Title: **AI Safety: A Brief Examination of Identity Scams and Implications**

Author: **James Newton-Thomas**

Introduction:

In this paper, I discuss a small but significant aspect of AI safety: identity scams. This specific issue illuminates the larger challenges we as a society face with the introduction of artificial intelligence.

Historical Context:

When Alan Turing proposed the imitation game as a test of intelligence in 1950, his intentions were positive. Turing aimed to provide an objective measure of when a machine could be considered "intelligent". He suggested that if we could not distinguish a machine from a human after a short period of time, it would be reasonable to regard the machine as "intelligent".

Development of AI Platforms:

Building upon Turing's work and that of countless others, we now possess the knowledge and platforms necessary to create AI systems and access the vast amount of data needed to train them. These include AIs capable of generating text, speech, imagery, etc., which we refer to as generative AIs due to their productive capabilities. These systems have proved remarkably effective and are almost certain to become increasingly prevalent.

Implications of Advancements in AI:

The prowess of generative AIs is such that they now appear capable of passing the Turing Test — at least for some of the people, some of the time. This ability to imitate has evolved from a simple tool for subjective classification to a red flag in the context of AI safety. As AI improves, it inadvertently aids the surge of AI identity scams, the propagation of fake news, potential political manipulation, and significant workplace substitution and disruption. Generative AIs can be tasked just as easily to create misinformation as they can valid information, making them an important focus in an era where screen sizes and time are limited and the volume of information matters. They matter.

The Future of Scams and AI:

In line with Turing's insights, and due to the lucrative potential, scams utilizing AI to emulate human activity are likely to become more sophisticated and prevalent. I use the term 'scam' in its strictly criminal sense here, as a dishonest scheme or fraud designed to trick people out of their money. There will certainly be other unethical, though not

necessarily illegal, practices that AI will increasingly be used for, but these fall outside the scope of this brief paper.

The Nature of AI Identity Scams and the Role of Corporations and Government

Scams, and those who perpetrate them, are by nature adaptive and innovative. They are very reflexive to defensive measures against them. Unconstrained by morality and often operating outside the bounds of specific legal jurisdictions, they exploit the existing infrastructure provided by telecommunication companies and websites to gain access and credibility. These systems can be technically sophisticated, suggesting that the most effective defenses will likely continue to be mounted by those on the line of contact - the banks, telecommunications companies, and software houses. These organizations constructed the platforms that scammers now exploit, making them the most technically competent to combat this issue. Nevertheless, they will likely need to be compelled by legislation to do so and supported by government educational programs to promote societal awareness.

However, public education can never fully replace a robust corporate defense strategy. There will always be a segment of the community with less access to education, and therefore more vulnerable and targeted. The government's response should be to increasingly incentivize these frontline organizations by legislating for, and progressively delegating the responsibility to them, and their insurers, for scams conducted on their platforms.

The Larger Danger Posed by AI and its Societal Impact

Whilst the threat and fears of this class of identity scam are real and require a coordinated response, I argue that they pale into insignificance against the larger danger that AI poses. That is, the rapid and mass substitution of human mental labour in our current workspace. The psychological importance of gainful employment cannot be overstated. A lack of it is correlated with crime, substance abuse, physical and mental health deterioration, impoverishment, homelessness and general immiseration and therefore must be considered a safety issue in and of itself.

Whether we like to admit it or not, poverty is also hereditable. Therefore, already disadvantaged members of the Australian community are likely to be disproportionately impacted.

In our 2018 senate paper on the Future of Work and Workers, my colleague and I highlighted the impending danger posed by AI. At the time, various academics and representatives of large tech giants opposed this view. They repeated the common mantra that "historically, technological innovation has tended to increase employment" – an argument known as "The Luddite Fallacy."

The Luddite fallacy argument fails in two significant respects in this context. Firstly, the argument was initially formulated to demonstrate the economic benefits that came

about during the Industrial Revolution, when machines displaced people from agricultural labor to more pooled labor in cities. There is no denying that we are better off now because of what happened then. However, it's crucial to remember the hundreds of thousands of displaced peasants of that time who suffered greatly from this shift. They lived the hunger, homelessness, long treks, disease, crowded factories, low paid work, long hours, pollution, prevalent child labor and for whom our argument is moot. It was a terrible time for them and whilst we may see that time in history as being a mere eyeblink, and hand wave it away as being for the greater eventual good, they lived it.

Secondly, the Luddite fallacy describes a transition from a rural lifestyle to an urban one, where eventual technological progress and efficiencies of scale improved living standards. This does not necessarily mean a similarly positive transition from our current work to future work actually exists. At the very least, students encouraged by the government to enter STEM professions are likely to find the nature of their actual work significantly different from what they trained for, and probably less remunerative.

Evidence of AI Substitution and the Current Tech Landscape

The fact that we are now seeing AI identity scams, or that terms like deep-fake are now in common usage, or that actors, writers and journalists are striking in an attempt to keep themselves relevant is all evidence that AI substitution is not only possible, it is happening. The inescapable conclusion must be that Turing was correct and that Intelligence is not something unique to biological entities. There is also no evidence that intelligence somehow peaks with us, we know that typically animals with a larger brain to body mass are more intelligent. We know that animals with a higher neural density such as birds are smarter relative to their brain size than those with lower neural density. Loosely there appears to be three dimensions at play, speed, scale and complexity and none of these are biologically constrained. On the contrary, quantum computers, fibre optics, meta-materials and many other technological advances exploit properties and capabilities that are biologically not realiseable, at least in scale.

Of the five most valuable companies in the World four of them are tech giants - Apple, Microsoft, Google and Amazon, and Aramco, an energy company being the fifth. Tech matters and the increasing value of those companies shows that it increasingly matters. These same companies are amongst those leading this push to AI utilisation. Five years after the Sentate hearing I addressed, those same tech giants are still pointing to the Luddite Fallacy argument while quietly slashing their own work-forces by upwards of 20%. This is happening even as they pay hundred-million-dollar remuneration packages to senior staff, which significantly boosts their share prices, partly due to payroll reductions. Undoubtedly, it's a good time to be a software company when the software writes itself.

AI Integration, Its Implications, and Future Steps

The mass integration of AI systems into our society is inevitable and simply an extension of the mass integration of communications and knowledge availablility. However, because AI outputs are intrinsically documented and measurable, and because they are repeatable, and they have a much higher coefficient of sameness, aka 'quality' of their output over human powered systems. AI systems will be adopted as substitutes for people as they match the selection criteria within ISO-9000. ISO-9000 is driving us towards automation and AI uptake. In addition AIs never get sick, they never have *bad* days, they never retire. They work 24 hours. They are getting better and they will exponentially benefit those who deploy them. Ultimately because of the speed of communications, you only need one of them. OpenAI's ChatGPT is currently showing the fastest platform uptake in history.

We must prepare as a nation for the widespread introduction of AI and find ways to distribute the generated wealth equitably. We must develop social programs that allow people to maintain relevance and psychological health during this transition. We can not rely on the “magic of capitalism” to distribute this wealth although it can still be a valuable asset in its creation. We need to recognise that we are a Commonwealth and to ensure that our common wealth does not haemorrhage offshore or simply redistributed to benefit a first adopter minority.

Recommendations

1. Recognise both the opportunities and threats of AI and prepare for societal change.

2. Do not attempt to regulate access to AI. The best defence against AI is likely another AI therefore sponsor, develop, deploy and maintain these so people can be protected in real-time whilst having the assistive benefits AI can bring.

3. Do not allow foreign firms to monopolize AI. The Australian Government should commission its own state approved, developed and continuously updated, large language model and make it available to all Australians as a public utility.

James Newton-Thomas

2023