ARE WE IN THE DIGITAL DARK TIMES? HOW THE PHILOSOPHY OF HANNAH ARENDT CAN ILLUMINATE SOME OF THE ETHICAL DILEMMAS POSED BY MODERN DIGITAL TECHNOLOGIES

Damian Gordon, Anna Becevel

Technological University of Dublin (Ireland)

Damian.X.Gordon@TUDublin.ie; Anna.Becevel@TUDublin.ie

ABSTRACT

Philosophers are not generally credited with being clairvoyant, and yet because they recognise, record and reflect on trends in their society, their observations can often appear prescient. In the field of the ethics of technology, there is, perhaps, no philosopher whose perspective on these issues is worth examining in detail more than that of Hannah Arendt, who can offer real perspective on the challenges we are facing with technologies in the twenty-first century. Arendt, a thinker of Jewish-German origin, student of Martin Heidegger and Karl Jaspers, encountered her life turning point when she was forced into becoming a refugee as the world was shaken by a force of unimaginable brutality that she was one of the first to name "totalitarianism" (Baerh, 2010). She was an independent thinker, separating herself from schools of thought or ideology. Investigating totalitarianism was her ruling passion, and as such her political thought often overshadows her major contribution to other branches of philosophy. Arendt is best known for her accounts of Adolf Eichmann and his trial, and the concept of "banality of evil", though her perspective on politics was driven by a precise and original theory of action. While the latter is inextricably connected to her political perspective, it is also supported by a sharp ontological reflection of social structures and anthropological reflections.

INTRODUCTION

In her 1963 book "*Eichmann in Jerusalem: A Report on the Banality of Evil*" Arendt introduced the notion of the "banality of evil", meaning that very often evil acts are not committed by fanatics or psychopaths, but instead by very normal people who rely on simplistic clichés to justify their actions, rather than thinking for themselves. The concept is not meant to indicate the the deeds of Eichmann and others were in any way ordinary, but that the self-justification they used and complacency of their acts was wholly unexceptional Later in 1978 Arendt warned that "*clichés, stock phrases, adherence to conventional, standardized codes of expression and conduct have the socially recognized function of protecting us against reality, that is, against the claim on our thinking attention that all events and facts make by virtue of their existence*." This, unfortunately, seems to marry well with statements made by ex-employees of social media companies in the past 10 years, and most notably on the 2020 Netflix docudrama "The Social Dilemma" where former employees of social media companies like Facebook, Google, Twitter, Mozilla, and YouTube describe how their companies unthinkingly nurture addiction to their product, and help spread conspiracy theories and disinformation. This film also explores the issue of the impact of social media on user's mental health (and particularly the mental health of adolescents and rising teen suicide rates). The film cites key statistics such as a 62% increase in hospitalizations for American females aged 15–19 and a 189% increase in females aged 10–14 due to self harm, beginning in 2010–2011, which the companies are aware of, but their employees justify

their actions with mindless clichés like "*it's good for my career*", "*I'm just doing my job*", or "*I was only following orders*".

PROPAGANDA

Arendt (1951) said "*In an ever-changing, incomprehensible world the masses had reached the point where they would, at the same time, believe everything and nothing, think that everything was possible and that nothing was true. ... Mass propaganda discovered that its audience was ready at all times to believe the worst, no matter how absurd, and did not particularly object to being deceived because it held every statement to be a lie anyhow. The totalitarian mass leaders based their propaganda on the correct psychological assumption that, under such conditions, one could make people believe the most fantastic statements one day, and trust that if the next day they were given irrefutable proof of their falsehood, they would take refuge in cynicism; instead of deserting the leaders who had lied to them, they would protest that they had known all along that the statement was a lie and would admire the leaders for their superior tactical cleverness*."

Gaber and Fisher (2021) looked at political lying in general, and specifically at the lies told during the referendum over *Brexit* (the proposed withdrawal of the United Kingdom from the European Union and the European Atomic Energy Community), as well as the political campaigns of Donald Trump in 2016 and 2020. They argue that until recently politicians would avoid telling outright lies as they felt the political consequences would be too severe, however in the past decade there has been a change, and politicians have learned to use "strategic lies" (lies that are both attention-grabbing and agenda-setting) with apparently no consequences. They argue that this approach has its roots in the practice of "spin" which grew significantly in the political sphere in the 1990s (Street, 2011), where media advisors presented biased interpretations of events to influence public opinion about a particular issue. Since then, with the growth of social media and increased professionalization of media advisors, "spin" has been transformed into "strategic lying". An early example of this new form of lying was Donald Trump's claim to have "proof" that Barack Obama was not born in the United States (starting the "birther" movement), and he claimed that he was going to send a team of private investigators to Hawaii to explore the truth of these claims, and would donate $5 million to charity if definitive evidence was found that President Obama was, in fact, born on the USA. There is no record of a team of private investigators to Hawaii nor is there no record of Donald Trump donating $5 million to charity in spite of the fact that President Obama did publish his birth certificate, however, Donald Trump continued to reiterate his claims after the evidence, to create the impression of Obama "otherness" in the mainstream media. A similar example occured when current British Prime Minister Boris Johnson, who in his former role as a journalist made false claims about the European Union, including that they were going to ban certain electrical appliances (including vacuum cleaners, kettles, toasters and lawnmowers), and ban bananas that were too bendy. These claims were fact-checked as false a number of times, but nonetheless Johnson continued to reiterate them, including during the runup to the Brexit referendum. Additionally Johnson repeatedly made a claim that the UK sends £350 million a week to the EU, which is a gross figure, in reality the UK sent a net figure of £210 million a week to the EU. Johnson even went so far as to print this claim on the side of a bus, and claim that money could go to the UK national health services instead (See Figure 1 below).

Figure 1. The "Brexit" Bus

As Alastair Campbell (former Press Secretary and Director of Communications to former British Prime Minister, Tony Blair) stated: "*I am afraid we have entered a post-truth, post-shame world. The Washington Post says Donald Trump tells 12 lies a day. His predecessors would have been hounded out of office for one in a term. Boris Johnson won a referendum by lying. His reward? He was made Foreign Secretary and he is now going to be the Prime Minister. There is no shame!*" [21].

Unfortunately this appears to be correct for some people, it appears that for some of the voting public, they have been convinced that we live in a "post-truth" era, where people believe that truth is a relative concept, and they feel empowered to choose their own version of reality, where existing beliefs and prejudices are more important than facts, particularly if the existing beliefs and prejudices are reiterated and amplified by their political leaders. This also means that those individuals have abandoned conventional criteria of evidence and fact-checking, and in exchange they are not obliged to have to think about difficult or unsettling realities (Lewandowsky, et al., 2017). This "filter bubble" is comforting to those who live in them, and the residents of these isolationist spaces tend to reward the politicians who help maintain these bubbles.

SOCIAL MEDIA

Combining two quotes from Arendt, in 1963 she said that "*The trouble with Eichmann was precisely that so many were like him, and that the many were neither perverted nor sadistic, that they were, and still are, terribly and terrifyingly normal. From the viewpoint of our legal institutions and of our moral standards of judgment, this normality was much more terrifying than all the atrocities put together*" and in 1978 she warned "*The sad truth of the matter is that most evil is done by people who never made up their minds to be or do either evil or good*". This could very well be used as a lens for both the developers of social media, and the users of social media. It is important to recognise that social media companies want to keep their users on their sites for as long as possible, therefore they use manipulative approaches, such as "Digital Nudges" (Acquisti 2009) which are small interventions that guide choices without restricting them, such as timely reminders, personalized messages, or small digital rewards. As users are using social media, more and more behavioural data is being collected about them so that an increasingly complex and comprehensive digital model of each individual is created, and the correct means to extent that user's session time will be identified to expose that user to as much advertisement as possible. This, in and of itself, might seem innocuous, but when people are using a range of social media platforms, this can have unintended, catastrophic consequences.

[21] Alistair Campbell, Depression and the politics of mental health: Alastair Campbell on ABC Radio National Breakfast, July 22, 2019

Research by McHugh, et al. (2018) suggests that social media usage can cause symptoms of post-traumatic stress disorder (PTSD) in adolescents, and they also found that these adolescents engage in coping mechanisms to help to reduce the long-term negative effects of exposure.

In the context of lies being spread by social media, research by Vosoughi et al. (2018) indicated that lies spread "*significantly farther, faster, deeper, and more broadly than the truth*". They indicate that there are two key reasons for this: (1) *Confirmation Bias*, people tend to notice and remember things that help confirm their own worldview, and (2) *Repetition*, the more often a lie is repeated, the more likely it is to be believed, even it is refuted each time, because this lie has already been processed by the brain it takes no additional cognitive load to process it again (e.g. the way Donald Trump kept claiming the election was "stolen" in 2020). The individuals creating the lies know that they are doing harm, but the (potentially) millions of people who believe the misinformation, who spread this misinformation, and who embroider the misinformation, are not intending to do harm, but are part of a larger process that is detrimental to all participants.

MACHINE LEARNING

Famously Arendt (1962) wrote "*I have always believed that, no matter how abstract our theories may sound or how consistent our arguments may appear, there are incidents and stories behind them which, at least for ourselves, contain as in a nutshell the full meaning of whatever we have to say. Thought itself - to the extent that it is more than a technical, logical operation which electronic machines may be better equipped to perform than the human brain arises out of the actuality of incidents, and incidents of living experience must remain its guideposts by which it takes its bearings if it is not to lose itself in the heights to which thinking soars, or in the depths to which it must descend.*". This is a profound insight into the problems of machine learning, Arendt is arguing that real thinking can only occur through the lens of human experience, and an abstract representation of ideas do not in fact encompass the totality of thinking. The world Arendt describes is a vively and turmoiled one, where each individual acts freely in their environment while simultaneously creating a shared political space, a world that our current technologies seem unable to describe at this stage due to the limitations of machine learning. The concept of Machine Learning was developed by Samuel (1959), and generally consists of the following steps (Langley, 2011):

1. Collecting data about a significant number of examples of particular scenario; the data usually consists of key descriptors or characteristics;
2. The data is analysed using a computer program that attempts to uncover rules or relationships between the descriptors;
3. The rules are then used to predict the outcomes of new examples of the scenario that haven't been presented to the computer program yet.

This approach has led to a growing catalogue of disastrously poor results, for example, in 2014 Amazon began developing a computer program to help in personnel recruitment, and after a year they discovered that the system was sexist in operation, and would always prefer male candidates to female ones, and eventually they abandoned that system. What a subsequent analysis found was that because a significant majority of existing successful candidates were male, the system was fed an abundance of data on male candidates and less on female candidates (Fumiko, et al., 2020). In 2013 IBM partnered with *The University of Texas MD Anderson Cancer Center* to develop a new "Oncology Expert Advisor" system that would ultimately lead to a cure for cancer. Unfortunately, the resulting system gave

erroneous, and downright dangerous cancer treatment advice, and had to be finally abandoned in 2018, simply because the IBM engineers trained their software on synthesized data, rather than real patient data (Strickland, 2019). Hendrycks, et al. (2019) set out to show the limitations of machine learning algorithms, by selecting 7,500 specifically curated images of a large dataset of images of animals, insects and other natural phenomena, they reduced the effectiveness of a machine learning algorithm from 92% to 2%.

The problem with these systems is that they rely almost completely on data to draw their conclusions, and if data is misconfigured, then the rules that the system deduces are flawed. Additionally, it is only possible for some machine learning systems to express the rules that they have deduced in a manner that a human being can understand, for other systems the manner in which they deduce and encode these rules cannot be expressed as text, and they are therefore said to lack *explainability* (London, 2019). This is a very serious issue, if the systems can't even explain why they are making decisions, it makes trusting those decisions more difficult, so much so that the European Union is regulating the use of machine learning, and requiring that it must be of the explainable variety (Hamon, et al., 2020). As well as bias in data, other issues that appear to cause poor decision-making includes:

- *Underfitting*, is where the rules that the systems deduced aren't a sufficiently detailed model of the complexity of the data presented to the system.
- *Overfitting*, is where the rules that the system deduced are too specifically tailored for the data presented to the system, and can't accurately generalise the lessons learned.
- *Undersampling*, is where the distribution of data in one characteristic of the dataset doesn't reflect the population under investigation because one group is under-sampled, for example, if one race of people is under-represented in a dataset about a group of people.
- *Oversampling*, is where the distribution of data in one characteristic of the dataset doesn't reflect the population under investigation because one group is over-sampled, for example, if one race of people is over-represented in a dataset about a group of people.
- *Proxy Variables*, is where you have to use a stand-in variable because it isn't possible to represent a characteristic directly. So, for example, if you can't measure people's level of health, it might be easier to measure how much money people spend on health, as a proxy to measure level of health. Unfortunately, this doesn't take into account wealth level.
- *Missing Variables*, is where the characteristics selected in the dataset are not everything that should be taken into account to have a representation sample.
- *Underspecification*, identified in 2020, is where the characteristics chosen in the dataset don't represent the totality of the key features required to model the data (D'Amour, et al., 2020).
- *Data Scarcity*, is where insufficient data is presented to the system, and therefore, there isn't enough variation in the data to represent all of the potential cases the system will encounter.

These all simply point to an inherent flaw in the development of machine learning systems, that unless the exact parameters of the problem are already fully understood, it might not be possible to identify the correct dataset characteristics to accurately represent the problem. The truth of the situation is that the datasets used by these systems cannot capture the full diversity of real-world experience. When considering the phenomenological nature of action (Dal Lago, 2016), not being able to describe the complexity of human experience doesn't only mean missing on diversity, but missing on the chance to obtain it at any stage. Human experience is the way through which agents reveal themselves and

simultaneously accept the risks implied by this revelation. The exposure of human experience is a necessary and sufficient condition to create a political space where the individuals can work-together and regulate themselves in environments not regulated by governments such as the internet. (Arendt, 1958). Moreover, the people who create and curate datasets bring with them a series of tacit assumptions, and even cognitive biases, about the problem that make a representative dataset less possible. One common erroneous assumption that many people make is how frequently unusual events occur (Paulos, 1988), and this can led to the creation of unrepresentative datasets; again as Arendt says: "*incidents of living experience must remain its guideposts by which it takes its bearings*".

Unfortunately, modern technology is contributing to cognitive biases, for example, since 2009 the Google search engine has incorporated a "Personalized Search" which means that results returned are not the same for everyone, instead they are based on each individual user's personal behaviour and interests as well as those of the user's social circle (Zamir and Korn, 2020). This creates a "filter bubble" that creates polarization and echo chambers, and results in an exogenous isolation effect, as well as a lack of full discussion of the topics (Min, et al., 2019). This issue was highlighted by Arendt's when she stated that: "*To hold different opinions and to be aware that other people think differently on the same issue shields us from Godlike certainty which stops all discussion and reduces social relationships to an ant heap*".

CONCLUSIONS

These issues are a small sampling of the perspective and insight that Arendt can give us on computer ethics, and her reflections can be both thought-provoking and illuminating in terms of how we should develop and use new technologies. As mentioned at the start, philosophers are not generally credited with being clairvoyant, and yet Arendt's perspectives might provide a way forward in the modern world. And her work, and the work of other 20th century philosophers, urgently need to be re-examined in the light of the serious political decisions that are being made by so many in such a mindless way.

KEYWORDS: Digital Ethics, Hannah Arendt, Machine Learning, Conglomerations.

REFERENCES

Acquisti, A. (2009) "Nudging privacy: The behavioral economics of personal information", *IEEE Security & Privacy*, 7(6), 82-85.

Arendt, H. (1951). *The Origins of Totalitarianism*. New York: Schocken.

Arendt, H. (1958). *The Human Condition*. Chicago: University Chicago Press.

Arendt, H. (1962). Action and the Pursuit of Happiness. In A. Dempft, H. Arendt & F. Engel-Janosi (Eds.), *Politische Ordnung und Menschliche Existenz. Festgabe für Eric Voegelin zum 60 Geburtstag* (pp. 1–16). Munich: C. H. Beck.

Arendt, H. (1963). *Eichmann in Jerusalem*. New York: Viking Press.

Arendt, H. (1978). *The Life of the Mind*. Harcourt Brace Jovanovich.

Baerh, P. (2010). *Hanna Arendt, Totalitarianism, and the Social Sciences*. Stanford: Stanford University Press.

D'Amour, A., Heller, K., Moldovan, D., Adlam, B., Alipanahi, B., Beutel, A., Chen, C., Deaton, J., Eisenstein, J., Hoffman, M.D. & Hormozdiari, F. (2020). Underspecification presents challenges for credibility in modern machine learning. Retrieved from https://arxiv.org/abs/2011.03395v2

Dal Lago, A. (2016). [Introduction]. In H. Arendt, *Vita Activa: la condizione umana* (pp. 6-38). Milano: Bompiani.

Farhall, K., Carson, A., Wright, S., Gibbons, A., Lukamto, W. (2019) "Political Elites' Use of Fake News Discourse Across Communications Platforms", *International Journal of Communication*, 13, 23.

Fumiko, K., Arai, H. & Ema, A. (2020). Ethical Issues Regarding the Use of AI Profiling Services for Recruiting: The Japanese Rikunabi Data Scandal. Retrieved from https://arxiv.org/abs/2005.08663

Gaber, I., Fisher, C. (2021). ""Strategic Lying": The Case of Brexit and the 2019 UK Election", *The International Journal of Press/Politics*, 1940161221994100.

Hamon, R., Junklewitz, H. & Sanchez Martin, J. (2020). Robustness and Explainability of Artificial Intelligence. EUR 30040 EN, JRC119336. Luxembourg: Publications Office of the European Union. Retrieved from https://publications.jrc.ec.europa.eu/repository/handle/JRC119336

Hendrycks, D., Zhao, K., Basart, S., Steinhardt, J., & Song, D. (2019). Natural adversarial examples. Retrieved from https://arxiv.org/abs/1907.07174

Langley, P. (2011). The changing science of machine learning. Machine Learning, 82(3), 275–279.

Lewandowsky, S., Ecker, U. K., Cook, J. (2017) "Beyond misinformation: Understanding and coping with the "post-truth" era". *Journal of Applied Research in Memory and Cognition*, 6(4), pp. 353-369.

London, A. J. (2019). Artificial Intelligence and Black-Box Medical Decisions: Accuracy versus Explainability. *Hastings Center Report*, 49(1), 15-21.

McHugh, B. C., Wisniewski, P., Rosson, M. B., Carroll, J. M. (2018) "When social media traumatizes teens: The roles of online risk exposure, coping, and post-traumatic stress", *Internet Research*, 28(5), pp. 1169-1188.

Min, Y., Jiang, T., Jin, C., Li, Q., & Jin, X. (2019). Endogenetic Structure of Filter Bubble in Social Networks. *Royal Society Open Science*, 6(11), 190868.

Paulos, J. A. (1988). *Innumeracy: Mathematical Illiteracy and its Consequences.* New York, NY: Hill and Wang.

Samuel, A. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, 3(3): 210–229.

Strickland, E. (2019). IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. *IEEE Spectrum*, 56(4): 24-31.

Street, J. (2011) *Mass Media, Politics and Democracy*, Palgrave Macmillan.

Vosoughi, S., Roy, D., Aral, S. (2018) "The Spread of True and False News Online", *Science,* 359(6380), pp. 1146–51

Zamir, O., & Korn, J. (2020). *U.S. Patent No. 10,691,76*5. Washington, DC: U.S. Patent and Trademark Office.