Privacy in the Age of AI: Navigating the Ethical Dimensions of Machine Learning

In the modern day constantly the achievements and merits of artificial intelligence are talked of, but quite often neglected is the ethical toll of such development. The ethics behind machine learning is a topic spanning multiple areas of engineering and philosophy (causing new initiatives and branches of research such as ANU's recent field of *Cybernetics*[1]) and attempts to explore these ramifications. At the forefront of these moral issues is the use of intellectual and artistic property to train many frameworks of popular machine learning based A.I. Due to the newness of it all, laws addressing ownership are fragile, undeveloped and vague. The general consensus (reference law here) is that copyrighted property requires permission from the holder for use in training an AI[2]. But while that's the legal side and like most things law does not dictate morality. The privacy of the individual is another problem at the forefront of Machine learning. The accessibility and handling of publicly accessible data by algorithms poses security and moral issues.

The valuation of physical goods in our current era are becoming impertinent, with our society shifting to that of intellectual property. AI models have been wrongfully trained and threaten the privacy of the properties creators. Companies used the LAION database to access copyrighted works that would lead to a class action lawsuit against them[3], it is completely unethical to steal these works and artists' likeness and works without either consent or compensation. Currently the laws regarding copyright seem to allow for this use of property to dodge punishment[4]. To take the art and works of someone and then feed it to the very thing that causes their obsolescence, lacks any moral ground. This issue contributes as well to the wrongful stripping away of privacy, having personally owned intellectual property and information to be used without your consent or knowing extends to this morally dubious violation. The largest issue with having it legal to use these works is that it affects ownership. Although copyrights would previously protect this it does tear away this veneer of privacy and security previously granted. Ultimately the training of aspects of these models could avoid this problem if we were to enforce proper policy regarding copyright for Machine Learning and consent laws regarding users information.

Contemporary algorithms are based primarily on user information. Most social medias use it for the "good of the consumer"[5] to allow them to consume content tailored specifically for them. While this may seem to lack malicious intent, the information is often sold and transferred usually with the users knowing[6]. While we often consent to this use and eventual sale of information, it seems unethical to take user info and monetise (without any form of reimbursement) it in common instances forcefully[7]. It's considered the cost of using a website to donate your location, personal details and credential and identifying information. But while it has become accepted it is still unethical to strip away one's privacy. Having to sell away this data, while seeming at first a choice is becoming less and less of one as our reliance on the internet as a whole only grows. Identity theft is more prevalent than ever with over 10 billion lost each year from cybercrime[8]. This can be attributed to both our culture around the publication of personal information as well as algorithmic processing[9]. Machine learning's terrifying ability to process large swathes of data poses a threat to our privacy. The handling of this data by these algorithms contributes to both criminal issues as well as a slowly disappearing privacy among the general population. The level of publicly available information on each individual is both a security concern and a moral one.

---

[1] ANU School of Cybernetics

[2] This paper considers any task performed by a non-human source done with human level intelligence and not made by non biological factors to be AI.

[3] "The artists had not consented to have their copyrighted artwork included in the LAION database; they were not compensated for their involvement, even as companies including Midjourney charged for the use of their tools" - TheNewYorker

[4] "it is much more likely than not" that training systems on copyrighted data will be covered by fair use. But the same cannot necessarily be said for generating content" - The Verge

[5] An example of cookie usage described by Twitter. Most of the reasons are targeting the user using "you" language and how it helps them. Save and honor your preferences.
"Personalize the content you see. Protect you against spam and abuse. Show you more relevant ads."

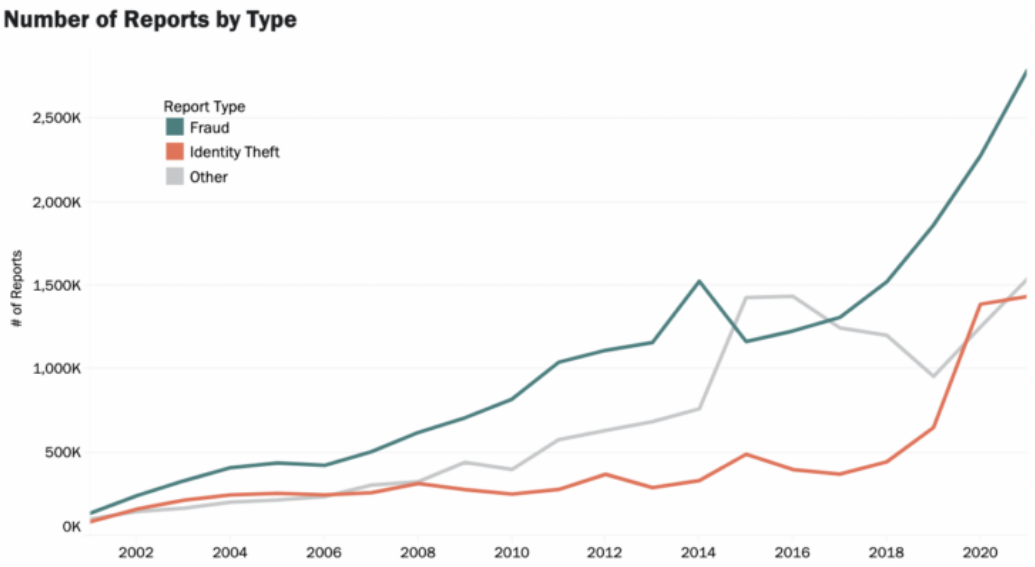
[6] The CCPA (California Customer Privacy Act) is an example of preventative measures being put in place regarding this information. In most places such measures do not exist. In places where laws are not in place there is no requirement to even tell the user that their data is being sold.

[7] According to W3Techs 23.8% of all websites use persistent cookies.

[8] See figure 1 and other information by IdentityTheft.org.

[9] Private information being as easy to access as ever before. I would consider my digital footprint to be small to insignificant considering how rarely I post personal information online, a quick google of my name reveals my face, former school and year group (at the time). Google being the algorithm used to locate and handle act as an example of how these ML models contribute to the access to this information and perpetuation of this culture.

Figure 1:



The main issues with the current use of AI in popular models is privacy. AI art's rise to prominence coming at the economic expense of its successors is completely unethical. The atomisation of an industry comes with its ethical repercussions. The theft by multiple of the largest companies targeting artists' intellectual property showcase this unethical disregard for intellectual property. This, in tangent with the stripping away of the privacy of users of most if not all parts of the internet. The rise of identity theft issues coupled with the increasing proliferation of cookies and other user data tracking followed by the then sale of this information all show this amoral disregard fuelled by the use of ML models. Ultimately, the largest concern of AI is simply its efficiency of handling data. AI isn't the reason for the aforementioned violations of privacy and unethical use of data, but rather its how AI is used to streamline ill-gotten information. Machine learning like anything is a tool, it can be used correctly, but will commonly be used as a tool for profit at the expense of both consumer and creator.