GRIFFITH JOURNAL
OF LAW & HUMAN DIGNITY
Law and Human Dignity in the Technological Age

GRIFFITH JOURNAL OF LAW & HUMAN DIGNITY

Law & Human Dignity in the Technological Age

Special Issue

2019

Published in May 2019, Gold Coast, Australia by the *Griffith Journal of Law & Human Dignity*
ISSN: 2203-3114

We acknowledge and thank Zarek Hennessy for creating our Special Issue cover art. We extend our appreciation to Ashleigh Leake for her indispensable role in shaping the vision and content of the Special Issue in its early stages. We also extend thanks to Jessica Antal for her marketing assistance throughout the production of this Special Issue.

CONTENTS

VANESSA ANTAL	*EDITORIAL*	01
RAÚL MADDEN	*EQUITY, "REVENGE PORN," & CAMBRIDGE ANALYTICA: THE DOCTRINE OF CONFIDENCE AS A PROTECTION FOR HUMAN DIGNITY IN THE TECHNOLOGICAL AGE*	05
MARK BRADY	*IS AUSTRALIAN LAW ADAPTABLE TO AUTOMATED VEHICLES?*	35
PROFESSOR JULIAN WEBB	*INFORMATION TECHNOLOGY & THE FUTURE OF LEGAL EDUCATION: A PROVOCATION*	72
ZEINA ABU-MEITA & NICK INGLIS	*FINANCIAL EQUALITY, THE IGNORED HUMAN RIGHT: HOW E-CURRENCIES CAN LEVEL THE PLAYING FIELD*	105
LACHLAN ROBB	*THANATOPOLITICS THROUGH TECHNOPHOBIA: USING CHARLIE BROKER'S BLACK MIRROR TO REFLECT UPON HUMANITY IN THE FACE OF ADVANCED TECHNOLOGY*	143
DR DAVID TUFFLEY	*HUMAN INTELLIGENCE + ARTIFICIAL INTELLIGENCE = HUMAN POTENTIAL*	170
PAMELA FINCKENBERG-BROMAN, MORGAN BROMAN, & MARK BRADY	*LAW & TECHNOLOGY: THE LEGAL & SOCIAL IMPLICATIONS OF SENTIENT ROBOTS*	190

HUMAN INTELLIGENCE + ARTIFICIAL INTELLIGENCE = HUMAN POTENTIAL

DAVID TUFFLEY*[1]

Artificial Intelligence ("AI") in the twenty-first century is a powerfully disruptive technology, one whose influence in society is growing exponentially. It is a technology with the potential to bring enormous benefit, but also great harm, if not properly managed. How then may we reap the benefits of AI while ensuring we are not harmed by it? How do we frame the correct relationship with AI to ensure the primacy of human dignity as technology in general accelerates exponentially into the future? I assert that AI is neither good, nor bad in and of itself. It is simply a tool, an extension of human intelligence — not an externalised threat to be feared as presented in popular culture. Clearly, it is the strategic uses to which AI is put that determines its value. The potential abuses of AI — for example — in rogue autonomous weapons, are a manageable risk and should not place unreasonable restraint on its development when the potential benefits arguably much outweigh the harm.

*Dr David Tuffley is a Senior Lecturer in Applied Ethics and Sociotechnical Studies at the School of ICT at Griffith University. He is primarily concerned with the social impact of technology, with a special interest in Artificial Intelligence and the future of employment. David comes to the academic world from the IT industry where he spent the first half of his career at various large public and private sector organisations in Australia and the United Kingdom.

CONTENTS

I	AN AGE OF DISSOLVING BOUNDARIES	03
II	PERVASIVE ARTIFICIAL INTELLIGENCE	04
III	ARTIFICIAL INTELLIGENCE IN THE WORKPLACE	05
IV	HOW IS AI HELPING LAWYERS?	06
V	PRINCIPLES FOR ETHICAL TECHNOLOGY USE	06
VI	A NEW RELATIONSHIP	07
VII	POPULAR CULTURE GENERATES A CLIMATE OF FEAR	08
VIII	TECHNOLOGY AS A FORCE FOR GOOD	10
IX	THE RISKS OF ARTIFICIAL INTELLIGENCE	11
X	PRACTICAL GUIDELINES FOR DEVELOPERS	13
XI	SUMMONING THE DEMON	15

I AN AGE OF DISSOLVING BOUNDARIES

Commercial computing in the modern world is focussed on making sophisticated but easy-to-use products that enhance the lifestyle of the user. Surprisingly enough, some of this technology comes at little or no cost. Devices such as GPS and video-conferencing are now software that is included in the purchase price of smartphones. Over a decade ago, you would have paid exorbitant prices to acquire the same physical products.[2] Ultra-high definition video cameras ten years ago were only used by TV studios at a cost of hundreds of thousands of dollars.[3] Today, they are software bundled on mid-range smartphones.

So integrated into our lives has such technology become, that it is sometimes difficult to know where the user ends and the technology begins. The technology has become a functional extension of our human selves. It is not uncommon for people to look at their smartphone a hundred times a day.[4] To some, it is the last thing they do at bedtime, and the first thing they do upon waking. With this degree of dependence, a person is already a functional cyborg: a blend of humanity and technology. One does not need to have the technology implanted internally to be a cyborg: it is enough that a person extends their mind into it.[5]

Smartphones first made their appearance in June 2007 when Steve Jobs brought out the first iPhone.[6] This first-generation iPhone was revolutionary because it had an internet browser, a music player and a telephone all integrated into a user-friendly package.[7] It set a new standard of design that quickly dispensed with the competition, leading to the establishment of a new paradigm of mobile computing. It has contributed to Apple becoming one of the highest valued commercial organisations on the planet.[8] The success

[2] Jenna Wortham, 'Sending GPS Devices the Way of the Tape Deck?', *New York Times* (online, 7 July 2009) <https://www.nytimes.com/2009/07/08/technology/08gps.html>.
[3] Lee Frederiksen, 'What Is the Cost of Video Production for the Web?', *Hinge* (online, 15 October 2018) <https://hingemarketing.com/blog/story/what-is-the-cost-of-video-production-for-the-web>.
[4] Stephen Willard, 'Study: People Check Their Cell Phones Every Six Minutes, 150 Times A Day', *Elite Daily* (online, 11 April 2019) <https://www.elitedaily.com/news/world/study-people-check-cell-phones-minutes-150-times-day>.
[5] See, eg, Andy Clark and David Chalmers, 'The Extended Mind' (1998) 58(1) *Analysis* 7.
[6] Charles Arthur, 'The history of smartphones: timeline', *The Guardian* (Online, 25 January 2012) <https://www.theguardian.com/technology/2012/jan/24/smartphones-timeline>.
[7] Ibid.
[8] Rob Davies, 'Apple becomes world's first trillion-dollar company', *The Guardian* (online, 3 August 2018 2019) <https://www.theguardian.com/technology/2018/aug/02/apple-becomes-worlds-first-trillion-dollar-company>.

of the smartphone in general, is an indication of the degree to which people value being able to intuitively extend their capabilities to accomplish what they could not do alone.

II PERVASIVE ARTIFICIAL INTELLIGENCE

AI is a recognised technology in its own right, but perhaps more importantly, it is an *enabler* of other technologies, many of which may not have even been invented yet. AI today has been compared with electricity a hundred years ago.[9] In earlier periods, machines that were operated mechanically or manually were enhanced by the addition of an electric motor.[10] Washing machines, refrigerators, and a variety of other devices came into existence because of the enabling technology of electricity. This trend progressed over time creating new products and a plethora of new jobs.[11]

Likewise, with AI today, when added to previously unconnected devices, creates the "internet of things" in everyday objects through an embedded computer chip that communicates with the internet, otherwise known as "the cloud". From this connection, AI can be delivered on-demand from various cloud services at little or no cost. Adding AI to previously unconnected devices will foreseeably create new industries and a multitude of jobs, following the same pattern as electricity.[12] As American humourist Mark Twain reputedly observed, '[h]istory doesn't repeat itself but it often rhymes'.[13]

AI is already integral to many of the functions in smartphones. Google's *Assistant*, Apple's *Siri*, Microsoft's *Cortana* and Amazon's *Alexa* are all early generation AI-enabled digital assistants designed to help people organise their everyday lives while coordinating various background applications not visible to the user. It allows people to communicate with their computer in natural language. It may not dispense with the keyboard and mouse altogether, but the inherent user-friendliness of natural language means it will likely become the preferred way to communicate with technology in the future.

---

[9] Kevin Kelly, *The Inevitable: Understanding the 12 Technological Forces that will Shape our Future* (Penguin Books, 2016) 33.
[10] Ibid.
[11] Ibid.
[12] Ibid.
[13] Charles C Doyle et al, *Dictionary of Modern Proverbs* (Yale University Press, 2012) 121.

III ARTIFICIAL INTELLIGENCE IN THE WORKPLACE

Examples of AI in everyday life have focussed on smartphones and consumer-level products, since these represent a growing, now pervasive class of technology. But what of the impact of AI in the workplace? Many people fear their jobs will be automated while their children will be hard-pressed to find employment.[14] Futurists augment said fears by implying that if a job *can* be automated then it *will* be automated.[15] However, this is far from true. Many jobs will continue to be done by people because the human touch is preferable.[16] Or perhaps creativity is called for.[17] Or maybe because it is politically dangerous for governments to put bread-winners out of a job.[18]

It will be decades before AI can replace humans in all of our idiosyncratic complexity. As for those jobs that are at risk, they are likely to be the 'dull, dirty or dangerous jobs', the ones that employers already have difficulty filling;[19] jobs like sewer reconnaissance, underground mining and repetitive factory work. Manual welding, for example, can produce highly toxic fumes – a prime candidate for automation.[20]

AI in the workplace can greatly improve job performance. In Japan, the first recorded case of AI saving someone's life was recently seen when a woman with a rare form of leukaemia who was initially misdiagnosed by a team of human doctors.[21] A diagnostic AI was put to work and in under 20 minutes had analysed the woman's genome, compared it with 20 plus million oncological studies, arrived at the correct diagnosis and recommended a treatment regime which was subsequently proved correct. It was the *combination* of human doctors and a diagnostic AI that succeeded where the human doctors alone had failed. When AI is used as an extension of human intelligence, the

[14] David Tuffley, 'In 10 years, your job might not exist. Here's how to make sure you're still employable', *The Washington Post* (online, 5 January 2015) <https://www.washingtonpost.com/posteverything/wp/2015/01/05/in-10-years-the-job-market-will-look-totally-different-heres-how-to-make-sure-youre-ready/?utm_term=.78224bb81dab>.
[15] David Tuffley, 'We Should Learn to Work with Robots and not Worry about them Taking our Jobs', *The Conversation* (Web Page, 19 February 2018) <https://theconversation.com/we-should-learn-to-work-with-robots-and-not-worry-about-them-taking-our-jobs-91004>.
[16] Ibid.
[17] Ibid.
[18] Ibid.
[19] Ibid.
[20] Ibid.
[21] David Tuffley, 'How can Doctors use Technology to Help them Diagnose?', *The Conversation* (Web Page, 25 October 2016) <https://theconversation.com/how-can-doctors-use-technology-to-help-them-diagnose-64555>.

partnership can be a powerful one indeed. It is this partnership and enhanced personal capability that should be focussed on when considering the future of AI.

IV HOW IS AI HELPING LAWYERS?

Large law firms have junior associates whose unenviable job is to perform document discovery.[22] The greater number of associates engaged in discovery likely gives the larger firms a strategic advantage over smaller firms with fewer associates. But that advantage is diminishing as the smaller firms make greater use of AI-enabled data analytics to do the routine discovery work of many associates in a fraction of the time.[23]

Likewise, courtroom analytic tools already offer advanced data search capabilities to highlight relevant rulings, judge preferences, multiple motion types and so on.[24] This technology is a force multiplier of a lawyer's ability by emulating a human lawyer with the qualities of meticulous research, deep understanding of case law, and the ability to mount a solid argument.

V PRINCIPLES FOR ETHICAL TECHNOLOGY USE

Is there a set of universally applicable rules for ethical technology use? I propose the following principles that draw upon the earlier work of philosopher Immanuel Kant, whose ideas continue to exert a strong influence on ethics today.

These simply stated principles are general enough to work in both the virtual world and in the physical world. Their simplicity gives them universality while making their intention transparent.

I propose that Kant's categorical imperatives be adapted for ethical technology development and use. Categorical imperatives are unconditional requirements that are always true:

---

[22] See, eg, Natasha Gillezeau and Elouise Fowler, 'What it's like working as a young corporate lawyer at a top tier firm', *Financial Review* (online, 25 January 2019) <https://www.afr.com/leadership/workplace/what-it-is-like-working-as-a-young-corporate-lawyer-at-a-top-tier-firm-20181212-h190vt>.

[23] See eg, Katie Walsh, 'The $1m web service showing law firms are embracing the AI beast', *Financial Review* (online, 1 March 2018) <https://www.afr.com/business/legal/the-1m-web-service-showing-law-firms-are-embracing-the-ai-beast-20180225-h0wme4>.

[24] Ibid.

1. Before I do something with this technology, I ask myself: would it be acceptable if everyone did it?
2. Will this technology harm, diminish or dehumanise anyone, including people I don't know and will never meet?
3. Do I have the informed consent of those who will be affected?[25]

If the answer to any of these questions is 'no', then it is arguably unethical to use technology in that manner. These are rational principles that hold true in both the virtual world and physical world, applying the same standards to both, maintaining a consistent stance in an increasingly blurred boundary between virtual and physical worlds.

VI A NEW RELATIONSHIP

The risks of AI should not *per se* prohibit or limit the use of this potentially beneficial technology. The correct perspective is to see AI as a capability-extending adjunct to human intelligence, allowing us to do things that could not be done unaided. We already make extensive use of various technologies in this way. For example, a human walking on their own two feet might travel eight kilometres — in a car they might drive 800 kilometres in the same time; on a plane they could travel 8,000 kilometres. The driver of the car or pilot of the plane is still responsible for the actions of their vehicle. Arguably a simple and transparent principle could be adopted: that laws applying to the unaugmented human might be extended to include the augmented human on the basis that the augmentation is not independent of the person, but an extension of them.

The underlying computer code that tells AI how to behave is the product of human endeavour. While it is possible to instruct AI to write its own code, it would be an act of unbelievable folly to create a situation where the AI is allowed to develop malicious tendencies, much less allow the AI to express malice by creating harm in the world. It is questionable that AI is capable of malice, independent of its human programming.[26] Regardless of this, it must be made illegal to program an AI to do harm either through

[25] Immanuel Kant, *Fundamental Principles of the Metaphysic of Morals* (eBooks@Adelaide, 2014) [Tr Thomas Kingsmill Abbott].
[26] See, eg, Aatif Sulleyman, 'Stephen Hawking warns Artificial Intelligence 'may replace humans altogether'', *The Independent* (Web Page, 2 November 2017) <https://www.independent.co.uk/life-style/gadgets-and-tech/news/stephen-hawking-artificial-intelligence-fears-ai-will-replace-humans-virus-life-a8034341.html>.

negligent development practices or intentional design with the possible exception of narrowly constrained military applications.

VII POPULAR CULTURE GENERATES A CLIMATE OF FEAR

Popular culture and in particular, the Science Fiction ("SF")genre have a long history of anxiety-producing visions of the future that involve existentially threatening relationships between humans and technology. These anxieties are nothing new. They have been manifest since the technological advances of the Industrial Revolution, one consequence of which was the emergence of a new literary genre — SF. HG Well's classic *War of the Worlds* is an example of anxiety about technology being externalised as a perceived existential threat.[27]

Victorian science fiction writers have proved influential in the modern world. In the mid-twentieth century, there is clear evidence that the work of Jules Verne influenced the science and technology of the period. Verne's *Twenty Thousand Leagues Under the Sea*[28] features a submarine *Nautilus* — it is no coincidence that the US Navy's first nuclear submarine which was capable of staying submerged for weeks was called *Nautilus*.[29] Likewise, the US moon shot in 1969[30] was shaped by the narrative of Verne's *From the Earth to the Moon*.[31] Former American president John F Kennedy gave voice to Verne's vision in his famous speech, '[w]e choose to go to the Moon in this decade and do the other things, not because they are easy, but because they are hard'.[32]

More recently, Phillip K Dick wrote *Do Androids Dream of Electric Sheep*[33] that still resonates today in the public imagination if the success of the *Bladerunner* series[34] is an

[27] H G Wells, *War of the Worlds* (Penguin Books Ltd, 2011).
[28] Jules Verne, *Twenty thousand leagues under the sea* (Oxford University Press, 1998).
[29] Kyle Mizokami, 'Meet USS Nautilus: America's First Nuclear Powered Submarine Was A Game-Changer', *National Interest* (online, 7 April 2019) <https://nationalinterest.org/blog/buzz/meet-uss-nautilus-americas-first-nuclear-powered-submarine-was-game-changer-51347>.
[30] NASA, 'July 20, 1969: One Giant Leap for Mankind', *NASA* (online, 20 July 2017) <https://www.nasa.gov/mission_pages/apollo/apollo11.html>.
[31] Jules Verne, *From the Earth to the Moon* (Bantam Doubleday Dell Publishing Group Inc, 1996).
[32] John F Kennedy, 'Moon Speech' (Speech, Rice University, 12 September 1962).
[33] Phillip K Dick, *Do Androids Dream of Electric Sheep?* (Simon and Schuster, 2014).
[34] Scott Bukatman, *Blade runner* (Bloomsbury Publishing, 2017).

indication. Recent television shows like *Black Mirror*[35] and *Westworld*[36] as well as films like *Terminator*[37] have a strong dystopian spin. It can be suggested that the viewing public has an apparently insatiable appetite for scaremongering.

In contrast to this deterministic if not aspirational view of SF, is the Neo-Marxist perspective of Darko Suvin whose ideas became influential in the 1970's and are still influential today.[38] In his *Metamorphoses of Science Fiction*, Suvin reduces SF to an assembly of stories characterised by 'cognition and estrangement'.[39] Cognitive estrangement is experienced by the reader as they try to reconcile the fictional world with the world in which they live. Moreover, the SF genre is populist in nature — a low-quality "paraliterature" expressive of periods of rapid technological change. The technology of the fictional world are mere props to make the story more interesting. Suvin's perspective is evident in the popular SF culture of mass consumption of cinema.[40] As such, it would be inaccurate as well as unjust to describe as "paraliterature", the works of writers like Arthur C Clark, Isaac Asimov, Ursula LeGuin, Frank Herbert, Phillip Dick, Robert Heinlein, Douglas Adams, Kurt Vonnegut and others.

The anxiety-producing view of technology predominates in the popular genre despite there being innumerable instances of AI safely operating in the world today as digital assistants on consumer level electronics and more specialised AI operating in domains such as transport, finance, and medicine.[41]

It is true that all technologies embody *some* degree of risk. Nuclear technology, though it poses a greater existential risk than AI, has given the world bountiful, inexpensive energy to meet the needs of communities.[42] And yet it possesses the destructive potential seen

---

[35] Netflix, 'Black Mirror', *Netflix* (Television Show, 4 December 2011) <https://www.netflix.com/au/title/70264888>.
[36] IMDB, 'Westworld', *Westworld* (Web Page, 2 October 2016) <https://www.imdb.com/title/tt0475784/>.
[37] IMDB, 'Terminator', *Terminator* (Web Page, 20 December 1984) <https://www.imdb.com/title/tt0088247/?ref_=fn_al_tt_1>.
[38] See, eg, Darko Suvin, 'On the Poetics of the Science Fiction Genre' (1972) 34(3) *College English* 372.
[39] Darko Suvin, *Metamorphoses of Science Fiction* (Peter Lang AG, 1979) viii.
[40] Ibid ix.
[41] See, eg, Victorian All-Party Parliamentary Group on Artificial Intelligence, 'Artificial Intelligence Primer', *Parliament of Victoria*, (Web Page, March 2018) <https://www.parliament.vic.gov.au/images/stories/AI-Primer_Feb2018.pdf>.
[42] Seán Ó hÉigeartaigh, 'Technological Wild Cards: Existential Risk and a Changing Humanity', *OpenMind* (Online, 2017) <https://www.bbvaopenmind.com/en/articles/technological-wild-cards-existential-risk-and-a-changing-humanity>.

at Hiroshima which held the world ransom for the duration of the Cold War.[43] Medical science is busy extending people's life but it also creates the potential for biological weapons.[44] It can be said that one cannot have the potential for benefit without the potential for harm.

Considering the above then, the question is not *should we use this technology?* but, *how may we use it for the greater good?*

From a utilitarian perspective, if a million lives are saved for every one lost, the response should not be to ban the technology but to manage the risk to a lower level. Banning a technology because it poses a manageable risk is unreasonable. There are many examples, from transport to medical science, of beneficial technologies that are occasionally dangerous but which have proven safe for widespread use.[45] Humanity has been doing this since we discovered the utility of fire — indeed, there is the risk that careless use could burn down a village, but the benefits of cooked food, heat, light, protection from predators, and the strengthening of social bonds, has made it nonetheless worthwhile.

VIII TECHNOLOGY AS A FORCE FOR GOOD

If developed and used according to ethical principles, technology in general has the potential to help individuals to approach a fuller expression of their human potential. This is tied to what psychologist Abraham Maslow, best known for creating Maslow's hierarchy of needs, called 'self-actualisation'.[46] AI responsibly governed can be seen as a catalyst for self-actualising individuals.

Imagine Mozart in a world before the technology of the piano had been invented; Van Gogh in a world before inexpensive oil paints or George Lucas before the technologies of film. Today, there are potentially millions of children being born for whom their

[43] Ibid.
[44] Jeanne Guillemin, 'Scientists and the history of biological weapons: A brief historical overview of the development of biological weapons in the twentieth century', *US National Library of Medicine National Institutes of Health* (online, July 2006) <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1490304/>.
[45] Stephen H Unger, 'Is Progress in Technology Always Beneficial?', *Columbia University* (online, 26 May 2014) <http://www1.cs.columbia.edu/~unger/articles/technologyProgress.html>.
[46] Ann Olson, 'The Theory of Self-Actualization', *Psychology Today* (online, 13 August 2013) <https://www.psychologytoday.com/au/blog/theory-and-psychopathology/201308/the-theory-self-actualization>.

technology of self-expression has not yet been invented. The critical issue is to institutionalise principles for ethical technology use in the AI development community to promote life-affirming uses.

IX THE RISKS OF ARTIFICIAL INTELLIGENCE

The risks of AI can be accommodated in three broad areas: programming errors, cyber-attacks, and taking instructions too literally. None of these needs to necessarily be a problem if properly managed. Software Engineering ("SE") has evolved model-based development practices that when followed reduce programming errors in safety critical applications to a very low level. SE is a mature discipline that today can guarantee quality in the same way that a civil engineer, for example, can guarantee that a bridge will not fall down if the defined process is followed.[47] Likewise, developers of the AI software that are flying the planes we travel on can be relied upon to get us safely to our destination.[48]

In relation to cybersecurity we are certainly not defenceless against those who would subvert the orderly running of software systems. The fact that our complex IT-enabled world continues to operate with an acceptable degree of reliability, if not efficiency, is testament to the efforts of cybersecurity researchers, and practitioners. One emerging security method is proving effective in critical areas: the "blockchain" principle.[49] When you have the same record replicated on multiple disparate computers, the attack is only successful if *every* instance of that record is changed ahead of the next cycle — an impossible feat in the short time frames allowable. There is no room for complacency or cutting-corners, but it is clearly feasible that with proper safeguards, an AI system can be protected from attack by using, for example, a blockchain security device.[50] Our complex, computer-driven modern world could not function as it does every day unless the issue of cybersecurity is not being successfully managed.

[47] Walt Scacchi, 'Process models in software engineering' in John J Marciniak (ed), *Encyclopaedia of Software Engineering* (Wiley, 2nd ed, 2002) 3.
[48] See, eg, Jack Stewart, 'Don't freak over Boeing's self-flying plane—robots already run the skies', *Wired*, (online, 6 September 2017) <https://www.wired.com/story/boeing-autonomous-plane-autopilot/>.
[49] See, eg, Omri Barzilay, '3 Ways Blockchain is revolutionizing cybersecurity', *Forbes* (online, 21 August 2017) <https://www.forbes.com/sites/omribarzilay/2017/08/21/3-ways-blockchain-is-revolutionizing-cybersecurity/#754418f22334>.
[50] Ibid.

The risk of taking instructions too literally can be likewise managed with the same safeguards applied to every safety critical system. One example is the scenario of a furniture factory under the control of an AI that has been instructed to not only make furniture, but also to optimise the process for maximum efficiency. The AI decides that because the human workers are less efficient than robot workers, the humans should be removed from the process. Sometime later the AI decides by extension that all humans should be eliminated for the purpose of achieving greater overall efficiency. Ultimately the planet becomes one big furniture factory when all humans have been exterminated. It is an interesting thought experiment that should not be taken literally. Those who set up the factory and programmed the AI would need to be both highly intelligent yet abjectly and wilfully ignorant of the inherent problems.

More realistically perhaps, consider the scenario of a hospital that has an expert AI in charge of maintaining optimal treatment of intensive care patients. It is unlikely, in light of the reality of law suits and the 24 hour news cycle, that the hospital would allow the AI to make life and death decisions in relation to that patient without involving the medical staff.

Data accuracy is another issue. AI can extract useful information from very large data sets using "predictive algorithms" to identify trends that can predict future trends with some degree of accuracy. When combined with other predictive methods, the reliability can be improved to a usable level.

I maintain that while AI-related risks do exist and must be taken seriously, they are nonetheless manageable as demonstrated by our ability to not destroy ourselves, since the time in which we learned to use fire, and more recently with an increasingly lethal arsenal of weapons.[51] Given the social and economic benefits of AI generally, we must not be afraid to explore the possibilities of AI helpers in a supportive role, with humans always in overall control.

[51] Tuffley (n 15).

X PRACTICAL GUIDELINES FOR AI DEVELOPERS

A set of high level ethical principles is proposed to guide developers of AI and other technologies:

1. Before I do something with this technology, I ask myself: would it be acceptable if everyone did it?
2. Will this technology harm, diminish, or dehumanise anyone, including people I don't know and will never meet?
3. Do I have the informed consent of those who will be affected?

But how should these principles be implemented and what practical form would they take? We need a set of universal design principles to produce AI that poses little or no threat to humanity. Being universal, the guidelines should represent the consensus view of all stakeholders. The guidelines are expected to evolve over time to negotiate perspectives that differ from the majority. It is clearly a bottom-up approach as distinct from a top-down, rule by decree approach.

With around 420,000 members in 160 countries, the Institute of Electrical and Electronics Engineers ("IEEE"), is well-placed to produce a consensus-driven set of design principles. IEEE is the largest association of technical professionals in the world.[52] After consultation over an extended period with members and interested others, the *IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems* has been produced.[53] Available under Creative Commons Attribution-Non-Commercial license, it represents a consensus view of the Artificial Intelligence and Autonomous Systems (AI/AS) communities globally. The principles are a work in progress and will no doubt evolve over time to better reflect the views of those who do not fully endorse the current principles.

Eight principles relating to AI/AS (Automated Systems) are thus set forth that outline the proposed moral DNA of Artificial Intelligence, as summarised below:

[52] Institute of Electrical & Electronic Engineers, 'About IEEE', *Institute of Electrical & Electronic Engineers* (Web Page, 2019) <https://www.ieee.org/>.
[53] Sarah Mattingly-Jordan, 'The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems' (2017) *Institute of Electrical & Electronic Engineers* 1.

1. **General Principles**: AI/AS that embodies the highest ideals of human rights, prioritises maximum benefit to humanity and the natural environment, and mitigates negative impacts as AI/AS evolve into socio-technical systems. Essentially, AI/AS develops into a benign extension of human intelligence.
2. **Embedding Values into Autonomous Intelligence Systems**: Identify the norms and values of a specific community affected by AI/AS, implement the norms and values of that community within AI/AS and, evaluate and correct the compatibility of those norms and values between the humans and AI/AS within that community.
3. **Methodologies to Guide Ethical Research and Design**: The modern AI/AS organisation should ensure that human wellbeing, empowerment, and freedom are at the core of AI/AS development.
4. **Safety and Beneficence of Artificial General Intelligence (AGI) and Artificial Superintelligence (ASI)**: Future highly capable AI systems are likely to have a transformative effect on the world on the scale of the agricultural or industrial revolutions which could bring about unprecedented levels of global prosperity.
5. **Personal Data and Individual Access Control**: A key ethical dilemma regarding personal information is data asymmetry. It is necessary to define, access, and manage personal data to guarantee a person's unique identity.
6. **Reframing Autonomous Weapons Systems**: Autonomous systems that are designed to cause physical harm have additional ethical ramifications as compared to both traditional weapons and autonomous systems that are not designed to cause harm.
7. **Economics/Humanitarian Issues**: Careful attention to the range of technologies, methodologies, and systems that reduce the need for human intervention in our day-to-day lives.
8. **Law**: The early development of AI/AS has given rise to many complex ethical problems. These ethical issues almost always directly translate into concrete legal challenges. Some of them create difficult collateral legal problems such as liability for accidents involving autonomous vehicles. [54]

[54] Ibid.

XI SUMMONING THE DEMON

If one is to take an even-handed approach to the question of the future of AI, the views of advocates like Elon Musk, Bill Gates, and Stephen Hawking must be taken seriously when they warn of the existential risk of AI. Musk famously compared AI research and development with "summoning the demon".[55] What are we to make of this incongruous rhetoric? One interpretation is that it is an attempt to scare the public, and in so doing, puts pressure on governments to legislate stricter controls over the future of AI development. As Musk pointed out more reasonably in a later interview, he has had to negotiate a labyrinth of government regulations constraining the operation of autonomous vehicles and of aerial vehicles, such as his SpaceX rockets. He poses the question of whether the public would be happy to see other forms of AI implementation not similarly regulated in the public interest.[56]

A prosperous future with improved quality of life depends on us coming to terms with the challenges of AI. What is particularly important is for us to pay attention to the dynamic tension that is generated as we make the transition from a human to a "post human" society.[57] As Bostrom observes, an extinction event is certain to occur eventually, either through seismic catastrophe, asteroid impact or global contagion.[58] It is technology or AI in particular that represents our best chance to save ourselves.

[55] Matt McFarland, 'Elon Musk: With Artificial Intelligence we are Summoning the Demon', *Washington Post* (Web Page, 24 October 2014) <https://www.washingtonpost.com/news/innovations/wp/2014/10/24/elon-musk-with-artificial-intelligence-we-are-summoning-the-demon/>.
[56] Jamie Condliffe, 'Elon Musk Urges U.S. Governors to Regulate AI Before "It's Too Late"', *MIT Technology Review* (online, 17 July 2017) <https://www.technologyreview.com/s/608296/elon-musk-urges-us-governors-to-regulate-ai-before-its-too-late/>.
[57] Nick Bostrom, 'Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards' (2002) 9(1) *Journal of Evolution and Technology* 1, 58.
[58] Ibid.

REFERENCE LIST

A *Articles/Books/Reports*

Arthur, Charles, 'The history of smartphones: timeline', *The Guardian* (online, 25 January 2012) <https://www.theguardian.com/technology/2012/jan/24/smartphones-timeline>

Barzilay, Omri, '3 Ways Blockchain is revolutionizing cybersecurity', *Forbes* (online, 21 August 2017) <https://www.forbes.com/sites/omribarzilay/2017/08/21/3-ways-blockchain-is-revolutionizing-cybersecurity/#754418f22334>

Bostrom, Nick, 'Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards' (2002) 9 *Journal of Evolution and Technology*

Bukatman, Scott*, Blade runner* (Bloomsbury Publishing, 2017)

Clark, Andy and David Chalmers, 'The Extended Mind' (1998) 58(1) *Analysis* 7

Condliffe, Jamie, 'Elon Musk Urges U.S. Governors to Regulate AI Before "It's Too Late"', *MIT Technology Review* (online, 17 July 2017) <https://www.technologyreview.com/s/608296/elon-musk-urges-us-governors-to-regulate-ai-before-its-too-late/>

Davies, Rob, 'Apple becomes world's first trillion-dollar company', *The Guardian* (online, 3 August 2018) <https://www.theguardian.com/technology/2018/aug/02/apple-becomes-worlds-first-trillion-dollar-company>

Dick, Phillip, K, *Do Androids Dream of Electric Sheep?* (Simon and Schuster, 2014)

Doyle, Charles C Mieder, Wolfgang, Shapiro, Fred R, *Dictionary of Modern Proverbs* (Yale University Press, 2012)

Frederiksen, Lee, 'What Is the Cost of Video Production for the Web?', *Hinge* (online, 15 October 2018) <https://hingemarketing.com/blog/story/what-is-the-cost-of-video-production-for-the-web>

Gillezeau, Natasha and Elouise Fowler, 'What it's like working as a young corporate lawyer at a top tier firm', *Financial Review* (online, 25 January 2019)

<https://www.afr.com/leadership/workplace/what-it-is-like-working-as-a-young-corporate-lawyer-at-a-top-tier-firm-20181212-h190vt>

Guillemin, Jeanne, 'Scientists and the history of biological weapons: A brief historical overview of the development of biological weapons in the twentieth century', *US National Library of Medicine National Institutes of Health* (online, July 2006) <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1490304/>

hÉigeartaigh, Seán Ó, 'Technological Wild Cards: Existential Risk and a Changing Humanity', *OpenMind* (Online, 2017) <https://www.bbvaopenmind.com/en/articles/technological-wild-cards-existential-risk-and-a-changing-humanity>

Kant, Immanuel, *Fundamental Principles of the Metaphysic of Morals* (eBooks@Adelaide, 2014) [Tr Thomas Kings,mill Abbott]

Kelly, Kevin, *The Inevitable: Understanding the 12 technological forces that will shape our future* (Penguin, 2016)

Kennedy, John F, 'Moon Speech' (Speech, Rice University, 12 September 1962)

Mattingly-Jordan, Sarah, 'The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems' (2017) *Institute of Electrical & Electronic Engineers* 1

McFarland, Matt, 'Elon Musk: With Artificial Intelligence we are Summoning the Demon', *The Washington Post* (online, 24 October 2014) <https://www.washingtonpost.com/news/innovations/wp/2014/10/24/elon-musk-with-artificial-intelligence-we-are-summoning-the-demon/>

Mizokami, Kyle, 'Meet USS Nautilus: America's First Nuclear Powered Submarine Was A Game-Changer', *National Interest* (online, 7 April 2019) <https://nationalinterest.org/blog/buzz/meet-uss-nautilus-americas-first-nuclear-powered-submarine-was-game-changer-51347>

NASA, 'July 20, 1969: One Giant Leap for Mankind', NASA (online, 20 July 2017) <https://www.nasa.gov/mission_pages/apollo/apollo11.html>

Olson, Ann, 'The Theory of Self-Actualization', *Psychology Today* (online, 13 August 2013) <https://www.psychologytoday.com/au/blog/theory-and-psychopathology/201308/the-theory-self-actualization>

Scacchi, Walt, 'Process models in software engineering' in John J Marciniak (ed), Encyclopaedia of Software Engineering (Wiley, 2nd ed, 2002)

Stewart, Jack, 'Don't freak over Boeing's self-flying plane—robots already run the skies', Wired, (online, 6 September 2017) <https://www.wired.com/story/boeing-autonomous-plane-autopilot/>

Sulleyman, Aatif, 'Stephen Hawking warns Artificial Intelligence 'may replace humans altogether'', *The Independent* (Web Page, 2 November 2017) <https://www.independent.co.uk/life-style/gadgets-and-tech/news/stephen-hawking-artificial-intelligence-fears-ai-will-replace-humans-virus-life-a8034341.html>

Suvin, Darko, *Metamorphoses of Science Fiction* (Peter Lang AG, 1979)

Suvin, Darko, 'On the Poetics of the Science Fiction Genre' (1972) 34(3) *College English* 372

Tuffley, David, 'How can Doctors use Technology to Help them Diagnose?' *The Conversation* (online, 25 October 2016) <https://theconversation.com/how-can-doctors-use-technology-to-help-them-diagnose-64555>

Tuffley, David, 'In 10 years, your job might not exist. Here's how to make sure you're still employable', *The Washington Post* (online, 5 January 2015) <https://www.washingtonpost.com/posteverything/wp/2015/01/05/in-10-years-the-job-market-will-look-totally-different-heres-how-to-make-sure-youre-ready/?utm_term=.78224bb81dab>

Tuffley, David, 'We Should Learn to Work with Robots and not Worry about them Taking our Jobs', *The Conversation* (online, 19 February 2018) <https://theconversation.com/we-should-learn-to-work-with-robots-and-not-worry-about-them-taking-our-jobs-91004>

Unger, Stephen, H, 'Is Progress in Technology Always Beneficial?', *Columbia University* (online, 26 May 2014) <http://www1.cs.columbia.edu/~unger/articles/technologyProgress.html>

Verne, Jules, *From the Earth to the Moon* (Bantam Doubleday Dell Publishing Group Inc, 1996)

Verne, Jules, *Twenty thousand leagues under the sea* (Oxford University Press, 1998)

Walsh, Katie, 'The $1m web service showing law firms are embracing the AI beast', *Financial Review* (online, 1 March 2018) <https://www.afr.com/business/legal/the-1m-web-service-showing-law-firms-are-embracing-the-ai-beast-20180225-h0wme4>

Wells, H G, *War of the Worlds* (Penguin Books Ltd, 2011)

Willard, Stephen, 'Study: People Check Their Cell Phones Every Six Minutes, 150 Times A Day', *Elite Daily* (online, 11 April 2019) <https://www.elitedaily.com/news/world/study-people-check-cell-phones-minutes-150-times-day>

Wortham, Jenna, 'Sending GPS Devices the Way of the Tape Deck?', *New York Times* (online, 7 July 2009) <https://www.nytimes.com/2009/07/08/technology/08gps.html>

B *Other*

IMDB, 'Terminator', *Terminator* (Web Page, 20 December 1984) <https://www.imdb.com/title/tt0088247/?ref_=fn_al_tt_1>

IMDB, 'Westworld', *Westworld* (Web Page, 2 October 2016) <https://www.imdb.com/title/tt0475784/>

Institute of Electrical & Electronic Engineers, 'About IEEE', *Institute of Electrical & Electronic Engineers* (Web Page, 2019) <https://www.ieee.org/>

Netflix, 'Black Mirror', *Netflix* (Television Show, 4 December 2011) <https://www.netflix.com/au/title/70264888>

Victorian All-Party Parliamentary Group on Artificial Intelligence, 'Artificial Intelligence Primer', *Parliament of Victoria*, (Web Page, March 2018) <https://www.parliament.vic.gov.au/images/stories/AI-Primer_Feb2018.pdf>