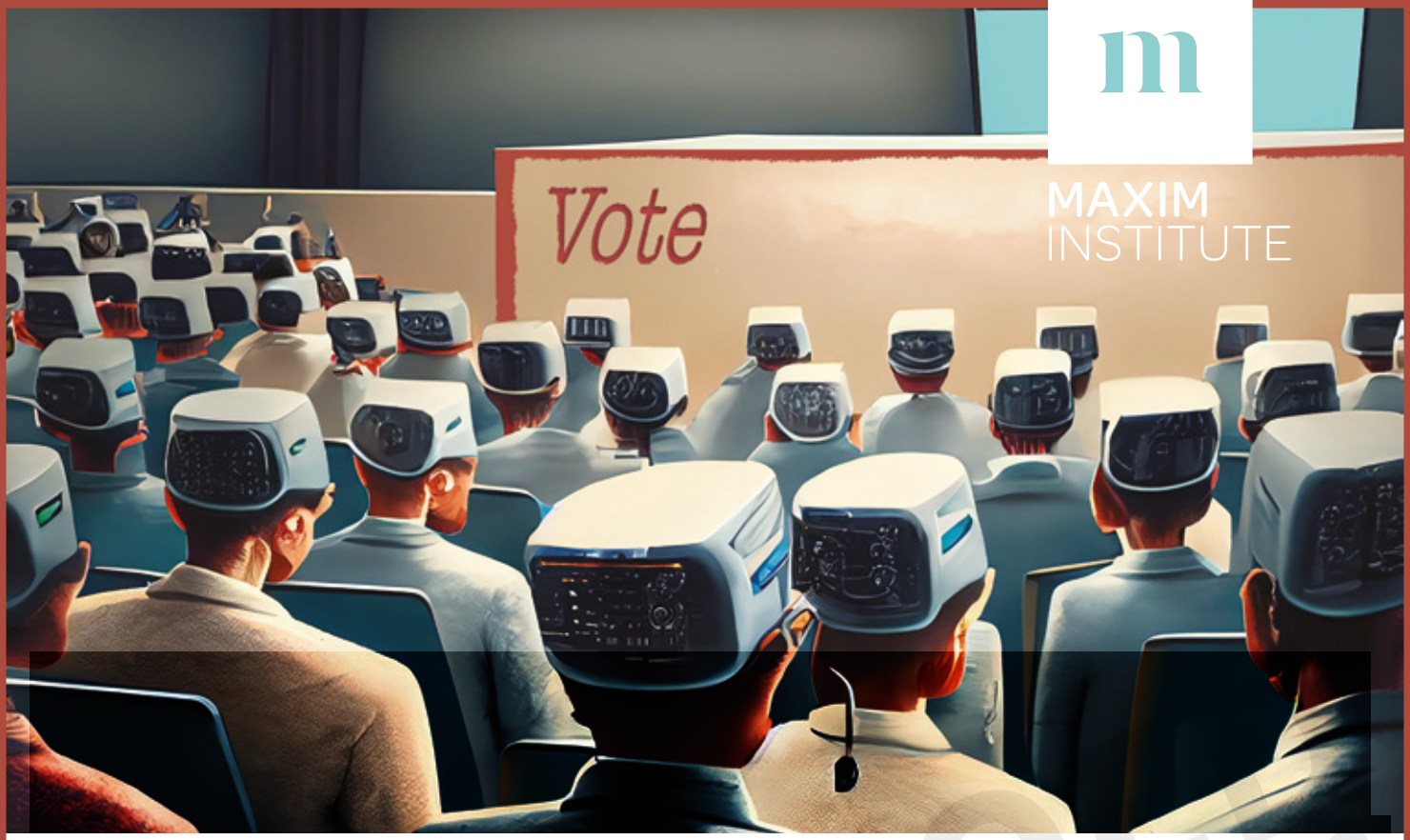


MAXIM INSTITUTE

DISCUSSION PAPER

HOW AI IS CHANGING DEMOCRACY: NUDGING, MICROTARGETING, AND EPISTEMIC BUBBLES

DR PAUL HENDERSON,
DR JONATHAN COLE
AND NATASHA BAULIS

First published in August 2023 by Maxim Institute
PO Box 49 074, Roskill South, Auckland 1445, New Zealand
Ph (0064) 9 627 3261 | Fax (0064) 9 627 3264 | www.maxim.org.nz

978-1-7386020-2-5 – PDF
978-1-7386020-3-2 – Hard copy

Maxim Institute is an independent think tank, working to promote the dignity of every person in New Zealand, by standing for freedom, justice, and compassion.

At the core of our work is an understanding that it is ideas that will shape society over time, so we're engaged in long-term research and analysis of the policies and ideas that inform our nation's social, political, and cultural practices.

To find out more, head to www.maxim.org.nz.

MAXIM INSTITUTE

HOW AI IS CHANGING DEMOCRACY: NUDGING, MICROTARGETING, AND EPISTEMIC BUBBLES

DR PAUL HENDERSON, DR JONATHAN COLE AND NATASHA BAULIS

Table of Contents

About the authors

Dr Paul Henderson, Senior Research Fellow. Paul's background is in education. Working from DPMC, under the Right Hon John Key, he was secretary to a cross-party committee tasked to review school choice and educational reform. He holds degrees from Aberdeen, Cambridge, and Laidlaw College and a PhD from Australian Catholic University. His research focuses on human and artificial intelligence.

Dr Jonathan Cole, Research Fellow. Jonathan is a researcher at Charles Sturt University, Canberra, who works on political theology: the intersection of politics and religion. He previously worked as an analyst in intelligence agencies: the Office of National Assessments and the Defence Signals Directorate.

Natasha Baulis, Researcher. Natasha comes from a background in anthropology, with six years of experience in the not-for-profit sector in East Africa and several years working in training for political engagement in Australia. She holds degrees in International Studies, Applied Anthropology and Participatory Development, and Anthropology, Art and Human Perception. She is currently completing a PhD at the University of Notre Dame, Australia.

The authors are grateful to Professor Aeron Davis of the Victoria University of Wellington and Andrew Gibson of Queensland University of Technology for their comments and suggestions on drafts of this paper. Any errors are the authors alone.

The paper in summary...

This is Maxim Institute's second paper on artificial intelligence (AI). Our first, *At the Cutting Edge: How Artificial Intelligence Will Change our Primary Sector Forever,* considered the impact of AI on our agricultural sector. This discussion paper considers the impact that AI will have, and is having, on our democracy.

The paper begins with an overview of the uses and benefits that AI brings to most facets of our society. We then turn to the various threats to our democratic processes that the use of AI poses. Some of the phenomena discussed, such as "nudging" and the use of misleading text and pictures, are not new in our political processes, but the advent of AI has increased their reach and speed. Other threats, such as the psychographic profiling and microtargeting of voters, have been substantially strengthened by AI and have become go-to tools for influencing elections, policy outcomes and even individual decision-makers.

We then turn to the external threats that are strengthened by the increasing sophistication of AI. The paper outlines the ways that foreign governments are already attempting to influence elections and public opinion around the world. It also explores the threat of cybercrime by non-state actors as well as the ability of multinational corporations to use AI to shape the public discourse and the regulatory/policy debate.

Finally, we turn to an internal threat to our democracy: algorithmic government. The temptation may come to turn over more of our governmental decision-making to AI to improve policy responses, productivity and efficiency. Although well-intentioned, there is a real risk that we slide from democratic accountability into digital authoritarianism with little to no voter input.

Having set out the opportunities and risks that AI poses our democracy and government, we finish the paper with three broad recommendations:

- **At the individual level.** We all need to be aware of the impact of AI and improve our critical thinking skills. We should all analyse our online sources and broaden our horizons when it comes to consuming media so that we can break free of our epistemic bubbles and limit the effectiveness of nudging, microtargeting and disinformation.
- **At the national government level.** The New Zealand Government should create a new business unit under the Department of Prime Minister of Cabinet: The AI Coordination Group. This Group would be responsible for strengthening protections around government-held data about New Zealanders, overseeing and guiding the use of AI by Government Ministries and leading the review of New Zealand's existing legislative and regulatory framework.
- **At the international level.** New Zealand should be part of the nascent international attempts to ensure that AI does not undermine our democratic systems. New Zealand should be arguing that democracy requires human accountability and that the normative centre of all public power is people, and not technology.

If these recommendations are followed then this will go a long way towards ameliorating the worst consequences of the widespread use of AI in our democracy and government. AI will feature heavily in our future campaign seasons. However, if we all know its effects and reach then we can protect our democratic systems from its undesirable effects.

A note on our research method: To explore the usefulness or otherwise of AI, we prepared this paper with limited and selective use of ChatGPT-3 and -4, Claude+, and Sage. We asked these programs such questions as "Explain psychographic profiling", "What destroys trust?", "What are epistemic bubbles and why are they a risk?", "What is the impact of disinformation on elections?", and "What are deep fakes?". The responses did not alter our literature review that had been conducted earlier and which forms the basis for the paper. Instead, the answers provided aided in the structuring of the argument and the points to be made. However, the responses required reordering, deletion of repetition, editing, and the inclusion of much additional material. The large language models were certainly not flawless. Overall, despite errors, gaps in detail, and ineffective referencing, the AI showed itself to be a promising aid to research and text generation. It may well soon become as word processors were to typewriters and handwriting: something useful which we hardly think about.

A note on our endnotes: There is a large and rapidly expanding body of research on the uses, risks and opportunities that come with AI. The most useful sources that we found in our literature review have been placed in the endnotes to this paper. We refrained from placing many quotations in the body of the text to make the paper more readable. However, interested readers will find a wealth of further material and quotations in the endnotes.

1. INTRODUCTION

1.1 AI among us

AI brings many benefits to society.[1] These benefits are remarkable and may come to affect nearly every area of human experience. Because of these benefits, AI is not going away. On the contrary, it will become more entwined in our lives, whether we are aware of its presence or not. Just as the developmental history of TVs, phones or cars has shown us, AI will increase in quality and capability over time. As humanity has experienced agricultural and industrial revolutions, so it is now caught up in the birth pangs of a new revolution—one which will define coming eras.[2] Where irrigation, soil science, seed, scientific method, metallurgy, coal, oil, and nuclear elements fuelled previous leaps in civilization, data will fuel this one. If data is not already our most precious commodity, it will become so.

AI is nearly ubiquitous. It will operate, perhaps autonomously if allowed, in most areas of human activity. Something new and powerful is emerging that will change our lives at every point. It might be thirty years before we see the scope and shape of this; we might not understand its significance for a further century. But we will probably come to recognise it as revolutionary in the grand sense of the word.[3]

1.2 A Growing awareness of AI

Programs like ChatGPT-4 and DALL-E have caught our imagination, raising awareness in generative machine learning. Now everyone is talking about AI.[4] Both platforms are amazing. It is quite easy to point to their failings, but that is to miss their significance.[5] Early platforms and technologies are precisely that: early. What comes later is more capable and helpful. Consider our use of Google, Netflix or Amazon compared to our experiences of sites like Yahoo in its early days. Text and image generation is in its infancy. What comes next will change the way we think about writing, and, for example, the power of Hollywood.

But it is important to realise that as impressive as these technologies are, even in their early stages, AI has been silently working across multiple sectors changing, and sometimes improving, our lives for the last decade or more by creating efficiencies. AI as machine intelligence and learning is enabling a quantum leap in productivity (as, for example, we documented in *At The Cutting Edge*, Maxim Institute's discussion paper on AI and agriculture). AI is creating massive labour efficiencies. We can do more, and more quickly than ever before. We have access to information instantly. Thanks to AI and the underlying technological infrastructure that enables it (for example, fast networking and graphics processing units), we can communicate through a range of media, sharing ultra-high-quality images with ease and speed.

Machine (or artificial) intelligences aid research in medicine, linguistics, engineering, social science, transport and communication, and astrophysics. They also expedite social connections. Further, working as recommenders, they not only predict who we will enjoy meeting but what we will enjoy eating, listening to, and watching. They profile us and make it easier for us to find what we want or to meet people with shared interests. For many this involves gaming communities or networks of discussion groups, book clubs, and so on. In all these domains AI is designed to make things easier, quicker, better, and more enjoyable. Furthermore, it aims to improve its own performance.

In terms of our government, democracy, and elective processes, AI works to similar ends. It looks to create efficiencies—to make things run smoothly with less wastage.[6] A detailed understanding of a political candidate's views and values can, for instance, be shared on social media helpfully. So too can a party's policies. At a government level AI's capacity to analyse data, offer predictions, and make recommendations arguably surpasses human ability. In health, transport, financial planning, defence—across all ministries and departments, AI acts as a powerful instrument. It has the possibility to enhance the workings of democracies and their governments.

1.3 AI is Everywhere

1.3.1 Business and Manufacturing

Not only is AI startling, it is also becoming ubiquitous. This is due to digital things (connected devices) and the *internet of things*[7]. The presence of machine intelligence and learning is an increasing reality in business and manufacturing. It not only assists employees but also in some instances replaces them. This is true of both blue- and white-collar workers. Robotics, drones, virtual assistants, voice recognition, large language models (LLMs, e.g., ChatGPT-4), AI logistics programs, augmented reality, and intelligent accounting are the types of technologies expediting this.

Whether in the production of planes, cars, phones, software, semiconductors, houses, clothes, or food, machine learning works with declining error, continuously, and for no salary. And when it is designed to do so, it is improving its performance.[8] In the next three years it is likely that AI will create notable efficiencies for businesses that adopt AI strategies as core to management. Machine intelligences may well oversee major utilities. They will continue to improve the automation of transport and delivery systems. We may soon see, for instance, pilots, uber and taxi drivers, and train operators having less agency in relation to the algorithms that underlie their vehicles, and perhaps even becoming redundant.[9]

Moreover, AI will generate and distribute entertainment—whether this takes the form of gaming, streaming movies, or music. It can create images, movies and sound from written commands. In so many areas, machine learning is changing business and manufacturing, and radically disrupting current employment arrangements.

1.3.2 Investment

Finance and investment have traditionally been tied to sages and big personality types; it has variously relied on associative memory, intuition, intersubjectivity, and "quants" (experts' statistical and mathematical methods that forecast financial markets' behaviour). But as Thomas Nagel observed, they have never achieved "the view from nowhere"—i.e., a godlike awareness of trends, information and events that affect the market, sometimes within microseconds.[10] Machine learning, however, seems to be approaching such a state, and it is increasingly being used for trading stocks, forex, banking, the development of financial instruments, and spotting fraud.[11]

Machine intelligence and learning systems analyse data and trends far faster than "experts". For instance, they are beginning to predict accurately a company's results and a market's response. They have superior immediacy of feedback and computational power compared to humans; and they can trade without fear, basing their decisions on the analysis of cold, hard data. They have also repeatedly and positively exploited counterintuition.[12] They are being shadow-used or actively employed by traders in the markets.[13]

Depending on regulation, AI will likely play an even greater role in finance, banking and investment in the next ten years. Humans may well have to fight not to be left out of the loop of decision making because they are too slow or simply unable to understand the rationale calculus for the nature of a transaction.

1.3.3 Family and Education

The unseen presence of machine learning is also thick in family life and education. As intimated earlier, children and families interact with it continuously, for instance, through posts, games, search engines and tailored information, Instagram, Facebook, YouTube and feedback loops used in formative assessment in classrooms (virtual or real). AI is already affecting teaching and learning. It is taking on a system that has long needed an overhaul.

In some countries, such as Japan, robotic intelligences support the elderly, comfort the lonely, provide dementia care, and more disturbingly engage in sex with humans.[14] A number of TV series and films already explore AI as "embodied", and how it might act and change our perspective on relationships.[15] Such stories pose the question of whether a person can form a strong, meaningful, and rewarding relationship with a non-human intelligence. It might sound ridiculous, but commentators argue that issues of personhood, emotions, and more profoundly, consciousness arise in this scenario.[16]

We will have to confront the degree to which we are comfortable with robotic intelligences, digital games, and interactive programmes being a part of our wider lives and in the raising of our children—especially if they are hackable. Family and education are increasingly entwined with AI.

1.3.4 Security and defence

The pivotal domains of defence and cyber security are also becoming more reliant on AI to protect against identity hacking, theft, phishing, stalking, privacy invasion (see further below in section 2.1).[17] The most crippling attacks in recent years on industry, hospitals, utilities, government departments, and banks have been on those which have not had advanced AI monitoring their security arrangements.[18] Machine intelligence scrutinises enormous amounts of data almost instantaneously to detect any irregularity. It has even developed the capacity to anticipate cyber-attacks based on changes in an environment. AI can be both poacher and gamekeeper: at the same time managing cyber security against AI-generated threats.[19]

On the kinetic front, machine intelligence is now capable of being deployed in multiple ways in the form of drones, attack vehicles, as well as smart and hypersonic missiles. If permitted to do so, it will operate autonomously. In "fire and forget" mode, drones with no human controller at a joystick carry out assassinations and targeted killings.[20]

1.3.5 Health

AI is also becoming widespread in health and medicine. It is used for monitoring health, anticipating, detecting, and diagnosing illness (blood analysis and evaluating CT scans, x-rays, and MRIs for malignant legions or tumours, etc.), decision support, and the discovery and development of medicines, vaccines, and cures.[21]

Some hold that a turning point is looming because biotechnological innovation driven by machine intelligence will bring about new forms of human augmentation. Genetic engineering will not only be used to target disease but also for human enhancement. Questions over how much and what type of augmentation is permissible and who benefits from such enhancements are already arising.[22]

The issue is made more complex by neural links or neural lacing and robotic, skeletal technology. The former holds out the sci-fi like possibility of uploading data to and from the brain. In consequence, some futurists, such as Elon Musk, believe humans might evolve into digital non-biological "persons". All joking apart, this is seen as an advantage as it suggests we would not easily be worn out by time, and therefore options like extended interplanetary travel and survival beyond an environmental Armageddon become possible.[23]

In sum, AI in the forms of robotic intelligence and machine learning is becoming ubiquitous. Keep this fact in mind when we turn to its effects on government, democracies and our elective processes. In business, finance, family, education, defence, cybersecurity, and health, AI is becoming a housemate, if not an essential tool. It is vital to grasp the fact that in each of these fields, competition and the need for economic survival will all but guarantee AI's continued uptake. To be competitive in product quality, marketing, distribution, and pricing, and to enjoy the benefits accruing from this, it will be necessary for individuals, businesses, and governments to embrace these technologies. Further, on the geopolitical front, the pressure to place AI at the centre of strategic thinking and action will inevitably intensify.

1.4 Is this a Nuclear Moment?

A number of legal, scientific, religious, and political leaders believe that advances in AI are bringing us to a nuclear moment, i.e., a time of profound discovery with its concomitant opportunities and risks.[24] This is for two reasons: first, we do not fully understand the technology we are developing but we have intimations of its power. Secondly, it is seemingly inevitable that we will give robotic and machine intelligences more authority or scope to manage our lives and political arrangements. We will, so to speak, put ourselves in the back seat. This is fine if AI remains within our ability to moderate and control. It is not so good otherwise.[25]

Relinquishing control of our lives carries enormous risk. But so does maintaining control.[26] The rise or AI appears inevitable and irresistible, which is why we need to align AI's values with our own and to place parameters around its power. Parameters will also be needed to safeguard our deeply cherished political processes, as we shall see.

2. HOW CAN AI UNDERMINE DEMOCRACY?

This section has three parts. Subsection 2.1 describes the methods and technologies that have been developed by tech companies and businesses for management, sales, and marketing etc., but which can be used politically either benignly or malignly. These include psychographic profiling, reverse engineering for identification, microtargeting and nudging. The subsection also includes commentary on misinformation and disinformation, and the realities and dangers of *epistemic bubbles* (knowledge bubbles). It also highlights risks to democratic processes, to voting and elections.

The second subsection (2.2) looks at threats to government and to democracies. It reflects on the external threats posed by foreign intelligence services and criminal networks. Additionally, it touches on the power of multinational companies. This subsection also examines the question of algorithmic government arising from either apathy or good intention.

Finally, subsection 2.3 addresses the issue of trust and how it is undermined by the malicious use of AI. The subsection points to the importance of trust in government (and wider society, business, education, etc.), and how its erosion damages the New Zealand body politic.

2.1 The threat to elections

This section describes methods and technologies which have been developed over the past fifty years, and which have been increasingly used during the last two decades for political purposes.

2.1.1 Psychographic profiling and predictive behaviour

Psychographic profiling and *microtargeting* are not new. We have always wanted to know who we are interacting with and what their interests are. We want to relate to them. Psychographic profiling generates the profile of a group of people based on their psychological and behavioural characteristics. It outlines their interests, lifestyles, personality traits, values, attitudes and opinions. Typically, psychographic profiling has been based on data collected through expensive, time-consuming surveys and focus groups. It is incorrect to suggest that in the context of AI psychographic profiling has agency. It is algorithmic and a product of machine intelligence, but the use of machine learning has brought astonishing levels of sophistication to psychographic analysis. It has augmented social interaction and business practices. It has also dramatically enabled the ambivalent practice of microtargeting (explained further below). More recently it has incorporated *scraped* data from our browsing records, searches, preferences (website, articles, and videos), and gaming habits. Machine intelligences examine, analyse, and record the shape and impress of our digital footprint.

Psychographic profiling also uses data from social media, studying the types of content, topics, and brands that draw our eyes and touch our wallets. It studies our spending habits, shopping preferences, and brand loyalties. Moreover, it uses data gleaned from our frequent flyer and rewards cards, gaining an insight into our personality that possibly surpasses our own.[27] It consumes information from a host of apps, devices, and platforms, which, we use daily without pausing for thought. In some iterations, depending on the jurisdiction or the type of government (autocratic, authoritarian, or democratic), psychographic profiling makes use of our health, education, employment, credit, and tax history.[28]

In terms of personality assessments, psychographic profiling also determines our personality type (Myers-Briggs, Big Five, Typefinder, DISC). It predicts behaviours such as introversion or extroversion, openness to experience, or conscientiousness. This enables marketers to finesse their messaging so that it appeals to specific personality types.[29] Micrographic profiling also targets our tikanga, values and attitudes because they too affect our decision-making. These might relate to the value of life, sustainability, immigration, religion, family, tradition, or education. Marketing campaigns which align with our values are more likely to be successful. Recent search suggests personality is dynamic. Character is to a degree fluid and values change. The traits, however, that psychographic profiling identifies, even if they prove to be seasonal, remain useful to marketers.

The same level of scrutiny washes over our interests. Our love of sport, cycle trails, comedy, the beach, the bar down the road, or of real-estate are data points for leverage. They help predict our behaviour and enable the generation of content that appeals to specific consumer segments. The same goes for lifestyle—the patterns of behaviour and decision-making that define how we live. Profiling takes

account of factors such as our work-life balance, health, exercise, wellness and travel choices. By understanding the lifestyles of their target audience, marketers create campaigns that fit our lives.[30]

It is important to grasp that all the information mentioned above is analysed nowadays by machine intelligences. This activity is expedited by the *Internet of Things* (IoT)—"an ecosystem of sensors in our homes, offices, vehicles, on our bodies, and in public places that collects raw data on us."[31] Indeed, Cisco estimates there are currently one trillion inter-connected IoT devices and that by 2040 there will be 45 trillion "things" gathering our data.[32] The bulk of the data relates to industry—to tracking products, and monitoring equipment and environments, etc. But we should not forget that on the human/social front information collected via phones, wearables, and home automation, is already being parsed by machine intelligences which find patterns and trends, predict behaviour, and create three-dimensional profiles of target groups for marketing. These profiles move far beyond traditional demographic segmentation, which focused on simple factors like age, gender, income, and education. They lead to highly targeted and effective marketing campaigns.[33]

In sum, the purpose of psychographic profiling is to gain a deeper understanding of the needs and motivations of a particular group of people, and to anticipate how they will behave in certain conditions. This enables marketers to tailor products, services, and messaging to better meet their needs. Over two decades, psychographic targeting has become increasingly sophisticated. More recently, narrow algorithmic machine intelligences using statistical modelling have advanced the practice further. Data scraping and data aggregation from a broad sweep of devices, platforms, services, and applications now provide panoptic insight into our lives and peccadillos, and divide us into hyper-specific social segments. These may consist of a few thousand or a few dozen people with a high degree of similarity. They are gold to marketing.[34] They result in higher conversion rates and sales while granting better customer satisfaction. More importantly for our purposes, and in relation to elections and government, psychographic profiling is becoming invaluable to advisors running political campaigns.[35] As we shall see in following sections, it enables them, for example, to target swing voters with precise, tailored messaging.

2.1.2 Reverse engineering for identity

Recent news articles have reported that machine intelligences have cracked ancient languages. Taking fragments of cuneiform text from thousands of tiny shards of clay, AI has reverse engineered millennia-old Sumerian, Akkadian and other etchings to decode the languages. A feat previously deemed impossible by solely human means.[36]

Using similar skills, a more disturbing phenomenon sees machine intelligence taking tiny particles of information from multiple platforms and isolated sources and re-assemble them into a single personal profile. AI is reverse engineering specks of data and bringing to life our digital identities. Using trace data and algorithmic behavioural prediction machine intelligences are creating an image of us. It is a simulacrum—it does not reflect our human complexity. It is, however, a development that is especially disturbing for law makers and policy writers who are trying to protect our privacy.[37] Previous legislation guarded us from powerful corporations on-selling data but did not anticipate the type of psychographic profiling that has emerged through this form of technology. It is difficult to pin liability on any one business that might store limited information on us, especially when it is available from other sources and is only useful when combined with many others.

2.1.3 Microtargeting

Psychographic profiling just discussed quickly tips into *microtargeting*. And like psychographic profiling, microtargeting has its origins in business. Banks and financial services, pharmaceuticals, retail and e-commerce, initially drove its development. More recently, subscription services, and social media and advertising have given it fresh impetus.[38] The latter has become especially important to elective processes. We have already noted that machine intelligence is operating across every sector. It is ubiquitous. But it is also core to contemporary microtargeting operations.

Thus, banks and credit card companies use microtargeting to offer custom-made financial products and services. They target individuals who will be interested in new savings accounts with specific features or a credit card with certain rewards. They also pay minute attention to our credit worthiness and financial risk profile.[39] Pharmaceutical companies and healthcare providers

also use microtargeting, pinpointing patients with specific conditions, treatment histories, or demographic profiles. They target ads for weight loss, diabetic treatment, or arthritic relief, etc., using data from people who have researched medication online or have a history of purchasing related products.[40]

Likewise online retailers have adopted microtargeting as a core strategy. They, too, provide personalised product recommendations based on our browsing and purchase history. Texts also spruik targeted sales on products chronicled in rewards and loyalty programs. And each particle of sales information is collated and rendered by a machine intelligence into something that might appeal to us and make us finger our credit card.[41] An action which in turn feeds the growing and general capabilities of AI.

It is the same story with streaming platforms like Netflix, Amazon Prime, and Spotify. Their success largely depends on the hunger for new content generated by microtargeting. Material is matched with preferences reflected in our viewing and listening history. This keeps us engaged with the platform and ensures our continuing subscription.[42] More generally, businesses are developing AI strategies that include microtargeting. Using machine intelligence technologies they leverage social media platforms like Facebook, WhatsApp, TikTok, and Instagram, and deliver targeted ads to specific audiences based on interests, behaviours, and demographic data, etc.

We need to make two observations here. First, the technology and practices around microtargeting have matured dramatically in the last five years. They are deeper, more powerful, and more comprehensive. Secondly, the availability and aggregation of data provides politicians, election campaign managers, domestic lobbyists, and foreign powers with an unprecedented ability to interfere in our elections.[43]

Microtargeting is becoming a go to tool for influencing elections, policy decisions, and even justices in High Courts, because it works.[44] Microtargeting precisely matches the psychographic profile of the people it takes aim at with its messaging. It presents them with party policies or the character traits of a leader that they value. Its messaging is on point. It is relevant and persuasive. More importantly, people receiving it see it as time-saving and useful.[45] Machine learning also continuously analyses how people engage and respond to different messages. It determines what is currently most effective for each targeted segment. Its models are unremittingly optimized for better results. It swaps out text, images, and videos to be more persuasive and to garner a greater number of followers. Customised content and political messaging resonate and find a home with audiences, resulting in higher click-through, conversion, and loyalty rates.[46]

Thus, microtargeting gives political parties and campaign managers a substantial competitive advantage. Big data, deep data insights, and precision targeting enable them to secure and retain voter segments.[47] Further, it opens new opportunities for them. It uncovers niche audiences, or hidden voter segments, that can form partnerships to defeat an incumbent or shore up the *status quo*. AI experts warn that "the use of Big Data and AI in digital media are often incongruent with fundamental democratic principles and human rights. The dominant paradigm is one of covert exploitation, erosion of individual agency and autonomy, and a sheer lack of transparency and accountability, reminiscent of authoritarian dynamics rather than of a digital well-being with equal and active participation of informed citizens."[48] It is worth highlighting the scalability of this. Although microtargeting is highly customised, machine automation (mailouts, bots, etc.) allows it to be executed at massive scale.[49] It can instantly reach millions of individuals with ringing endorsements of political leaders, impact statements on a proposed tax, or stats on emigration.

In sum, microtargeting is powerful and effective. It is addictive and political campaigners are becoming users. It reaches specific voter segments based on their political beliefs, values, interests, and demographics. Campaigners use microtargeting to send personalised messages to voters who care, for example, about environmental and sustainability issues, highlighting their candidate's green policies and achievements. Such information can be helpful to voters. But questions remain as to the integrity of information sent out, its comprehensiveness, and what it does not acknowledge in relation to other policy stances.[50] It can be misleading and trigger voters to make decisions and act on poor information.

2.1.4 Nudging

As we have seen, microtargeting uses psychographic profiling. It takes information, "frames" it, and targets us in order to "nudge" our decision-making on a particular issue. In the case of elections, it subtly influences our political choices.[51]

There has always been nudging in some form. In the early twentieth century blunt forms of nudging might have involved favouring candidates by listing them first in ballots. Or perhaps it may have involved by detailing their virtues in slightly larger information boxes in pamphlets, etc. Later it may have involved distributing news feeds that anchored public perception, releasing a poll with a candidate as a frontrunner, thus nudging voters to get onboard with the natural winner.

Nor are framing effects new. They work in tandem with predictive analytics. They present information or policy initiatives in a selective, emotionally resonant way, advancing specific candidates or policies which will find favour with particular voters. We respond positively to these frames of information because they play to our concerns. But they tend to place no expectation on us to evaluate the rest of the policies a party will introduce. Framing nudges us towards candidates by portraying them in the most compelling light.

For example, in the last Canadian federal election (2019), predictive analytics identified swing voters and regions where electoral races were tight. By analysing previous voting patterns, demographic data, and survey responses, campaigners were able to tightly frame the messaging in key battlegrounds, such as the Greater Toronto Area, where only a handful of close races determined the outcome of the election. The targeted approach enabled parties to carefully allocate their resources and nudge undecided voters towards candidates in crucial electoral districts.[52]

Campaigners also use machine intelligences to bombard us. The frequent flashing of a political candidate's name, slogan, or message nudges us down a desired path of action.[53] It does so through familiarity. As we generally prefer things we have been exposed to before, we become more inclined to vote in a certain way. To use repetition in microtargeting is to deploy a basic but powerful nudge.[54]

To this end, bots and chatbots have flooded the internet. They nudge us with titbits and gobbets of information. They can encourage us to change our minds to great effect.[55] In the 2017 French presidential election, Emmanuel Macron's campaign used A/B testing (which variant do you prefer A or B?) to optimise its digital marketing efforts. By experimenting with bots using different email subject lines, ad creatives, and social media content, the campaign hit the most effective messaging strategies for engaging supporters and then motivating them to vote. The data-driven approach allowed Macron's campaign to adapt its tactics in real-time and maximise the impact of nudges. These arguably contributed to his election victory.[56]

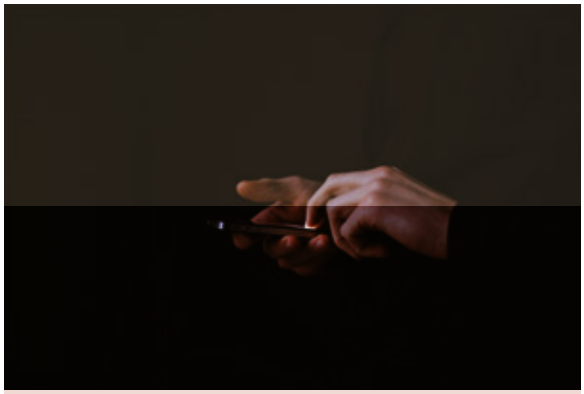
Another example can be seen in the 2019 Indian general election, when the Bharatiya Janata Party (BJP) and the Indian National Congress, used chatbots extensively on WhatsApp and Facebook Messenger to secure voters. They flooded users with information on party policy, candidates, and voting logistics (when, where, and how to vote). By offering "personalised" and importantly, interactive experiences, the chatbots raised awareness among voters on key issues and candidates who would address them. They nudged citizens to vote in specific party directions. It is worth noting that bots are now advanced enough to "identify keywords in public posts or conversations and then populate the content with their own posts (if they act as spambots) or conversations (if they act as chat bots)."[57]

Personalisation is a slightly different form of nudging to framing. It is more sophisticated. Using large data sets and machine learning it microtargets voters with messages on various issues like crime, immigration, and the cost of living, etc. Through extensive data scraping and profiling, it raises issues of interest to us. It then intentionally nudges us towards a given candidate whose values and priorities appear to align with ours.[58] Working in conjunction with "sentiment analysis" which captures degrees of positivity or negativity towards something, this type of personalisation can be helpful. It can act as a positive filter. But it can also be construed as deception and manipulation. And it can leave us feeling used (we are being targeted and manipulated) and erode trust (see subsection 2.3 below).[59] For the last decade, political campaigns have leveraged huge amounts of data to personalise nudging at scale. And they have been very effective.[60]

In the 2012 United States presidential election, Barack Obama's campaign used data analytics to tabulate voters based on their gender, age, ethnicity, and issue preferences. It was the first step to targeted personalised messaging. Young voters were beset with messages relating to student debt and affordable healthcare, while older voters received messages about social security and Medicare. The personalised approach saw Mr. Obama's campaign connect with discrete segments of voters and is widely believed to have contributed to his re-election victory.[61]

In a similar manner, but with more sophistication, the UK Labour Party used sentiment analysis in the 2017 UK general

election to gauge people's feelings through social media. By analysing the emotions and opinions expressed in tweets, Instagram and Facebook posts, the party identified key issues that resonated with members of the British public, such as healthcare, education, and housing. This enabled the Labour Party to focus its messaging on issues that bands of voters cared about most, helping it gain seats in the election and contributing to a hung parliament.[62]



New Zealand political parties, candidates and lobby groups use Facebook and Google to advertise their campaigns. Both the Labour Party and the National Party use detailed targeting for their advertising, spending $132,802 and $75,172 respectively on their Facebook campaigns for the ninety days between 22 February - 22 May 2023.[63]

Both Google and Facebook require New Zealand political advertisers to verify their identity and location and to publicly disclose who is publishing and paying for content. Google only permits political advertisers in New Zealand to target advertising based on context—the types of webpages where an advertisement can be shown. However, in response to rising concerns globally regarding the use of micro-data and political nudging, as of 2022 Facebook no longer permits political advertisers to target demographics based on health, race or ethnicity, political affiliation, religion, or sexual orientation.[64] *Details of targeting by political advertisers are disclosed to the public via Facebook's 'Ad Library'. Facebook permits political advertisers in New Zealand to target their advertisements at specific groups based on location, age, gender, demographics (e.g., university graduates or parents), behaviours (e.g., frequent overseas travellers or commuters) and interests (e.g., electronic vehicles or classical music).*

Campaign managers also use social influencers and social media to amplify personalisation. We tend to be nudged by the views of our peers and social groups. So, reaching one person through personalisation can have the effect of reaching many.[65] Thus, campaigns often try to leverage social influence by sharing the endorsements of celebrities, sports stars, leading musicians and influential community leaders. Close friends and those we adulate are voting for so-and-so, and peer pressure pushes us to follow suit. In this way, social pressure nudges more and more voters down the same road. America has mastered this canvassing technique, but in recent elections in New Zealand, Australia, Canada, and the UK, political leaders have all employed the same technique, holding up the hands of celebrities and posting images that spread quickly across social media.[66] The strategy is used, too, for the introduction of cornerstone legislation and policy.

In summary, nudging works by using AI to tap into people's psychological predispositions, identify the thought processes they use when making complex choices, and predict how they will think and act in relation to certain information. It targets behaviour, but reaches into our thoughts and emotions. It deploys the techniques of favouring, framing, repetition, personalisation, and social pressure. Campaign managers aim to nudge voters towards their candidates in preference to others. But nudging runs the risk of becoming covert manipulation. While nudging is an age-old phenomenon, and not an evil *per se*, the technology driving it today allows for a subtlety, scale and speed of application that carries new dangers.

2.1.5 Disinformation and deepfakes

Disinformation is the deliberate creation and dissemination of false or misleading information.[67] Disinformation should be distinguished from misinformation. Both refer to the spread of false or inaccurate information. But their intent differs. Disinformation intentionally disseminates lies to deceive, manipulate, or harm people and institutions. Misinformation is unintentional and is without malice.[68]

Because disinformation is what we used to know of as *lies*, it is a phenomenon as old as time, but its reach is so much longer today: "AI tools available on the internet actively promote rumour cascades and other 'information disorders'."[69]

Disinformation campaigns during electoral processes are multiform and nasty. They can be text-driven, video, image-based or auditory. They are multimodal, often using a combination of all of these: e.g., a video accompanied by what looks like a news article, to make them appear more realistic.[70] Benzie and Montasari lament that "offsetting the benefits of a more connected world are a string of worries as far-reaching platforms have become breeding grounds for mis- and disinformation disseminated by humans and machines often with malicious underlying intentions."[71]

Bots further the deception through news sites and on social media platforms, creating the illusion of widespread support or hostility towards a particular candidate or issue. The manipulation distorts public perception and attempts to sway voter preference. Indeed, "European and global democracies are under a severe threat due to extensive spread of disinformation through social and traditional media. And the use of automated accounts and bots, psychographic microtargeting, and deepfakes to proliferate fake news during elections are making the problem even more alarming."[72]

Fake news typifies text-driven manipulation. It is a serious threat, but ironically, "as the likely quantity of fake news will become impossible for humans to identify and analyse rapidly enough, AI technology will certainly become key in detecting and identifying it." In the meantime, machine intelligences working as LLMs which generate text in multiple genres, will increase the quantity and probable quality of fake news. Its cousin, *manipulated content*, works by editing and metamorphosing images, videos, or audio recordings into something new—*Deepfakes*, which are a form of impersonation.

These deepfakes are a form of synthetic media generated by deep learning algorithms of *Generative Adversarial Networks—GANs*. These consist of two neural networks, a generator, and a discriminator, which work to produce lifelike synthetic media. The generator creates draft content, while the discriminator evaluates its authenticity. Through an iterative process, the generator expands its ability to produce increasingly realistic content, while the discriminator becomes better at finding fakes. Over time, the generator creates credible deepfakes difficult to distinguish from the person they are based on. Because of their power to deceive, deepfakes pose a serious problem to elective processes and democratic government, without even considering the serious issues of privacy and consent. The quality of resemblance to the deepfake's human counterparts will continue to increase. Seemingly perfect videos can now be created by written or oral command using AI video generators such as Synthesia.

This is not to say that deepfakes are wholly without legitimate use. They simultaneously draw on and advance the positives and pleasures of entertainment—of movies and gaming, etc. And it might be argued that deepfakes have a legitimate role in advertising. Certainly, and especially in relation to virtual reality in the classroom, education in the future might be permeated by deepfakes—with students travelling to a simulated Rift Valley, ancient wonder, or Beijing. Deepfakes are already being used in history classrooms to generate so called 'deep nostalgia', bringing to life the images of historical figures using old photographs and AI video generators.[73]

But the trouble is that synthetic media can be used perversely. And they will be in the future, too. In the political arena, Maja Brkhan describes two categories of deepfake: "videos aimed to harm political opponents and those seeking to enhance the candidates' political popularity. The first category includes videos depicting politicians involved in corruption or another controversial or criminal activity and uttering statements with inappropriate or offensive content. The second category could include fake videos of politicians attending high-level international meetings they never attended, shaking hands with prominent world leaders or offering support to vulnerable societal groups, such as homeless, sick or otherwise affected."[74] Deepfake content can have serious ramifications for political campaigns, particularly when it is released in the lead up to an election with insufficient time remaining to prove its falsity. In May of 2023 Turkish presidential candidate Muharrem İnce withdrew from the ballot less than a week from the election after the release of a porn video he claims is a deepfake.[75]

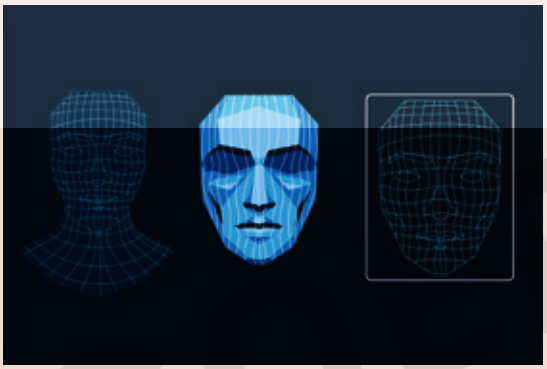
The power for destruction that these machine intelligences represent is particularly acute in relation to polarization and division. Deepfakes spread disinformation, which easily inflames existing social and political tensions. By targeting specific demographics with deepfakes, disinformation campaigns magnify divisive issues, heighten polarization, and freeze political discourse. They crowd out and frustrate the free-flow and transmission of accurate information. And they skew elections. Even social media platforms find it difficult to address the issue of deepfake crimes. In many cases of deepfake fraud in New Zealand, the hacked accounts were still not deactivated even weeks after they were reported.[76]

Deepfakes, and disinformation in general, have a caustic effect on public trust. In its 2018 report the Computational Propaganda Research Project "found evidence of formally organized social media manipulation campaigns in 48 countries, up from 28 countries [the previous year] .... Much of this growth comes from countries where political parties are spreading disinformation during elections, or countries where government agencies feel threatened by junk news and foreign interference and are responding by developing their own computational propaganda campaigns in response."[77] Disinformation campaigns sow doubts as to the legitimacy of electoral processes. We shall discuss the issue of trust further in subsection 2.3 below.

Several deepfake parodies have been produced of Government Minsters using AI generative technologies. In November 2021 a deepfake video of Jacinda Ardern smoking marijuana was posted to Facebook. Despite the poor quality of the video, comments made by Facebook users indicate that these images were mistaken for real footage by some members of the public.[79] The creator of the video, owner of the YouTube channel, 'Genuine Fake', has produced more than 85 videos of prominent New Zealanders. The most popular of these—featuring Jacinda Ardern as Buttercup from 'The Princess Bride'—has received more than 65,000 views. The marijuana video has been viewed more than 17,000 times to date.

Deepfakes technologies are also increasingly being used by political actors for more benign purposes, such as the production of images by the New Zealand National Party for campaign posters in the lead up to the 2023 election.[80]

Worryingly, deepfakes are increasingly being used to target New Zealanders in a range of serious cybercrimes. These scam videos use AI generated videos of prominent New Zealanders to trick people out of their money. In early 2023, a deepfake video of New Zealand influencer and former socialite Aja Rock was released via a hacked Instagram account deceiving several friends and followers into signing up for a financial investment scam. Other celebrities whose visages have been used in this manner include Sir Sam Neill, sporting stars Richie McCaw and Sonny Bill Williams and Newstalk ZB broadcaster Kate Hawkesby. The AI generative technologies used in these scams were far more sophisticated than those used by 'Genuine Fake', replicating both the images and the voices of these individuals convincingly.[81] The use of trusted public figures to successfully hoax the public threatens to further undermine public trust.



Although the production of deepfakes itself is not illegal in New Zealand, some instances of Deepfake production may be deemed 'harmful material' under the Harmful Digital Communications Act 2015 or considered a breach of other existing legislative frameworks, e.g., the Privacy Act 1993, the Broadcasting Act 1989 or the Crimes Act 1961. However, due to the nature of cybercrime itself—where a crime can be committed virtually anywhere in the world from virtually anywhere in the world—the abuse of deepfake technology is difficult to prosecute and many people do not bother to report a crime when it has occurred online. In a recent Ministry of Justice Survey, it was found that "Cybercrime offences were the least likely to be reported, with 98% of incidents not reported to the Police."[82]

This is further exacerbated by the complex and overlapping network of both government and charitable organisations responsible for responding to cybercrimes in New Zealand.[83] A simple Google search, "Where to report cybercrime in New Zealand", brings up the following list of organisations: CERTNZ (Computer Emergency Response Team), IDCARE (a New Zealand/ Australian charity that also receives referrals from government organisations), Department of Internal Affairs, Netsafe, New Zealand Police, National Cyber Security Centre and Serious Fraud Office.

In sum, disinformation employs deception. It can have far-reaching consequences for democracies. It works by leveraging social media platforms to promote or discredit political candidates, parties and policies. It manipulates public opinion and consequently affects voters decision-making and behaviour. Fake news and deepfakes are core to its success. Additionally, and dangerously, disinformation accentuates polarisation and social division, polluting the atmosphere of toleration that characterises democracies. Disinformation undermines trust in electoral processes, weakens democratic life, and strengthens the fist of authoritarian leadership. Consulting firm Gartner makes the sobering prediction that "the majority of individuals in mature economies will consume more false information than true information," and notes that "Brookings calls this 'the democratization of disinformation.'"[77][78]

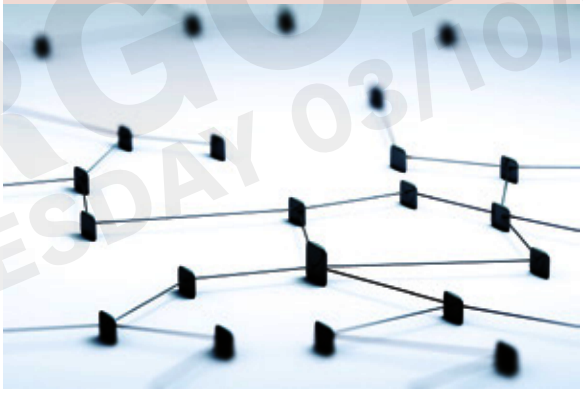
2.1.6 Epistemic bubbles

The science of psychographic profiling lends itself to microtargeting, which in turn nudges voters down certain political paths. In a similar way, disinformation and deepfakes make nigh impenetrable the layer which surrounds *epistemic bubbles* (knowledge bubbles).

Epistemic bubbles are communities of people who consume and share information which bolsters their presuppositions. To an extent we are all belong to one or more such bubbles. It is a natural consequence of following up on things that interest us. The danger comes when we block off exposure to counterfactuals or from different ways of evaluating a set of propositions, "facts", events, and conclusions. We become smugly isolated in our thinking or fail to come to grips with contradiction and complexity—sometimes simply because it is inconvenient to do so. It is not only dis- and misinformation that makes epistemic bubbles so hazardous, but also limited information and the wrong conclusions drawn from it.[84] Microtargeting, recommenders, and a plethora of serviceable media that are ready to buttress our "insights" only add to these perils.[85] And such bubbles can be profoundly anti-democratic.[86] They can silence debate and hamper personal and national development.

How are these dangers to be understood in electoral and political terms? First, epistemic bubbles tend to embrace a narrow perspective. This is not always a bad thing. Research, for instance, can have a very narrow focus. And specialisation, by its nature, limits the amount of reference material that can be resourced. But valid research remains open to falsifiability or a theory with greater explanatory and predictive power than that offered by a current model.

According to an AUT report,[91] trust in traditional media continues to decline in NZ: down from 53% in 2020 to 42% in 2023. Kiwis are increasingly turning to social media platforms for their news content, with Facebook ranking as the third largest source of news for New Zealanders (behind TVNZ and Stuff). Facebook uses an AI algorithm to rank posts (including news) on each user's platform. The algorithm has been developed over several years. Initially it performed fairly simple calculations to rank updates from friends so that items of most interest (e.g. changes in relationship status) appeared first. Later developments to the algorithm prioritised time spent on the platform, giving rise to the 'clickbait' article. Now the focus is to encourage interaction. One side-effect of this is the prioritisation of divisive content and news that aligns with the user's interests, thus driving the growth of the epistemic bubble.[92]



The risk with epistemic bubbles lies in the sustaining of false beliefs as narrow interests morph into closed, untested and biased ones.[87] At this point, it becomes easy to drift into bad, even conspiratorial ways of thinking. And machine intelligence that is malignly used compounds this phenomenon. Key questions, here, are whether epistemic bubbles, fortified by AI, advance or reflect the particular pathologies of their communities; and the relation between agency and technological determinism.[88]

Further, while the quality and objectivity of the free press (*The Fourth Estate*) can often be questionable, the growth of epistemic bubbles virtually guarantees its growing

irrelevance and likely demise.[89] And in a vicious circle its shrinking confines us to more bubbles. This loss marks a potential inability to inform the public in a broadly agreed upon way. And it is not surprising that public protests have recently become more regular, more strident, and seemingly less tolerant.[90] For a deliberative democracy to work, people need open access to accurate, shared information. The entrenchment of knowledge bubbles works to undermine this.

Epistemic bubbles might also endanger free and fair elections. When voters are targeted and nudged in knowledge bubbles, their political choices can be influenced by misleading or false information. Our votes run the risk of aligning with the information we have been fed or, more disturbingly, the political hand that feeds it to us.[93] And our ability to make informed decisions or freely choose between candidates or policies diminishes, while, for all the wrong reasons, our national electorate becomes more polarised.[94]

The broader significance of these effects is the degradation of social cohesion. A moment's reflection also suggests they dent friendships and diminish empathy. By definition, knowledge bubbles operate with differing premises and sets of "facts". As they solidify around a single narrative, it becomes harder for us to sustain a common horizon of national values and a shared understanding of the world.

In sum, epistemic bubbles pose a risk to the process of reasoning, the pursuit of truth, respectful disagreement, social cohesion, and the ability to have our actions shaped by compassion.[95] Epistemic bubbles erode trust in political candidates, the integrity of electoral commissions, traditional media, and judiciaries. And the attrition of trust leads to further civic disengagement.

All this translates into a serious threat to democracies. An electorate that is caught up in echo chambers will be plagued by poor leaders, bad policies, and tend to the decay of the underlying social cohesion which is necessary for democratic norms and values. Such an electorate will either wilfully or accidentally advance autocratic and despotic government. If we value democracy, we must address the narcissism of knowledge bubbles; their inflation and try to deflate them. The following section explains in more detail how profiling, microtargeting, nudging, GANS, and epistemic bubbles translate into dangers to democratic government both external and domestic.

2.2 The threat to government

Spying is an ancient practice stemming from the anarchic nature of the international order. No state can fully trust the intentions, aims and ambitions of friendly or neutral states, let alone hostile or pariah states. Thus, the need for intelligence in the contemporary world is pervasive and pressing. Governments, with a duty to protect their citizens and their interests, invest significant resources in detecting and thwarting threats from state and non-state actors. Corporations, particularly multinationals or those trading across borders, seek intelligence on governments and competitors to identify commercial opportunities, comparative advantage or to manage regulators. Even criminal syndicates and terrorist groups seek intelligence to identify targets and to evade law enforcement.

AI is set to transform the world of intelligence gathering and usage and will be adopted and employed widely by governments, corporations and transnational actors, not just criminal but also potentially international organisations, civil society organisations and the not-for profit sector.

2.2.1 The External Threat

The external threats to democracies mainly arise from foreign intelligence and multinationals. Transnational criminal organisations and terror groups also pose a risk, but in a parasitic and usually slightly less sophisticated manner. These external actors will be able to use the technology already described above to influence our politics for their own ends. Let us now look at how this will occur.

Foreign intelligence

Western democracies like New Zealand have long been the targets of espionage from hostile, or unfriendly, states. It is common knowledge that China and Russia pose consistent and serious espionage threats, routinely targeting government institutions, corporations and individual citizens, although other states also engage in espionage against Western democracies that is less well-known and well-broadcast.[96]

Advances in AI technology will enhance traditional espionage tradecraft through, for instance, psychographic profiling to identify potential agents with access to desired information and vulnerabilities capable of exploitation. However, it is the potential of AI to shape public opinion and the outcomes of elections that poses an emerging and far more serious threat.

Hostile states have long sought to employ their intelligence capabilities to influence and shape the public opinion, policy formation, legislation and diplomacy of democratic governments and societies deemed to be adversaries or impediments to the hostile state's national interest. The traditional targets of this type of espionage have been journalists, politicians, bureaucrats, lobbyists, diplomats and civil society organisations. However, with the digitisation of democratic governments and societies, with government-citizen interactions increasingly conducted online, along with political discourse and debate, particularly via social media platforms, there are new and emerging opportunities to manipulate both public opinion and policy making (including using the former to shape outcomes in the latter).[97]

We have already witnessed a concerted effort by one state actor, namely Russia, to interfere with the 2016 US election via Twitter, exploiting its anonymity and, at the time, relatively cheap and easy means of creating bot accounts or accounts managed by Russian actors impersonating American citizens.[98] It is difficult to discern the motive in this operation, given accounts offering both left-wing and right-wing perspectives were created and employed by Russian engineers. It may simply have been an effort to exacerbate political polarisation.[99] In any event, the operation should be seen as experimental, the testing of a new and nascent influence operation technique made possible by the advent of machine intelligence.

As LLMs become ever more capable and effective, further opportunities for foreign interference emerges. Hostile intelligence services, with the resources of state behind them, will turn bots into convincing autonomous political influencers. In fact, the capability of AI social media accounts to post human-like content, but also to do so in a more elegant and convincing way, will manipulate public opinion and debate on contentious political, moral and social issues, in ways that will be difficult, perhaps impossible, to detect.[100] Moreover, hostile foreign actors can now employ such AI accounts at scale cheaply, quickly and remotely. These accounts will be able to flood the national electorate with thousands of autonomous individualised voices indistinguishable from the average citizen, except, ironically, insofar as they might prove more influential than the average voice, as noted above.[101]

Furthermore, agents might be able to use the technology in very targeted ways, for instance, with a messaging campaign targeting an elected representative in a single electorate aimed at creating the impression of a constituency backlash against a policy proposal completely generated and implemented by AI, including intelligent and automated phone calls to electoral offices.[102]

Products of GANs are another rapidly developing tool that external powers will use to shape public opinion, policy formation and electoral outcomes. As discussed above, deep fake technology, at the moment incapable of completely eluding detection, is becoming advanced enough to convince an undiscerning observer.

Bad state actors targeting New Zealand will not only generate deep fake images (and increasingly footage and voice) but will also exploit AI's virality. They have analysed the mechanisms and environments that facilitate the phenomenon of viral social media and can drop strategically timed deep fake images days before an election.[103] While the bogus nature of the image might be detected relatively quickly, if the image or footage attains virality, it will achieve its intended effect in spite of discovery.[104] For example, this type of weaponization of images could involve a prime ministerial candidate in a compromising position, doing something illegal, reneging on a promise or changing a key policy at the heart of the election battle.

Cybercrime

Governments will not only have to contend with this new threat to the integrity of democratic institutions, processes and society from state actors, but also from corporate actors and criminals. They all have agendas and interests in the outcome of specific policy decisions and/or electoral outcomes. And the same tools will be available to anyone with the resources and motivation to affect decision-making and outcomes in democratic processes. Moreover, these tools can be used from virtually anywhere on earth. In the realm of AI-enhanced cyber espionage, digital manipulation and interference in the functioning of democracy is borderless.

Cybercrime predates the arrival of recent advances in AI technology, but is set to become far more sophisticated and costly thanks to machine intelligence. Governments in Western democracies are under increasing attack, and increasingly successful attack, by transnational cyber criminals. The current *modus operandi* of cyber criminals is to steal sensitive customer data held by businesses and then extort the business for millions of dollars. New Zealand and Australia have been the subject of a recent

spate of successful cyber-attacks, unprecedented in their damage. One egregious example was the hack of customer data held by Australian private health insurer Medibank Private in October last year (2022) by hackers in Russia. The criminals stole information relating to four million customers for extortion. When Medibank refused to pay up, the cyber criminals systematically released the customer information doing untold damage to the reputation and finances of Australia's largest health insurer and repository of some of the most sensitive data pertaining to citizens.[105]

In December 2022 a ransomware attack on managed service provider, Mercury IT, in New Zealand impacted several regulatory authorities including the Ministry of Justice and Te Whatu Ora. Tens of thousands of records were impacted or made inaccessible, including post-mortem examinations, cardiac and inherited disease registries and files relating to the transportation of bodies.[106]

When reporting to the Security and Intelligence Committee in March of 2023, Andrew Hampton, Director General of the New Zealand Government Communications Security Bureau (GCSB), testified that state-sponsored cybersecurity actors account for over 34% of recorded instances in New Zealand. In the same hearing, Phil McKee, Acting Director-General of Security, reported growing concerns that data was being collected on individuals speaking out against foreign regimes and used to target family members in their home countries.[107]

It is not just hacking, though. AI also enables cyber criminals to employ GANS, other algorithms or AI-enhanced techniques to target government agencies (and private corporations) at scale in cost-effective ways.

Multinationals (corporate leviathans)

Multinationals, like foreign governments, have the resources and incentives to intrude on democratic processes. They openly compete for political influence by employing government relations teams, lobbyists, PR companies and political advertising. Because of the resources at their disposal, the commercial interests involved in policy, legislative, regulatory actions, trade agreements and treaties and their quick adoption of technology, multinationals have the means and motive to enter the AI influence game in a big way.

For corporations, machine intelligence can shape public opinion and policy/regulator decision-making in profitable and undetectable ways as was discussed in section 2.1.[108] With reference to our earlier discussion on psychographic targeting, the point to note is corporations are no strangers to manipulation and perception shaping. Advertising has long employed manipulative techniques leveraging psychological research and communication expertise to sell products, support or tarnish a policy proposal (the outcome of which the corporation has a commercial interest in), and promote or repair, as the case may be, a corporation's brand and image. It is a small step for them to redeploy machine intelligence and learning to conduct psychographic profiling, and to develop and implement micro-targeting campaigns. The new element in this old game of corporate influence-peddling is due to the leap that AI gives in capability.

Citizens have long been wary of the power of corporations and the way their commercial interests at times side-step democratic decision-making. This caution has grown proportionally with the ties that have developed between politicians, lobbyists, corporate boards and their government relations teams.[109] Moreover, the potentially widespread and effective employment of AI technology to shape the public environment and/or influence decisions, especially in the regulation and tender spaces, and particularly when it is a foreign corporation or multi-national seeking to influence the decisions of a sovereign government, further undermines citizen trust in democracy.

The threat to democracy posed by AI technology is cumulative. Its employment by hostile state actors to target government institutions and decision-makers, by cyber criminals to target corporations and manipulate stock markets, by unscrupulous foreign corporations to target government decision-makers and regulators, to shape or manipulate public opinion, collectively has the potential to seriously undermine the trust that underpins democratic society. Our trust in the accuracy and reliability of information. Our trust in the integrity of the individuals we deal with. And trust in the institutions that are entrusted to maintain the health and sanctity of democratic order, such as parliaments and courts. We will turn to this subject in more detail in section 2.3. Before doing so, we offer a brief account of the risks that governments themselves pose in relation to AI, democracy, and elective processes.

2.2.2 The Internal Threat: Algorithmic government

External threats utilising AI-enhanced capabilities to corrupt democratic institutions and manipulate processes and public discussion are not the only AI specific threat on the horizon. Governments themselves will increasingly be tempted and/or pressured to adopt machine intelligence and learning because of its efficacy in data collection, analysis, decision-making and program administration.[110] The take up of this type of technology threatens democratic processes.[111]

Machine learning is already being employed by authoritarian regimes to enhance surveillance-based governance.[112] It will only entrench and, in some cases, strengthen authoritarian rule. This will increasingly impact the international order, and the status, role and influence of democracies on the global system of governance and relations.

More worryingly, democracies such as Aotearoa New Zealand cannot expect to remain immune to the threat of digital authoritarian systems at home. The risk lies with good intentions and the effort to leverage advances in machine intelligence for governance and administration.[113] But when these are coupled with weak or non-existent regulatory frameworks, AI naivety and/or ignorance, and citizen apathy or even enthusiasm for the uptake of the technology, they run the risk of creating pathways for democratic transformation into digital authoritarianism, or some hybrid form of governance that compromises our democratic health.[114]

In 2018, the New Zealand government commissioned a report into the use of algorithms by government agencies.[115] The report, based on the self-assessment of 14 government agencies, found that algorithms played a crucial role in significant aspects of decision making in the public interest including:

- *Use by police in assessing the risk of future offending in domestic violence cases*
- *Use by Probation Officers in assessing the risk of reconviction/reimprisonment*
- *Use by the Ministry of Health to rank priority for patients awaiting elective surgery*
- *Use by Stats NZ to create projections that inform agency planning*
- *Use by Immigration New Zealand to assign risk ratings to visa applicants*

Significantly, 11 out of 14 agencies reported that the algorithms they used were developed with the assistance of outside expertise, raising concerns that the agencies using the algorithms may not fully understand how results are generated.

Twenty-seven government agencies have signed up to the Algorithm Charter for Aotearoa New Zealand aimed at fostering transparency and accountability in the use of data. A key finding of a 2021 report into the charter was that it had introduced a complex range of considerations that government agencies had neither the capability nor the capacity to uphold. For example, the Charter requires signatories to 'identify and manage bias'. This is recognised as an essential component of algorithm oversight, but it is one that many of the agencies simply lack the subject matter expertise to deliver on.[116]

Accountability is an integral ingredient to a healthy and functioning democracy. It will come under increasing duress. First, because it is hard to see how a machine intelligence can be held to account by committee hearings, a royal commission or the like. Secondly, it will be difficult, if not impossible, for the media to play its traditional role in monitoring and reporting on, for instance, ministerial infractions if decisions are being made by AI programs. Furthermore, soon it will be impossible for humans to understand the rationale, mechanisms, or judgements machine intelligences use in decision making. Notions of accountability will flounder in their opaque operations.[117]

A particular challenge for New Zealand will be the increasing difficulty of maintaining a healthy information space, in which citizens are able to participate in public discussion and debate by drawing on accurate and reliable public information: "ground truth data".[118] The malevolent pollution of our democratic information ecosystem (described in section 2.1 above) risks exacerbating political polarisation across the country, and sowing confusion among our electorate. It threatens to destabilise a crucial component in a democracy: functional and political literacy.

AI also poses the greatest risk to our democracy by making voters irrelevant. Machine intelligence has the potential to make citizen opinion, preferences and participation obsolete.[119] Well-intentioned technocratically-minded elected bodies, tired with fickle voters, may be tempted to hand over more and more decision-making to powerful machine intelligences already enhancing productivity in multiple areas of government.[120] These will be more effective in policy development, for instance, in health, taxation or transport than their human counterparts. They will operate without the cost, confusion, and tedium of consultation.[121] Prudent governments have long relied on data and analysis to inform decision-making ("evidence-based policy"). Machine intelligence simply takes this to a higher level, and with very little financial cost. As Kaplan writes: "An AI-driven system could constantly collect and gather big data on the current opinions, preferences, and desires of a nation's people and citizens. Policy and decisions could reflect in-the-moment public interests and potentially represent these more accurately than within a system where political parties are elected for several years, drifting more or less away from public approval over time. The technology to do this already exists."[122]

But there is something crucial to note. We are not only talking about the use of tools and aides to augment human policy making decisions, but also the handing of decision-making itself over to an AI.[123] If we tread this route, we depart from democratic ways and beat a path towards an algocracy: government by algorithm.[124] Thus, the unalloyed adoption of machine intelligence may tear apart a system of government that relies on elected representation and citizen input into the policy-making process. It will likely deracinate the "core principle [of democracy]—that those affected by political decisions can understand themselves as their authors and that political decisions and outputs are tied back to citizens' views and preferences."[125] Some might argue the gains outweigh the losses, but this is a debate New Zealanders should have while we are able to do so.

In sum, all this means that advances in AI will present democratic governments with dilemmas in which the choice will be between efficiency and cost versus democratic participation and trust. The challenge for countries like New Zealand will be one of balance: maximising the certain benefits to democratic governance offered by machine intelligences while minimising and mitigating their potential harms. External influence from hostile state actors, criminals and multinationals, as well as internal pressure lend urgency to the task. Let us now turn to what the spread of AI means for citizens in a democracy.

2.3 The threat to citizens

The single biggest threat that the poor or malign use of AI technologies poses for New Zealanders is probably the erosion of trust. Trust is essential in relationships. It is critical for business. It is the premise for informed decision-making, medical treatment, and curriculum development and delivery for the next generation. It is a remarkable, vital and yet underrated phenomenon that makes the world go round.

Trust is undermined when we think or feel that we are being deceived, and when we have been used, betrayed, or subjected to a string of broken promises. Trust is also eroded by selfishness. AI either wilfully or accidentally through good intentions has the capacity to scorch trust among us. It can also foment distrust in our own intuitions and judgements, causing paralysis in decision-making and dithering when it comes to action.

The microtargeting and nudging techniques mentioned previously in section 2.1 can cause us to distrust what we have previously believed. This is not necessarily a bad thing. It can lead to critical analysis. But if nudging is

wrought through bot bombardment, and the dissemination of disinformation or the presentation of only half the story, it is problematic. It is deceptive. And such a lack of transparency works to undermine trust in our own ability to make sense of the world. It erodes our confidence.[126]

Indeed, hostile foreign actors exploit this. Intelligence units, criminal groups, and commercial entities use machine intelligence and learning to "ignite a sense of confusion or distrust in the individual… people believing falsehoods [is] a way in which common reference points within society are broken down, thereby undermining systems."[127] Doubting our judgements and the things we hear or read from others, tests our conscience and sense of discrimination. It can cause us to question the value of the right to assembly, to privacy, to freedom of religion, and even speech. Disinformation can powerfully erode commitment to the value of each liberty we enjoy, debasing our sense of self and the freedoms we take for granted. The sowing of distrust of parties, leaders, cabinet ministers and their advisers, the press, big tech, Google, etc., further damages society, democratic processes, and civic institutions.[128]

The distinctive feature of machine intelligence is the speed, breadth, and relentless pressure it brings to persuasion. Adverts, articles, and political messaging on TV are one thing; microtargeting, disinformation and AI quite another. They utterly outclass the former in terms of capability and effect.

In a nutshell, and with some irony, bad state actors and other entities use machine intelligences to exploit our trust and then use that exploitation to attack what we trust. In an increasingly sophisticated manner, they use what we read, see, or hear to assault our values and undermine trust in our own judgements in others, and in the democratic institutions and processes that sustain us as a nation. The risk for New Zealand is that its citizens wallow in low levels of trust, which in turn will stymie social relations and perhaps tempt people with more authoritarian forms of leadership.[129]

3. RECOMMENDATIONS

The primary purpose of this discussion paper is educative. It has described the opportunities and threats that come with AI, especially in relation to elections and government. It has highlighted the hazards AI poses to political candidates, to information, and to public trust. It has also addressed the risks of indifference towards the truth, a lack of appreciation for the democracy that we have, the pressures to maximise efficiencies without pausing to think of the human cost, and autocratic tendencies. These are all trends that could expedite the emergence of algorithmic government in democracies.

The following recommendations are nascent. They do not provide full answers to the challenges outlined in this paper, but they trace paths that could be taken to help safeguard our elective processes, and the practices of participation, consultation, representation, and government accountability to the voting public that lie at the heart of our democracy. They lie at three broad levels of responsibility: individual; national government; and international organisations and diplomacy.

3.1 Education of the Individual

The first recommendation is for all of us. The best counter to most of the risks outlined above comes through individual knowledge and accountability. A good first step is to be aware of the threats that AI can pose to our democracy. But we also need to all develop critical thinking skills, to analyse our sources, to consume a range of media, and engage with different perspectives. The threat AI poses to democracy shrinks if there is open and informed public discourse. This paper seeks to be a small part of this discourse.

Transparency will be the key to maintaining a clean and healthy informational ecosystem. Democratic governments will need to consider the establishment of new information regulatory regimes that embed transparency in an informational space in which AI technology is active. This might include rules around identifying the originators of information, along with their qualifications (establishment of authority), and most crucially, transparency regarding the involvement of AI in the generation and dissemination of the information in question, to enable citizens to properly assess different sources of information. (See, for example, the disclosure of our use of AI in the production of this paper at the beginning of this piece.)

3.2 National Government

The best defence against the erosion of trust engendered by AI in our body politic is an informed and wary public, irrespective of where the AI-generated disinformation or nudging or microtargeting is coming from. However, the New Zealand government also has a role to play in addressing the threats to our democracy posed by this new technology.

As we have shown, the impact of AI is felt throughout society. Its use is already widespread and will continue to grow as the technology grows more sophisticated. The importance of this new technology and its concomitant risks warrant the establishment of a new business unit under the Department of Prime Minister of Cabinet (DPMC), The AI Coordination Group, in order to coordinate and lead the Government's response.

The AI Coordination Group would have a number of functions, but its first role would be fulfilled at its birth: it would demonstrate both how pressing the potential problems that come with the widespread adoption of AI adoption are, and also how serious the consequences to our democracy might be. A business unit within the centrally important DPMC would demonstrate to the public, the Government and other ministries just how necessary it is to take this issue seriously. It would, by its very creation, be educative to the public. It would in effect announce that this is something that the Government takes seriously, and is something therefore that you should also take seriously.

The educative purpose would continue to be a key part of the new business unit's purpose: to help demonstrate to the New Zealand public the dangers of AI and how to protect against them. This could be fulfilled through a variety of ways: running seminars; information evenings; advertisements; public announcements etc. The goal would be to help the public to be savvy consumers of AI generated content so that we are less likely to be misled by deepfakes, disinformation and nudging. To the extent that this new AI Coordination Group can do that, it would have obviated the need for more invasive governmental intervention.

Within Government, this new business unit would be responsible for strengthening the defences around the data stored about New Zealanders. As we have seen, this data is vulnerable to cybercriminals and other international actors. One measure that would mitigate the harm from outside attacks would be to institute data siloes, whereby

information is kept separate from other government institutions. This limits the amount of information that can be gained from any one security breach. On the other hand, this would reduce the efficiencies gained by the pooling of information by government departments, so a balance would need to be considered and, hopefully, struck.[130] The other way in which the AI Coordination Group would help clean up the Government's internal processes is by having oversight and guidance over the use of AI by the different governmental departments. The business unit would provide the expertise about AI that the current ministries lack. They would be able to ensure that the use of AI was in accordance with the Algorithm Charter for Aotearoa New Zealand that the various ministries and agencies have signed up to and that the use of AI was transparent and without bias.

Finally, the new business unit would need to lead a review of New Zealand's existing legislative and regulatory framework to determine if it needs to be amended in the light of the advent of AI. This review would best be done by experts in both the legal and the technological fields, and probably best led by the Law Commission. Such a review would be large and wide-ranging and, in some senses, perhaps futile in that the technology is evolving so quickly that the law will always be playing catch-up. However, this is still a necessary exercise. The end result may not be perfect, but it will be better than what we currently have. The review would look at, for example, privacy laws in light of deepfakes, the effect AI will have on Intellectual Property law, and the requirement for codes of compliance by users of AI. The use of AI in electoral advertising would also need to be considered in conjunction with the Electoral Commission. The efforts of overseas jurisdictions in this area are an obvious first place to look: for example, the EU recently passed a law regulating AI systems to ensure that they safeguard public safety, user rights and data privacy.[131] This review would also need to consider the downsides of governmental overreach when it comes to regulating content to try and prevent AI-generated disinformation. The concerns around government-controlled speech and the issues around defining what is disinformation remain whether or not that disinformation is AI-generated. The AI Coordination Group's review should not be used as a way of limiting free speech under the guise of regulating AI.

3.3 International Organisations and Diplomacy

The profound effects of AI on human society demand a proportionate response. Local and national initiatives will not be enough on their own. Something with a wider reach, like the *Christchurch Call*, needs to be established. Or, with the prospect of an artificial general intelligence (AGI) looming, a new *Manhattan Project*. Whether *Call* or *Project*, the proposed entity has to be an embodiment of the most able ethical, political, and scientific leaders from healthy democracies. Although transnational communities of professionals, academics, ethicists, etc. are already at work on many of the ethical and safety concerns related to the development of AI technologies, there is a need for a higher-level entity with more bite. Much as *Five Eyes* aims at physical security, it would aim to safeguard democratic life.

There are already nascent attempts to bring (at least parts of) the international community together to discuss the safety of AI advances. Later on in 2023, the UK Government will hold the "first major global summit on AI safety" which will discuss how the risks of AI can be mitigated through "globally coordinated action".[132] Whether or not this will lead to anything concrete like a *Manhattan Project* for AI is yet to be seen, but New Zealand should be part of this summit so that its views and interests are heard. As a small nation, we cannot do much in response to global AI on our own. We must work with other countries to try and ensure that the serious risks that AI poses to our democratic systems are recognised and mitigated.

To this end, we should argue for a new *Call* or *Project* premised on the intuition that people, not technology, is the normative centre of all public power. It would describe the human vectors to defend (in relation to democracy) from the threat of AI. These would likely privilege epistemic rights which acknowledge: (1) that learning is indispensable to human growth and happiness—and that the acquisition of knowledge involves trial and failure; and (2) the right to know who or what we are interacting with.

It would also define and outline the necessary protections for democratic processes. AI, as it stands, focuses on information processing and on solving cognitive and coordination tasks. Participation and consultation are not its concerns. It processes data and makes recommendations on that basis. It is statistically and probabilistically driven. In contrast, a liberal democracy has special informational

requirements. These exist not just for *outputs*, but for safeguarding certain forms of human activity and decision-making (*throughputs*) as well as the continuous integration of public preferences (*inputs*). Participation, consultation, implementation, reaction, reflection, feedback, and emendation are human elements of democracy.

At root the project should tackle the central issue of accountability. One major risk of a drift to algorithmic government is the lack of accountability. The need then is to articulate and defend not just those elements which are necessary for democratic processes, but also at a more fundamental level it would defend what makes these specifically human requisites. Such a project would carry a defence of human work and the value of employment as well as an acceptance of human failure as a point of growth over and against algorithmic perfectionism.

A new *Call* or *Project* should also settle on a framework and/or hierarchy of values. While questions remain as to (1) the basis for human rights and (2) which values/whose values, the need for democracies to agree on a charter of rights and values in relation to AI is pressing. Given its actual and potential power, there is a case for seeding values we deem essential to human welfare and democratic health into AI's evaluative processes. As we have noted, experts in AI doubt it is possible to "load" AI with values or create human-AI value alignment, or "contain" AI, but they admit the importance of trying to do so. In this case, if something proves better than nothing, it will be to everyone's benefit. The International Bill of Human Rights (the Universal Declaration on Human Rights, the International Covenant on Economic, Social and Cultural Rights and the International Covenant on Civil and Political Rights) might be sufficient for our needs or provide the foundations for further development.

Necessary conditions for success in a *Call* or *Project* will comprise commitment, personnel, and also funding for research, testing, and evaluation, and all the associated costs of these. This funding should be provided proportionally by participating democracies.

4. CONCLUSION

We may be on the cusp of the AI revolution. In decades to come we may look back at the 2020s as the time that AI fundamentally changed our economy, our social interactions and our day-to-day living. We can already see the impact of AI on our democratic systems. Microtargeting, nudging and deepfakes are all helping to dissolve the bonds of trust between citizens and their elected representatives. AI has also increased the scope, power and reach of international state competitors, cyber criminals and global corporations. While AI may help erode our democracy from the outside, the lure of AI-determined policy may tempt us to voluntarily throw away democratic government for government by algorithm.

Aside from wrenching ourselves away from modernity to live "off the grid", there is not much we can do to escape the effect of AI. However, we can all lessen its impact upon our democracy by being aware of its reach and capabilities. This is what this paper is primarily designed to do: to educate and warn us to be "AI-savvy". There are also steps that the government can take to minimise AI's harm as much as possible. Internationally, we should be pushing for a second *Manhattan Project* to emphasise the importance of the human over the machine. By these means we will hopefully be able to look back at the 2020s as a time when we grasped the benefits of the AI revolution while warily sidestepping its dangers.

ENDNOTES

1 "The term AI, coined by John McCarthy in 1956, is elusive in its precise meaning but today broadly refers to machines that can go beyond their explicit programming by making choices in ways that mirror human reasoning. In other words, AI automates decisions that people used to make." Eileen Donahoe and Megan MacDuffee Metzger, "Artificial Intelligence and Human Rights," *Journal of Democracy* 30, no. 2 (2019), https://doi.org/10.1353/jod.2019.0029 at 115.

2 "New technologies have always challenged, if not disrupted, the social, economic, legal, and, to a certain extent, ideological status quo.…Constitutions have been designed to limit public (more precisely governmental) powers and protect individuals against any abuse from the state." Oreste Pollicino and Giovanni De Gregorio, "Constitutional Challenges in the Algorithmic Society," (2021), https://doi.org/10.1017/9781108914857 at 5.

3 And like Ernest Hemingway's comment on bankruptcy happening "gradually, then suddenly", AI's entry into ordinary life is now upon us.

4 AI is receiving a lot of attention in the media, but not all the press is positive. Donahue et.al. note: "In digitally connected democracies, talk of what could go wrong with AI now touches on everything from massive job loss caused by automation to machines that make discriminatory hiring decisions, and even to threats posed by "killer robots." These concerns have darkened public attitudes and made this a key moment to either build or destroy public trust in AI." Donahoe and Metzger, above n 1, at 115.

5 "While it is important to understand where AI currently threatens democracy, it is as crucial to appreciate its opportunities. To understand the openness of the use and potential of technology allows us to choose whether to develop the technology further and which path to take. When it comes to the democratization of AI, some general truths about democracy apply: Democracy is a process, not an achievable result. It can be lost very easily…AI is just another challenge that has the potential to bring society closer to the ideal lying behind Art. 12 section of the Bremen Constitution, as well as many other democratic provisions: to meaningfully put people in the normative center of all public power." Christian Djeffal, *The Democratization of Artificial Intelligence: Net Politics in the Era of Learning Algorithms*, ed. Andreas Sudmann (Columbia: Columbia University Press, 2019) at 280.

6 For example: "In public administration, AI serves to, for example, forecast health care needs, predict domestic violence, or to identify cases of fraudulent social benefit claims. On higher levels of policy making, AI tools may assist with decisions and help to achieve greater efficiency and effectiveness (Höchtl et al., 2016), as is the case with algorithmic systems used by finance ministries for policy simulations to help with policy planning (Kolkman, 2020). In a similar vein, the US Food and Drug Administration piloted an application that processes reports of adverse events after a drug has been launched on the market to detect undesirable effects and to adaptively inform its rulemaking (Coglianese & Ben Dor)." Pascal Konig, "Citizen Conceptions of Democracy and Support for Artificial Intelligence in Government and Politics," *European Journal of Political Research* 61, no. 4 (2022), https://doi.org/10.1111/1475-6765.12570 at 2.

7 *Artificial Intelligence in Society*, (Paris: OECD, 2021) at 47-80. Note, too, the quantity of data available for use: "The global volume of data circulating in 2020 [alone reached] 44 zettabytes, or, to give a slightly more understandable idea, to *8 billion* times the information collected in the library of Congress in Washington."

8 Nestor Maslej et al., *The AI Index 2023 Annual Report*, AI Index Steering Committee, Institute for Human-Centered AI (Stanford, 2023) at 70-124.

9 *Artificial Intelligence and Employment: New Evidene from Occupations Most Exposed to AI*, (Paris: OECD, 2021); *The Future of Jobs*, World Economic Forum (Geneva, 2023) at 37-42; Max Tegmark, *Life 3.0: Being Human in the Age of Artificial Intelligence* (New York: Penguin, 2018) at 82-133.

10 Amir Husain, *The Sentient Machine: The Coming of Age of Artificial Intelligence* (New York: Scribner, 2017) at 113; Donald MacKenzie, *Trading at the Speed of Light: How Ultrafast Algorithms are Transforming Financial Markets* (Princeton: Princetone University Press, 2021) at 172-205.

11 Husain, above n 10, at 109-29.

12 Tegmark, above n 9, at 83.

13 Bonnie G. Buchanan, *Artificial Intelligence in Finance*, The Alan Turing Institute (Seattle, 2019) at 15-18.

14 See, for instance, James Wright, *Inside Japan's Long Experiment in Automating Elder Care* (2023), https://www.technologyreview.com/2023/01/09/1065135/japan-automating-eldercare-robots/#:~:text=Japan%20has%20been%20developing%20robots,and%20development%20for%20such%20devices; Carina Soledad González-González, Rosa María Gil-Iranzo, and Patricia Paderewski-Rodríguez, "Human–Robot Interaction and Sexbots: A Systematic Literature Review," *Sensors (Basel)* 21, no. 1 (2021), https://doi.org/https://doi.org/10.3390%2Fs21010216.

15 Depicted in movies like *Zoe* (2018); *Her* (2013); *Blade Runner* (1982).

16 Tegmark, above n 9, at 280-315. Note the debates that have raged over Blake Lemoine, Google engineer, commentary by Ilya Sutskever, Chief scientist of the OpenAI, and neuroscientists like Daniel Dennett. See, also: Martin Kingwell, "Sentient AI Persons," in *The Oxford Handbook of Ethics* (Oxford: Oxford University Press, 2020). And on a lighter (?) note: Thomas Weber, "The Inventor Who Fell in Love with His AI," *The Economist* 2023, https://www.economist.com/1843/2023/04/04/the-inventor-who-fell-in-love-with-his-ai. In response see: Emily M. Bender et al., "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" (Conference on Fairness, Accountability, and Transparency, 2021).

17 Gina Granados Palmer, "AI Ethics: Four Key Considerations for a Globally Secure Future," in *Artificial Intelligence and Global Security* (Emerald Publishing Limited, 2020) at 167-76; Pauline Shanks and Kaurin and Casey Thomas Hart, "Artificial Intelligence and Moral Reasoning: Shifting Moral Responsibility in War?," in *Artificial Intelligence and Global Security* (Emerald Publishing Limited, 2020) at 121-26; Husain, above n 10, at 69-109.

18 Wannacry; Notpetya; Equifax; Mirai Botnet; SolarWinds. Husain, above n 10, at 66-79; Raghav Sandhane, "Artificial Intelligence in Cyber Security" (Journal of Physics Conference Series, 2021). Note, also, the use of machine intelligence in post-incident analysis: *Threat Report*, Blackberry Intelligence (Toronto, 2022) at 40-58.

19 *Threat Report*, above n 18, at 42-48.

20 Infamously, Yusuf Çetiner, "The Turkish Kargu-2 Carries Out The First Autonomous Drone Attack," *Overt Defence* 2021, https://www.overtdefense.com/2021/06/02/the-turkish-kargu-2-carries-out-the-first-autonomous-drone-attack-un-report-says at 1: "It is worth noting that many analysts think that defence will be the first arena to witness full handing of control over to an AI or Artificial General Intelligence (AGI)—i.e., giving an AGI autonomy in determining its goals and means for achieving them (the prelude to a sense of self). Even so, imagine a battle scenario, where for instance, the commander (an admiral) of an aircraft carrier with a fleet of warships supporting it is attacked simultaneously on multiple fronts by swarms of drones, missiles and cyber-attacks on communications and weapon systems. And further, the same commander, cycling through information available through a visor, can see the rising battle toll—increasing human injury and losses (personnel are wearing biosensors), damage to infrastructure, and flame-laden sinking ships. AI has given the commander massive situational awareness: now the choice is to give it full control for the defence of the fleet. And to go on the offensive in response to the attack. The decision is taken to do so, and AI saves the day, because it perceives, decides, and acts from a richer base of information, and a more creative and far, far quicker intelligence than that of the humans being slaughtered about it. But giving it command has meant surrendering control. It has diminished human agency and opened the doors to a possible world where AIs take the lead in thinking and acting, and humans do not." Taken from Tegmark, above n 9.

21 Husain, above n 10, at 57-68.

22 See, for example: James Gallagher, "Baby Born from Three People's DNA in UK First," *BBC* 2023, https://www.bbc.com/news/science-environment-65538866#.

23 Explored in the movie *Interstellar* (2014).

24 Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies* (Oxford: Oxford University Press, 2014) at 87-90; Roman V. Yampolskiy, "On the Controllability of Artificial Intelligence: An Analysis of Limitations," *Journal of Cyber Security and Mobility* 11, no. 3 (2022), 364-81, https://doi.org/10.13052/jcsm2245-1439.1132. And see: TED, "Superintelligence," (1 June 2023 2016) https://www.bing.com/videos/search?q=TED+talk+Nick+Bostrom+superintendent%3bigence&docid=60354205 0143143424&mid=DD5948BC03F57F002742DD5948BC03F57F002742&view=detail&FORM=VIRE.

25 Frank Landymore, "Godfather of AI Says There's a Minor Risk It'll Eliminate Humanity," *Futurism*, 27 March 2023, https://futurism.com/the-byte/godfather-ai-risk-eliminate-humanity. Or: Victor Ordonez, Taylor Dunn, and Eric Noll, "OpenAI CEO Sam Altman says AI will reshape society, acknowledges risks: 'A little bit scared of this'," 17 March 2023, https://abcnews.go.com/Technology/openai-ceo-sam-altman-ai-reshape-society-acknowledges/story?id=97897122.

26 In relation to the emergence of an artificial general intelligence and beyond this a superintelligence see: Bostrom, above n 24, at 329-41.

27 This is no exaggeration: "A tremendous amount of digital trace data—Big Data—is collected behind the screens using the trail we leave behind us while we navigate and interact with digital media, through clicks, tweets, likes, GPS coordinates, timestamps (Lewis and Westlund, 2015), and sensors by smart information systems, which register information of our behaviour, beliefs and preferences for profiling citizens. Those who have access to citizens' digital data and profiling know more about individuals than probably their friends, family members or even individuals themselves (Smolan, 2016). That data is then used to tailor the information we receive in an attempt to influence our behaviour for the sponsors' benefit, for securing, for example, financial profit or winning the elections, as was the case with Cambridge Analytica (Isaak and Hanna, 2018)." Eleni Christodoulou and Kalypso Iordanou, "Democracy Under Attack: Challenges of Addressing Ethical Issues of AI and Big Data for More Democratic Digital Media and Societies," *Frontiers in Political Science* 3 (2021), https://doi.org/10.3389/fpos.2021.682945 at 4.

28 For example: Ralph Schroeder, *Aadhaar and the Social Credit System: Personal Data Governance in India and China*, vol. 16 (2022), https://ijoc.org/index.php/ijoc/article/view/19059; Eunsun Cho, "The Social Credit System: Not Just Another Chinese Idiosyncrasy," *Princeton Journal of Public and International Affairs*, May (2020), https://jpia.princeton.edu/news/social-credit-system-not-just-another-chinese-idiosyncrasy.

29 Popular commentary includes: Andrew Drenth, *The 16 Personality Types: Profiles, Theory, & Type Development* (Inquire Books, 2017); Otto Kroeger, *Type Talk: Or How to Determine Your Personality Type and Change Your Life* (New York: Delacorte Press, 1988).

30 Barrie Gunter, *The Psychology of Consumer Profiling in a Digital Age (Routledge Studies in Marketing)* (New York: Routledge, 2016) at 105-28.

31 Suman Gupta, "Procedural Democracy and Electronic Agents," *AI Magazine* 42, no. 1 (2021) at 123.

32 Gupta, above n 31, at 123. And: "AI has the capacity to generate and analyse huge data sets; to guide a diverse group of hardware systems, ranging from mobile phones, robotic vacuum cleaners and intelligent washing machines, to surveillance cameras and autonomous weapons systems; and to draw inferences about humans and human behaviour." Birgit Schippers, "Artificial Intelligence and Democratic Politics," *Political Insight* (March 2020) at 15.

33 Note: "AI can make predictions about a person's behaviour, state of mind, and identity by sensing information that is not necessarily considered personal or private, such as facial expressions, heart rate, physical location, and other seemingly mundane or publicly accessible data. This can have the effect of being invasive of a person's sense of privacy, and can also have so-called 'panoptic effects' by causing a person to alter their behaviour upon suspicion it is being observed or analysed." Michael Chui et al., *The State of AI in 2022—and a Half Decade in Review*, McKinsey (McKinsey, 2022) at 15. And of political concern: "During the past years, democratic political systems in Europe and globally have been significantly endangered by spreading of disinformation (fake news), particularly during the run up to elections. .... The spread of disinformation through automated accounts and bots, coupled with psychographic micro-targeting, does not only reach an incomparably greater number of voters, but also appeals to their sensitivities, fears and psychological characteristics." Maja Brkhan, "Artificial Intelligence and Democracy: The Impact of Disinformation, Social Bots and Political Targeting," *Delphi* 2 (2019), https://doi.org/10.21552/delphi/2019/2/4 at 69.

34 "Indeed, personal data has become the most prized commodity of the digital age, traded on a vast scale by some of the most powerful companies in Silicon Valley and beyond. The result is the datafication of society." Andreas Kaplan, "Artificial Intelligence, Social Media, and Fake News," in *Digital Transformation in Media and Society*, ed. Ayşen Akkor Gül, Yıldız Dilek Ertürk, and Paul Elmer (Istanbul: Istanbul University Press, 2020) at 119.

35 "While conferring some consumer benefit, the principal function at present is to capture personal information, create detailed behavioural profiles and sell us goods and agendas. Privacy, anonymity and autonomy are the main casualties of AI's ability to manipulate choices in economic and political decisions." Nicholas Wright, *Artificial Intelligence and Democratic Norms: Meeting the Authoritarian Challenge* (Washington: Think Democracy, 2020) at 106.

36 Karl Manheim and Lyric Kaplan, "Artificial Intelligence: Risks to Privacy and Democracy," *The Yale Journal of Law and Technology* 21 (2019) at 119.

37 "At the same time, technologies are advancing at pace, giving rise to complex patterns of data use and decision-making. In this complex environment, data collected for one purpose can be rapidly repurposed or shared in ways that are opaque or unanticipated at the point of data collection....This growing complexity of data processing compounds the limitations of consent-based models of data governance, which have been well-characterised elsewhere....Existing policy frameworks are not necessarily well-placed to manage these net-work effects. They present different circumstances to those envisaged in the early stages of drafting the GDPR in the 1990s, where policymakers were primarily concerned with the use of data to inform decision-making in areas that might have a significant personal or social impact." Sylvie Delacroix, Joelle Pineau, and Jessica Montgomer, "Democratising the Digital Revolution: The Role of Data Governance," *Artificial Intelligence for Humanity* (2021), https://doi.org/10.1007/978-3-030-69128-8_3 at 45.

38 Pascal D. König and Georg Wenzelburger, "Opportunity for Renewal or Disruptive Force: How Artificial Intelligence Alters Democratic Politics," *Government Information Quarterly* 37 (2020), https://doi.org/10.1016/j.giq.2020.101489 at 5. And infamously: Christodoulou and Iordanou, above n 27 at 4.

39 See: Akshat Agarwal, Charu Singhal, and Renny Thomas, *AI-powered decision making for the bank of the future*, McKinsey & Company (New York, 2021), https://www.mckinsey.com/~/media/mckinsey/industries/financial%20services/our%20insights/ai%20powered%20decision%20making%20for%20the%20bank%20of%20the%20future/ai-powered-decision-making-for-the-bank-of-the-future.pdf. And hence responses such as: *Reducing bias in AI-based financial services*, Brookings Institute (Washington, 2020), https://www.brookings.edu/research/reducing-bias-in-ai-based-financial-services/.

40 Ralph Breuer et al., *The Rise of Digital Marketing in Medtech*, McKinsey & Co (New York, 2021), https://www.mckinsey.com/industries/life-sciences/our-insights/the-rise-of-digital-marketing-in-medtech. But note John Osborn, *White House Communications Microtargeting Aims To Further Reduce Vaccine Hesitant Population* (Forbes, 2021), https://www.forbes.com/sites/johnosborn/2021/09/01/white-house-communications-microtargeting-aims-to-further-reduce-vaccine-hesitant-population/?sh=c65f2d963cb8.

41 Indeed, "it would be naive to ignore that for most in our societies today, the reality of how they use the Internet and what the Internet delivers to them is shaped by a few mega corporations, as it would be naive to ignore that the development of AI is dominated exactly by these mega corporations and their dependent ecosystems." Paul Nemitz, "Constitutional Democracy and Technology in the Age of Artificial Intelligence", (Belgium: European Commission, 2018) at 3.

42 Xavier Amatriain and Justin Basilico, "Recommender Systems in Industry: A Netflix Case Study," in *Recommender Systems Handbook*, ed. Francesco Ricci, Lior Rokach, and Bracha Shapira (Boston, MA: Springer US, 2015) at 385-419; Netflix, "How Netflix's Recommendations System Works," (1 June 2023), https://help.netflix.com/en/node/100639.

43 "The Anglo-American company called Cambridge Analytica (CA) [were] harvesting data from over 50 million Facebook users, without their consent or knowledge. Assisted by a specifically designed algorithm, the profiles of Facebook users were analysed to identify personality traits and potential voting intentions. This information, in turn, helped to identify swing voters and micro-target them with messages that, according to critics, constituted deliberate misinformation. Today, as the US prepares for another presidential election and Britain has left the European Union after a referendum characterised by aggressive social media targeting, Cambridge Analytica has become a by-word for technology-driven interferences in elections…[and] has become emblematic of the growing use and impact of artificial intelligence (AI) on our democratic processes and institutions." Schippers, above n 32 at 32. And: "CA boasted to having up to four thousand distinct data points on each adult in the United States, by combining data from public sources, such as voter registration databases, third-party data brokers with information, such as magazine subscription, web page visits, social media behaviours, and zip code—which reveals key demographic information, such as race, age, income, and education levels. CA claimed to be a technological leap over other data analytics approaches because of their ability to generate "psychographic" profiles of people in their database…. CA worked with the Trump campaign to identify target voters. Alongside three companies contracted by the Republican National Committee, TargetPoint Consulting, Causeway Solutions, and Deep Root Analytics—which helped with field operations and television advertising—CA embedded over a dozen staff in Austin, Texas, with the digital staff of the Trump campaign to support data analytics and targeted messaging along with guidance from Facebook, Twitter, and Google." Jennifer Stromer-Galley, *Presidential Campaigning in the Internet Age*, Second edition. ed., Oxford Studies in Digital Politics, (New York, NY: Oxford University Press, 2019) at 204.

44 For example, very early on in the school voucher/charter school debates, Justice Sandra Day O'Connor was subject to a microtargeting campaign during Supreme Court deliberations on school voucher programs in *Zelman v. Simmons-Harris*, 536 U.S. 639 (2002).

45 P. G. Hansen and A. M. Jespersen, "Nudge and the Manipulation of Choice: A Framework for the Responsible Use of the Nudge Approach to Behaviour Change in Public Policy," *European Journal of Risk Regulation* 4, no. 1 (2013), https://doi.org/10.1017/s1867299x00002762; C. Schneider, M. Weinmann, and J. vom Brocke, "Digital Nudging: Guiding Online User Choices through Interface Design," *Communications of the ACM* 61, no. 7 (2018), https://doi.org/10.1145/3213765. And more cautiously: Dennis Hummel and Alexander Maedche, "How Effective is Nudging? A Quantitative Review on the Effect Sizes and Limits of Empirical Nudging Studies," *Journal of Behavioral and Experimental Economics* 80 (2019), https://doi.org/10.1080/12460125.2023.2198056.

46 Christodoulou and Iordanou, above n 27 at 4.

47 Stromer-Galley, above n 43, at 154.

48 Christodoulou and Iordanou, above n 27, at 1.

49 Indeed: "One of the most important features of bots is that they can achieve scalability, which enables them to massively spread information and hence to (artificially) enhance the importance of a particular idea or popularity of a political candidate… and from a technical perspective, the creation of social bots on social networks is becoming increasingly easier. On Twitter, for example, the creation of social bots is greatly facilitated through the open application programming interface (API) and it has been found that bots represent up to a quarter of all Twitter accounts. Facebook is equally populated with a considerable number of fake accounts…. [Moreover], automated social bots can be (mis)used to promote political candidates and convince the voters to vote for this candidate even if they do not spread disinformation, in particular if coupled with microtargeting." above n 33 at 66, 68.

50 Casonato writes: "As with internet advertising, in which people are increasingly offered items corresponding to their tastes, also for political and electoral information, the risk is that the person will be exposed to very limited information specifically custom-made on the basis of her profiled preferences." Carlo Casonato, "AI and Constitutionalism: The Challenges Ahead," in *Reflections on Artificial Intelligence for Humanity* (Switzerland (AG): Springer, 2021) at 132.

51 At a fundamental level, nudging aims "to influence the choices of people by changing their choice architecture, without constraining what Berlin calls their negative freedom, their freedom from constraint: people are still free to choose. However, their choice architecture and environment are structured in specific ways on purpose, to influence those choices. Thaler and Sunstein are right to say that people are externally and negatively free, in the sense that no-one is directly interfering with their freedom of choice; however, people's internal autonomy and agency seems to be compromised because they are influenced in subconscious ways." Mark Coeckelbergh, "Democracy, Epistemic Agency, and AI: Political Epistemology in Times of Artificial Intelligence " *AI and Ethics* (2022), https://doi.org/10.1007/s43681-022-00239-4 at 9.

52 Alison Jones, "Toronto and Surrounding Area Key to Liberal Electoral Victory," *City News* (Toronto) 2019, https://toronto.citynews.ca/2019/10/22/toronto-and-surrounding-area-key-to-liberal-electoral-victory/; Jill Mahoney and Jeff Gray, "Federal Election 2019: Liberals Maintain Hold on Key Ridings in Vote-rich Ontario," *The Glode and Mail* 2019, https://www.theglobeandmail.com/canada/article-federal-election-2019-liberals-maintain-hold-on-key-ridings-in-vote/.

53 Brkhan, above n 33, at 67.

54 Aumyo Hassan and Sarah J. Barber, "The Effects of Repetition Frequency on the Illusory Truth Effect," *Cognitive Research: Principles and Applications* 6, no. 38 (2021), https://doi.org/10.1186/s41235-021-00301-5.

55 "A group of bots working collaboratively is known as a botnet (Burkhardt, 2017) … information may also be directed to those within the user's network, creating confined bubbles of information. Burkhardt argues that individuals within these echo chambers may feel as though their ideologies reflect those of the majority of the population, as they no longer see information that conflicts with their opinions (Burkhardt, 2017). The benefit of bots for parties wishing to disseminate information is that they work around the clock and work much faster than humans. What is more, humans assist bots with their duties by engaging with the content by sharing across their own networks." Annie Benzie and Reza Montasari, "Artificial Intelligence and the Spread of Mis- and Disinformation," *Artificial Intelligence and National Security* (2022), https://doi.org/10.1007/978-3-031-06709-9_1 at 8.

56 Benzie et al. suggests "that it is not the case that the content produced by AIs are able to outright change one's opinion or political beliefs, but rather they ignite a sense of confusion or distrust in the individual. In other words, it sows the seeds which may later grow (Weedon et al., 2017)." Benzie and Montasari, above n 55, at 5.

57 Brkhan, above n 33, at 66.

58 "Algorithmic sorting furthermore makes it possible for third parties to purposefully supply users e.g., of social network sites with tailored messages. The combination of publicly available voter data with consumer and social media data and enhanced analytical capacities allow for generating detailed voter profiles, such that political actors develop unprecedented possibilities of targeting citizens with messages (Franz, 2013; Hersh, 2015)." König and Wenzelburger, above n 38, at 5.

59 Kaplan, above n 34, at 155.

60 To the extent that "sophisticated manipulation technologies have progressed to the point where individuals perceive that decisions they make are their own, but are instead often "guided" by an algorithm." Manheim and Kaplan, above n 36, at 110.

61 Journalism and Media Staff, *How the Presidential Candidates Use the Web and Social Media*, Pew Research Centre (2012), https://www.pewresearch.org/journalism/2012/08/15/how-presidential-candidates-use-web-and-social-media/. And note: "The Trump [2016] campaign capitalized on the power of digital advertising to reach the public to engage in unprecedented mass-targeted campaigning. His campaign spent substantially more on Facebook and other digital media paid ads than Clinton. Yet, the company that Trump worked with, Cambridge Analytica, closed up shop in 2018 under a cloud of controversy about corrupt officials and voter manipulation in several countries, as well as ill-begotten data of Facebook users that drove their micro-targeting practices. The Clinton campaign modelled itself on data-driven successes of the Obama campaign, yet the algorithms that drove their decision making were flawed, thereby leading her campaign to underperform in essential swing states." Stromer-Galley, above n 43, at 179.

62 Moreover, in the 2018 Brazilian presidential election, Jair Bolsonaro's campaign made extensive use of social media and targeted online advertising to reach specific voter segments. By analysing users' online behaviour and preferences, the campaign was able to serve personalised ads to groups. For instance, messages highlighting Bolsonaro's tough stance on crime were targeted towards voters concerned about public safety, while ads focusing on economic liberalization appealed to those worried about the country's financial future. A targeted approach helped Bolsonaro build a diverse coalition of supporters and secure his election victory. Worryingly, analysts note that "disinformation was spread with the help of bots in Germany during debates on the UN migration pact and during the 2017 German elections, in Sweden during the 2018 elections… and also in the run up to the Catalan independence referendum, [as well as] the framework for a debate on immigration in Italy, and in the run-up to recent European elections." Brkhan, above n 33, at 68.

63 "Meta Ad Library Report: New Zealand," Meta, 2023, accessed 26/05/2023, https://www.facebook.com/ads/library/report/?source=nav-header.

64 "Removing Certain Ad Targeting Options and Expanding Our Ad Controls," *Facebook Announcements*, *Meta*, 09/11/2021, 2021, https://www.facebook.com/business/news/removing-certain-ad-targeting-options-and-expanding-our-ad-controls. Note, too: "The 2016 US campaign was unprecedented on many levels. Seventeen Republicans and five Democrats all vied for their party's nomination, making the primaries the most hotly contested in modern history. The social media platforms had reached a state of maturity, with Facebook rivalling cable television in terms of reach and gave campaigns powerful new tools for advertising on their platforms and elsewhere." Stromer-Galley, above n 43, at 210.

65 Stephen Feldstein notes: "Social-media platforms use content-curation algorithms to drive users toward certain articles—and keep them addicted to their social-media feeds…[people] can exploit such algorithms to push out pro-regime messaging using bot and troll armies for hire. AI can help to identify key social media "influencers," whom can then [be] co-opted into spreading disinformation. Emerging AI technology can also facilitate the deployment via social-media platforms of automated, hyper personalized disinformation campaigns—targeted at specific individuals or groups." Steven Feldstein, "The Road to Digital Unfreedom: How Artificial Intelligence is Reshaping Repression," *Journal of Democracy* 30, no. 1 (2019), https://doi.org/10.1353/jod.2019.0003 at 44.

66 Barry King, "Cultural Studies and the Politics of Celebrity," in *Routledge Handbook of Celebrity Studies*, ed. Anthony Elliott (London: Routledge, 2018).

67 Benzie and Montasari, above n 55, at 4.

68 "Misinformation may be defined as 'a claim that contradicts or distorts common understandings of verifiable facts' (Guess & Lyons, 2020). It has also more generally been defined as 'false or misleading information' (Lazer et al., 2018). While misinformation is false information which may be disseminated mistakenly, the OED defines disinformation as 'the dissemination of deliberately false information' (European Parliament, 2015)." Benzie and Montasari, above n 55, at 4.

69 Benzie and Montasari, above n 55, at 2.

70 Note: "While most disinformation is based in one singular modality such as text or video, multimodal strategies are often used to make news articles more believable by having accompanying text alongside a fake video, for example." Benzie and Montasari, above n 55, at 7. Moreover: "Besides the orchestrating efforts of nudging (Helbing et al., 2019), digital media constitutes a fertile ground for the spread of misinformation through the profound absence of any form of gatekeeping. Misinformation, hate speech and conspiracy theories have found a way to reach thousands of citizens through digital media, especially social media, threatening political and social stability (Frank, 2021). Although concerns regarding the deliberate dissemination of the information in order to affect public perception were evident before (Bauer and Gregory, 2007), those issues have been amplified with the use of AI in digital media." Benzie and Montasari, above n 55, at 4.

71 Benzie and Montasari, above n 55, at 2.

72 Brkhan, above n 33, at 66.

73 Erik Ofgang, "How to Teach with Deep Fake Technology," *Tech & Learning*, *Tech & Learning*, 2022, https://www.techlearning.com/news/how-to-teach-with-deep-fake-technology.

74 Brkhan, above n 33, at 69. Note, too: "Technological advancements have made it possible to use generative adversarial networks to replace faces of a person in video content with the face of another person entirely.… George (2019) found that in cases where deepfakes are used to undermine politicians and political campaigns, this has the potential to have a resulting impact on the public sentiment towards the politician or political party in question (George, 2019). In terms of deepfake detection, interesting research strategies have been investigated. One such strategy is by Li et al. (2018), who sought to detect abnormalities in blinking." Benzie and Montasari, above n 55, at 7.

75 Ruth Michaelson, "Turkish Presidential Candidate Quits Race After Release of Alleged Sex Tape," *The Guardian*, 11/05/2023 2023, https://www.theguardian.com/world/2023/may/11/muharrem-ince-turkish-presidential-candidate-withdraws-alleged-sex-tape.

76 Kirsty Wynn, "Aja Rock and Kiwi Women Targeted in Deepfake Videos, Used to Scam Friends and Followers out of Thousands," *New Zealand Herald*, 11/02/2023.

77 Samantha Bradshaw and Philip N. Howard, *Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation*, The Computational Propaganda Project (University of Oxford, 2018).

78 Adding, "Google's YouTube also profits nicely from fake news. Its "recommendation algorithm" serves "up next" video thumbnails that its AI program determines will be of interest to each of its 1.5 billion users. The algorithm, which "is the single most important engine of YouTube's growth," revels in promoting conspiracy theories. While most of the attention has been directed at Facebook and Twitter, "YouTube is the most overlooked story of 2016.… Its search and recommender algorithms are misinformation engines." One exposé has found that "YouTube systematically amplifies videos that are divisive, sensational and conspiratorial." Manheim and Kaplan, above n 36, at 146, 148.

79 "Fact Check - Clip of Jacinda Ardern Smoking Cannabis is a Deepfake," *Reuters Fact Check*, *Reuters*, 2021, https://www.reuters.com/article/factcheck-newzealand-ardern-idUSL1N2S71BZ.

80 Tess McClure, "New Zealand's National Part Admits Using AI-Generated People in Attack Ads," *The Guardian*, 24/03/2023 2023, https://www.theguardian.com/world/2023/may/24/new-zealand-national-party-admits-using-ai-generated-people-in-ads.

81 "Editorial: Alternative Reality of Deep Fakes," *New Zealand Herald* 2023, https://www.nzherald.co.nz/nz/editorial-alternative-reality-of-deep-fakes/VY7LBGLNXFEKFC7AFO3YARP67E/.

82 *Survey Findings - Cycle 4 Report: Section 7 - Reporting to the Police*, (New Zealand Ministry of Justice, 2022), https://www.justice.govt.nz/assets/NZCVS-Cycle-4-Core-Report-Section-7-Reporting-to-the-Police-fin.pdf at 3.

83 Anusha Bradley, "The Rise of Cybercrime and NZ's FragmentedResponse," *New Zealand In Depth* 2021, https://www.rnz.co.nz/news/in-depth/444735/the-rise-of-cybercrime-and-nz-s-fragmented-response.

84 "We trust web-sourced information provided by platforms, assuming they are scientifically accurate or at least trustworthy. However, this trust has nothing to do with science or education. Platforms simply use powerful algorithms that learn behavioural patterns from previous preferences to reinforce individuals or groups in filtering overwhelming alternatives in our daily life. The accuracy of these algorithms in predicting and giving us helpful information with their results only occurs because they confirm – feeding a "confirmation bias" – our beliefs or, worst, our ideological positions ("bubble effect")." Andrea Simoncini and Erik Longo, "Fundamental Rights and the Rule of Law in the Algorithmicm Society," in *Constitutional Challenges in the Algorithmic Society*, ed. Hans- W. Micklitz et al. (Cambridge: Cambridge University Press, 2021) at 40.

85 König and Wenzelburger note: "Filter bubbles and epistemic bubbles are created by means of recommender algorithms that select content that matches users' profile and online history. Users can also deliberately block people with opposing views, the filtering can be intended or not. The creation of epistemic bubbles can also happen through search engines which increasingly have a recommender dimension, create a bubble based on the users' searches through personalization." König and Wenzelburger, above n 38, at 6.

86 As such: "The pluralistic structure of information and the sources of that information are strained and the free marketplace of ideas which supports any plural political competition is weakened and deteriorated. As a result, the circuit of political responsibility and, ultimately, the very core of democratic logic is in danger." Casonato, above n 50, at 132.

87 Thus, machine intelligences "perform an algorithmic sorting and filtering that is designed primarily to capture the attention of an audience and to engage users, and not primarily to convey an accurate and informative picture of politics (Carlson, 2018, pp. 1765–1766). This is based on the capacities of machine learning to process information about individuals' preferences and behaviours in order to achieve personalized messages and information offers that optimally resonate with individual predispositions (Yeung, 2017a). Citizens may then increasingly be kept within the confines of what has notoriously been termed filter bubbles (Pariser, 2011)." König and Wenzelburger, above n 38, at 5.

88 Axel Bruns writes: "...the echo chamber and filter bubble metaphors represent at the very least a technologically determinist fallacy that is likely to have consequences for how current crises are addressed: as Meineck puts it, these metaphors are "the desperate attempt to make technology responsible for ... societal problems. Whoever speaks of filter bubbles evidently sees the causes of radical users in algorithmic newsfeeds or monstrous online platforms that push their helpless visitors in despicable ways into bubbles of opinion"– an unacceptable oversimplification that fundamentally, cynically deprives users of their personal identity and agency. Meineck therefore calls the filter bubble "the dumbest metaphor of the Internet" (ibid.), and this criticism is not unreasonable." Axel Bruns, "It's Not the Technology, Stupid: How the "Echo Chamber" and "Filter Bubble" Metaphors Have Failed Us", Paper presented at the IAMCR 2019 conference in Madrid, Spain, 7-11 July 2019. Submission no. 19771 at 9.

89 Hence, "beyond informal influence based on money, these corporations increasingly control the infrastructures of public discourse and the digital environment decisive for elections.... And their internet services increasingly become the only or main source of political information for citizens, especially the younger generation, to the detriment of the Fourth Estate, classic journalist publications, with the ambition to control power, so important to democracy. Their targeted advertising business model drains the Fourth Estate, journalism, concentrating today more than 80% of new advertisement revenue previously the basis for the plurality of privately financed press in the hands of just two companies, namely Google and Facebook." Nemitz, above n 41, at 3.

90 "Mass protests increased annually by an average of 11.5 percent from 2009 to 2019 across all regions of the world" with technology considered a key catalyzing factor. Samuel J. Brannen, Christian S. Haig, and Katherine Schmidt, *The Age of Mass Protests: Understanding an Escalating Global Trend*, CSIS Risk and Foresight Group (Center for Strategic & International Studies, 2020) at iv.

91 M. Myllylahti and G. Treadwell, *Trust in News in Aotearoa New Zealand 2023*, AUT Research Centre for Journalism, Media and Democracy (JMAD) (2023), https://www.jmadresearch.com/trust-in-news-in-new-zealand.

92 "The phenomenon of epistemic bubbles is also problematic with regard to epistemic agency. First, algorithms control my epistemic environment, not me. While selection of information is always necessary, here it is not me who does the selection but an algorithm (and perhaps others, humans, who use that algorithm). Second, if I am not exposed to different views, the quality of my belief formation is low." Coeckelbergh, above n 51, at 6.

93 König and Wenzelburger warn that "the rise of private media corporations and a trend of commercialization means that recipients are primarily addressed as consumers of content and less as citizens with an interest in developing political judgment (Hjarvard, 2013)." König and Wenzelburger, above n 38, at 7.

94 König and Wenzelburger highlight that: "The core principle of democratic rule demands that those affected by political decisions can understand themselves as their authors and that political decisions and outputs are tied back to citizens' views and preferences (Scharpf, 1999)....Two main features are essential to this translation process: electoral institutions, ensuring free and fair elections, and political parties, competing for government office (Dahl, 1998)." König and Wenzelburger, above n 38, at 2. "The use of Big Data and AI in digital media are often incongruent with fundamental democratic principles and human rights. The dominant paradigm is one of covert exploitation, erosion of individual agency and autonomy, and a sheer lack of transparency and accountability, reminiscent of authoritarian dynamics rather than of a digital well-being with equal and active participation of informed citizens." Christodoulou and Iordanou, above n 27, at 4.

95 Note also: "The implications deriving from the use of algorithms may have consequences on individuals' fundamental rights, such as the right to self-determination, freedom of expression, and privacy. However, fundamental rights do not exhaust the threats which these technologies raise for constitutional democracies." Pollicino and Gregorio, above n 2, at 4.

96 Patrick Diotte, "The Big Four and Cyber Espionage: How China, Russia, Iran and North Korea Spy Online," *Canadian Military Journal* 20, no. 4 (2020).

97 Wright notes that the Cold War era "did not afford authoritarian regimes so many opportunities for action within democracies. At home, Beijing, Moscow, and others have used twenty-first-century tools and tactics to reinvigorate censorship and manipulate the media and other independent institutions. Beyond their borders, they utilize educational and cultural initiatives, media outlets, think tanks, private sector initiatives, and other channels of engagement to influence the public sphere for their own purposes, refining their techniques along the way. Such actions increasingly shape intellectual inquiry and the integrity of the media space, as well as effect emerging technologies and the development of norms. Meanwhile, autocrats have utilized their largely hybrid state-capitalist systems to embed themselves in the commerce and economies." Manheim and Kaplan, above n 26, at i.

98 "Election meddling "was not a one-time event limited to the 2016 election. It's a daily drumbeat. These [fake accounts] are entities trying to disrupt our democratic process by pushing various forms of disinformation into the system." Manheim and Kaplan, above n 36, at 142.

99 “Influence campaigns, by the Russians and others, have matured to the point where they are overwhelming social media’s efforts to keep their platforms accountable. The polemics span the political spectrum with the goal of engendering online outrage and turning it into offline chaos.” Manheim and Kaplan, above n 36, at 142. And note, for instance: *“What Happened*: The FBI and intelligence services from other Western governments sabotaged a sophisticated malware known as Snake used by *Turla*, a hacking arm of Russia’s FSB intelligence service, Reuters reported on 9 May 2023. *Why It Matters*: *Turla* is one of the most sophisticated and persistent adversarial threat actors in cyberspace, and the disruption of one of its malwares is a significant success for Western law enforcement and cybersecurity efforts. The mitigation of *Turla’s* Snake may undermine Russia’s cyberespionage activities in the short term, although the threat actor is likely to pursue alternative tools to continue its operations in the long term. *Background*: Turla has been observed in cyberspace for more than 20 years, targeting hundreds of computers in at least 50 countries worldwide. The threat actor has primarily engaged in cyberespionage activities, targeting research, diplomatic and military organizations across NATO countries and the Commonwealth of Independent States.” “Western Coalition Disrupts Russian Cyberespionage Tool,” *Stratfor*, 9 May 2023, https://worldview.stratfor.com/situation-report/global-western-coalition-disrupts-russian-cyberespionage-tool. Or: “*What Happened:* One of the largest observed cyberespionage campaigns by Chinese threat actors dubbed “Volt Typhoon” targeted a broad array of critical infrastructure in Australia, Canada, New Zealand, the United Kingdom, the United States and the U.S. territory of Guam, according to a May 25 report by Microsoft. Microsoft also assessed with moderate confidence that Volt Typhoon was pursuing the development of capabilities that could disrupt critical communications infrastructure between the United States and the Asia region during future crises. “China, U.S.: Widespread Cyberespionage Campaign Targets U.S. Critical Infrastructure,” *Stratfor*, 25 May 2023, https://worldview.stratfor.com/situation-report/china-us-widespread-cyberespionage-campaign-targets-us-critical-infrastructure.

100 “Bots can spread truthful information as well as disinformation alike, depending on the (potentially manipulative) goal for which they are being used. Finally, social bots can be efficient in mobilising citizens for the so-called ‘astroturf’ campaigns. Astroturf campaigns give the impression to be grassroot campaigns run by non-profit organisations or citizens, whereas in reality their driving force are businesses or politicians who do not reveal their identity.” Brkhan, above n 33, at 67.

101 “Creating content to increase confusion and political divides can be labour-intensive at best for any human aiming to consistently churn it out, thus bots are employed to complete this task efficiently.” Benzie and Montasari, above n 55, at 11.

102 Google Duplex is one such AI system capable of conducting real time, fully automated phone calls on behalf of users. The voice technology and natural language capability used are so advanced that the responses it generates are almost indistinguishable from a human.

103 Brkhan, above n 33.

104 “There was no difference between engagement levels of tweets containing misinformation or those containing accurate and reputable information (Himelein-Wachowiak et al., 2021). This emphasises that once information has been posted online, regardless of accuracy, there is the same level of interaction with the information by human users.” Benzie and Montasari, above n 55, at 10.

105 Max Mason, John Davidson, and Ayesha de Kretser, “Inside Australia’s Most Invasive Data Hack,” *Australian Financial Review*, 09/11/2022.

106 C. Keall, “Ransomware Attacks: Privacy Commissioner Plans Investigation as Justice, Health Hit,” *NZ Herald*, 06/12/2022.

107 Intelligence and Security Committee, *Statement to Intelligence and Security Committee: Acting Director-General of the New Zealand Security Intelligence Service*, (27/03/2023).

108 This is because: “There is little oversight of AI development, leaving technology giants free to roam through our data and undermine our rights at will. We seem to find ourselves in a situation where Mark Zuckerberg and Sundar Pichai, CEOs of Facebook and Google, have more control over Americans’ lives and futures than do the representatives we elect. The power of these technology giants to act as “Emergent Transnational Sovereigns” stems in part from the ability of AI software (“West Coast Code”) to subvert or displace regulatory law (“East Coast Code”).” Manheim and Kaplan, above n 36, at 111. And note: “Technology firms are motivated to work in the political space for marketing, advertising revenue, and relationship-building in the service of lobbying efforts. To facilitate this, these firms have developed organizational structures and staffing patterns that accord with the partisan nature of American politics. Furthermore, Facebook, Twitter, and Google go beyond promoting their services and facilitating digital advertising buys, actively shaping campaign communication through their close collaboration with political staffers.” Daniel Kreiss and Shannon C. McGregor, “Technology Firms Shape Political Communication: The Work of Microsoft, Facebook, Twitter, and Google With Campaigns During the 2016 U.S. Presidential Cycle,” *Political communication* 35, no. 2 (2018), https://doi.org/10.1080/10584609.2017.1364814 at 155.

109 Thus: “The UN Guiding Principles on Business and Human Rights (UNGPs), adopted by the UN Human Rights Council in 2011, more substantially articulate the role and responsibilities of private-sector businesses in protecting human rights. Within the UNGP framework, the general legal obligation to protect human rights remains with states, while private firms have responsibilities to respect and protect human rights (and to remedy violations of them) when the firm’s own products, services, and operations are involved.” Donahoe and Metzger, above n 1, at 121.

110 “Massively enhanced informational basis in the age of “Big Data” leads to better-informed decisions by policy makers (e.g. Janssen & Kuk, 2016; van der Voort, Klievink, Arnaboldi, & Meijer, 2019), because algorithmic systems are able to systematize and structure huge amounts of data. For instance, applications of machine learning can be used for simulations of policy action and for continuously sourcing information from citizens that can then be used as the basis of policy-making (Chen & Hsieh, 2014; Williamson, 2014) and help to produce more evidence-based decisions (Lepri, Oliver, Letouzé, Pentland, & Vinck, 2018).” König and Wenzelburger, above n 38, at 5.

111 “As China and other authoritarian regimes construct digital authoritarian systems at home and propagate these models abroad, they are competing with democracies to shape global standards and infrastructure. How can liberal democracies harness the massive benefits of AI-related technologies without infringing on fundamental rights and risking a long-term shift toward authoritarianism? The challenge for democracies and democratic civil society is to build digitized systems that enable economic and social development but do not afford a shift to authoritarianism.” Wright, above n 35, at iii.

112 For example, “China broadly embraces artificial intelligence technology in order to track, monitor, and control its citizens. The Chinese government collects individuals’ big data from a variety of sources such as finance, tax, and health records. It monitors online purchasing behaviour, social media activity and information resulting from facial recognition via the country’s approximately 200 million surveillance cameras. This data is then used to calculate an individual’s “social credit score” which is supposed to give incentive to lawful behaviour and good citizenship. The score goes up for good behaviour such as donating blood, volunteering in a hospital, or repaying your loan on time. The opposite happens if you get a speeding ticket, if you fail to pay your taxes, or if you drop litter. Consequences of bad scores are non-eligibility for jobs in public administration, being turned down for loans, or even not being allowed to board an airplane. Good scores, in contrast, provide exemplary citizens with discounts on their electricity bills or a privileged access to health care (Marr, 2019).” Kaplan, above n 34, at 153.

113 “Four ways that (near) existing technology may be used to augment or enhance society’s democratic processes. In brief terms, these are: (1) The simple transfer of voting for representatives online; (2) The use of online voting to pass or reject bills proposed in the legislature; (3) The use of (anonymised) individuals’ preferences to directly inform legislative decision-making; and, (4) The wholesale replacement of the (physical) legislature and the individuals within it with a legislature composed of algorithms representing the voting public.” Paul Burgess, “Algorithmic Augmentation of Democracy: Considering Whether Technology can Enhance the Concepts of Democracy and the Rule of Law Through Four Hypotheticals,” *AI & Society* 37 (2022), https://doi.org/10.1007/s00146-021-01170-8 at 98.

114 “Advancing AI to ‘improve’ political decision-making risks to misconstrue the nature of democratic politics, which is not merely about optimizing outputs, but also about process and about regulating and containing unavoidable pluralism and conflicting views in society (Hildebrandt, 2016).” Konig, above n 6, at 2.

115 *Algorithm Assessment Report*, Stats NZ (Wellington: Stats NZ, 2018), https://data.govt.nz/use-data/analyse-data/ government-algorithm-transparency.

116 *Algorithm Charter for Aotearoa New Zealand Year 1 Review*, Taylor Fry (Taylor Fry, 2021), https://www.data.govt.nz/assets/data-ethics/algorithm/Algorithm-Charter-Year-1-Review-FINAL.pdf.

117 Decision making algorithms in AI systems are sometimes referred to as ‘Black Box’: “The literature on responsible AI has identified and continues to discuss, however, the unique role of epistemic challenges ensuing from the poor ‘traceability’ (Mittelstadt et al., 2016) and ‘explicability’ (Floridi et al., 2018) of ‘opaque’ (Burrell, 2016) AI systems….This can make it difficult even for AI developers themselves to forecast or reconstruct how data inputs are handled within such systems, how decisions are made, and how these decisions impact domains of application in the long term (Mittelstadt et al., 2016).” Alexander Buhmann and Christian Fieseler, “Deep Learning Meets Deep Democracy: Deliberative Governance and Responsible Innovation in Artificial Intelligence,” *Business Ethics Quarterly* (2022), https://doi.org/10.1017/beq.2021.42 at 2." <style face=”italic”>Business Ethics Quarterly</style> (2022 And witness the tragic debacle of the Australia’s Robodebt scandal, which saw suicides and led to a Royal Commission. *Royal Commission into the RoboDebt Scheme*, Department of the Prime Minister and Cabinet (2023), Canberra, https://robodebt.royalcommission.gov.au/publications/report.

118 “The challenges that will need to be faced during the new round of digital constitutionalism. Most notably, it is necessary to design a frame that describes the relationship between the three parties that Balkin puts at the heart of the information society: platforms, states, and individuals. In other words, a digital habeas corpus of substantive and procedural rights should be identified, which can be enforced by the courts as they are inferred from existing rights protected under current digital constitutionalism. Therefore, a new set of rights can be derived by such revisited understanding of individuals in the new digital context – among others, the right that decisions impacting the legal and political sphere of individuals are undertaken by human beings, and not exclusively by machines, even the most advanced and efficient ones.” Pollicino and Gregorio, above n 2, at 20.

119 “If a person finds AI acceptable in political decision making, it seems very likely that she has a reductionist view of politics as technical problem solving and sees AI as a means to obtain better solutions. Hence, AI may conflict with basic principles of liberal-democratic politics, but if citizens do not value these principles as a part of democracy, they may accept or embrace AI in political decision making, even in high-level politics, for example, to take decisions for politicians.” Konig, above n 6, at 6.

120 “You would not even need to issue precise queries to the citizenry on the relevant topics. A study by Wu, Kosinski, and Stillwell (2015) showed that computer-based judgments of one’s personality are far superior to those of human beings. With only 10 of your likes, Facebook’s algorithm will better predict your opinions than one of your colleagues. With 150 likes, it will do so more accurately than your own family. And with 300 likes, it will do so better than your spouse. This shows the future potential for AI-driven instant democracy.” Kaplan, above n 34, at 159.

121 “Bringing such technical problem-solving to political decision making to guide or even replace it, amounts to prioritizing effectiveness and efficiency and thus means placing output-based legitimacy over input-based legitimacy, that is, that political decisions reflect preferences of citizens (Scharpf, 1999). Greater reliance on AI could mean that citizens’ preferences as inputs to the democratic process become less relevant and that this is justified with results. Even more important consequences of AI for democracy concern the dimension of throughput legitimacy (Schmidt, 2013), because a greater reliance on AI, especially if conceded greater authority on higher levels of decision making, can change the form of the democratic political process and the way decisions are brought about.” Konig, above n 6, at 4.

122 Kaplan, above n 34, at 159.

123 However, “Algorithmic decision-making (ADM) systems and other forms of AI are geared towards information processing and thus solving cognitive and coordination tasks; and liberal democracy as a political system is marked by special informational requirements. Such requirements exist not only with regard to producing adequate outputs. Rather, due to standards of democratic legitimacy, informational needs arise also for safeguarding a certain form of decision-making (throughput level) as well as the continuous integration of citizen preferences (input level). This is important because liberal democracy is concerned with how outputs are produced.” König and Wenzelburger, above n 38, at 2.

124 This overreach of algorithms in governance raises several concerns: “The use of personal information for this purpose leads one to wonder whether individuals should have the right not to be subjected to a decision based solely on automated processing, including profiling which produces legal effects concerning him or her or similarly significantly affects him or her. These data subjects’ rights have been primarily analysed from the perspective of the right to explanation. Scholars have pointed out possible bases for the right to explanation such as those provisions mandating that data subjects receive meaningful information concerning the logic involved, as well as the significance, and the envisaged consequences of the processing.” Pollicino and Gregorio, above n 2, at 9.

125 König and Wenzelburger, above n 38, at 2. And: “Data colonialism works to dismantle the basic integrity of the self…. It is the self in this bare sense that must be salvaged from data colonialism, using all available legal, political, and philosophical means. None of the ideals desired in today’s societies—their democratic status, their freedom, their health, or otherwise—make any sense without reference to an autonomous self. Autonomy in [a] basic sense is the core value underlying any human social order that has the potential to be good. Yet, when all the complex detail of data practices has settled, it is this value that is being threatened today, not by technology itself but by the mutual implication of life and technology we call data colonialism. We can go further. Moving decisively away from individualistic notions of autonomy, freedom, and privacy, we can argue that the datafied social governance and incursions into the self’s minimal integrity constitute an environment that is toxic for human life.” Nick Couldry and Ulises A. Mejias, “Data Colonialism: Rethinking Big Data’s Relation to the Contemporary Subject,” *Television & new media* 20, no. 4 (2019), https://doi.org/10.1177/1527476418796632 at 347.

126 “Fake news and misinformation is not just a problem at the level of what knowledge citizens need for democracy (for example, one could argue that democracy needs truth), but is especially damaging at the procedural, how level, since it destroys trust in the socio-epistemic environment and in the end in the one’s own epistemic capacities: when fake news is ubiquitous, I can no longer believe others but I also can no longer believe my own eyes. …” Coeckelbergh, above n 51, at 3.

127 Benzie and Montasari, above n 55, at 5.

128 The loss of trust in society has serious implications for democracy: “Truth is a necessary, though not sufficient condition for democracy… Truth is the enemy of authoritarian and especially totalitarian regimes, which aim to control the epistemic environment in a way that support the establishment and maintenance of their power. [Cf. Ministry of Truth, George Orwell, 1984]” Coeckelbergh, above n 51, at 4.

129 Importantly, "[Another] effect of the abuse of AI on democracy is the frustration of citizens who at some point become non-voters and will no longer participate in political or democratic life. Due to no longer knowing what the truth is and what is fake, people might abandon participation completely. The more deceptions that occur and the more difficult the verification of content becomes, the more likely that people's trust in their institutions will continue to decrease." Kaplan, above n 34, at 155.

130 Data silos are one means of reducing the exposure of data: "Data can instead be shared between silos on a case-by-case basis with appropriate permissions—this is the difference between authoritarian mass surveillance and the limited surveillance necessary to combat crime or terrorism and perform other legitimate government functions. Siloing data also enhances security in the face of external threats. The disastrous Chinese hack that stole intimate data on about 22 million government employees, including those with security clearances, from the U.S. Office of Personnel Management illustrates the inherent risks of building a giant repository of valuable information." Wright, above n 35, at 6.

131 Lisa O'Carroll, "EU moves closer to passing one of world's first laws governing AI," *The Guardian*, 14 June 2023, https://www.theguardian.com/technology/2023/jun/14/eu-moves-closer-to-passing-one-of-worlds-first-laws-governing-ai (accessed 23 June 2023).

132 Dan Milmo and Kiran Stacey, "Rishi Sunak's AI Summit: what is its aim, and is it really necessary?" *The Guardian*, 9 June 2023, https://www.theguardian.com/technology/2023/jun/09/rishi-sunak-ai-summit-what-is-its-aim-and-is-it-really-necessary (accessed 21 June 2023).

MAXIM
INSTITUTE

PO Box 49 074, Roskill South, Auckland 1445, New Zealand / 49 Cape Horn Road, Hillsborough, Auckland / Ph +64 9 627 3261 / **www.maxim.org.nz**