tay which was an artificial intelligence chatter bot. Tay was designed to be an interactive bot that can learn from its interactions with twitter users. It was supposed to become smarter as it engaged with people. Instead, it took less than 24 hours for users to turn it into a racist, antisemitic bot and Microsoft had to take it down[11]. Another example is the incident with *PULSE: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models.* In this work, the authors aimed to transform blurry, low-resolution images into sharp, realistic, high-resolution images. In order to train the algorithm they used Flicker Face HQ Dataset[12]. Later, it was found out that this algorithm turns black people to white people, including Barack Obama[13]. And the reason for it was the fact that the dataset it was trained on was overwhelmingly white people. Therefore, it can be said that the amount and quality of data that a start-up has is going to be one of the most influential factors that determine the success of the business in the coming years.

However, the trouble is not over when we have the high quality data. Furthermore, it can be said that it actually worsens. With high quality data, we can create high quality AI systems. But what happens if it is too good? Clearview is a US-based start-up that specializes in facial recognition. They use public images to train their system and justify this under the first amendment of the US Constitution. Their system was so exceptional, it could actually find a high-school photo of a CNN

Journalist online, during their interview[14]. That begs the question how much can AI system know about us. This was illustrated by Target, which is a US-based chain store. Target uses ML to predict what its customers' need based on their purchase history and sends discount coupons to these customers. In one incident, Target predicted that a high-school girl being pregnant and sent discount coupons for baby-related material, before her own family realized that she was pregnant[15]. If a chain store can predict this, what can a company like Google or Facebook predict about us? What do they know about us? And what can be done with this information? Unfortunately, these were illustrated by the Cambridge Analytica incident, in which a company, with this name, used Facebook's data to help the Republican Party to sway key voter bases during 2016 Presidential Election in US[16]. It is astonishing that a company with the premise of just connecting people can have such profound effects on the electoral cycle. This leads to the conclusion that AI will change the nature of start-ups in the future. Just a decade ago, a start-up could be considered as a pure commercial endeavor. But today, with the possibility of such profound effects on our lives, the start-ups are going to be burdened with immense ethical implication of their work which can even cause massive unemployment waves. Therefore, the author believes AI already made and is going to keep making the start-ups much more than a commercial endeavor. They can have effects worldwide.

The author is aware of the dark landscape he drew in the scope of this work. However, he also believes the artificial intelligence is one of the, if not the greatest scientific endeavor that humankind ever undertook until today. Despite all its potential dangers, it also has the potential to change the world for the better. Probably, a handful of start-ups are going to play a vital role in the coming decades to transform the world as we know it. The author aims to spearhead these transformations.

References

1. McCarthy, J., Minsky, M. L., Rocherster, N., & Shannon, C. E. (1955). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. *AI Magazine, 27*(4).
2. Hinton, G. E., Osindero, S., & Teh, Y. (2006). A fast learning algorithm for deep belief nets. *Neural Computation, 18*(7), 1527-1554. doi:10.1162/neco.2006.18.7.1527
3. Olga Russakovsky*, Jia Deng*, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei. (* = equal contribution) **ImageNet Large Scale Visual Recognition Challenge**. *IJCV,* 2015
4. Langlotz, C. P., Allen, B., Erickson, B. J., Kalpathy-Cramer, J., Bigelow, K., Cook, T. S., . . . Kandarpa, K. (2019). A roadmap for foundational research on artificial intelligence in MEDICAL Imaging: From the 2018 Nih/rsna/acr/the Academy Workshop. *Radiology, 291*(3), 781-791. doi:10.1148/radiol.2019190613
5. Markoff, J. (2011, February 16). Computer wins on 'jeopardy!': Trivial, it's not. Retrieved March 01, 2021, from https://www.nytimes.com/2011/02/17/science/17jeopardy-watson.html
6. AlphaGo: The story so far. (n.d.). Retrieved March 01, 2021, from https://deepmind.com/research/case-studies/alphago-the-story-so-far
7. AlphaFold: Using AI for scientific discovery. (n.d.). Retrieved March 01, 2021, from https://deepmind.com/blog/article/AlphaFold-Using-AI-for-scientific-discovery
8. Physiology. (n.d.). Retrieved March 01, 2021, from https://www.britannica.com/science/information-theory/Physiology
9. Banko, M., & Brill, E. (2001). Scaling to very very large corpora for natural language disambiguation. *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics - ACL '01*. doi:10.3115/1073012.1073017
10. Halevy, A., Norvig, P., & Pereira, F. (n.d.). The unreasonable effectiveness of data. Retrieved March 1, 2021, from https://research.google.com/pubs/archive/35179.pdf
11. Vincent, J. (2016, March 24). Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day. Retrieved March 01, 2021, from https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist
12. Menon, S., Damian, A., Hu, S., Ravi, N., & Rudin, C. (2020). PULSE: Self-supervised Photo Upsampling via latent space exploration of generative models. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. doi:10.1109/cvpr42600.2020.00251
13. Vincent, J. (2020, June 23). What a machine learning tool that TURNS Obama White can (and can't) tell us about AI bias. Retrieved March 01, 2021, from

https://www.theverge.com/21298762/face-depixelizer-ai-machine-learning-tool-pulse-stylegan-obama-bias

14. https://www.youtube.com/watch?v=pGJNXG2vmZw&t=7s&ab_channel=CNNBusiness
15. Duhigg, C. (2012, February 16). How companies learn your secrets. Retrieved March 01, 2021, from https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=1&_r=1&hp
16. Confessore, N. (2018, April 04). Cambridge Analytica and Facebook: The scandal and the fallout so far. Retrieved March 01, 2021, from https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html