AI–human interaction: Soft law considerations and application

Received (in revised form): 18th February, 2022

Catherine Casey
Chief Growth Officer, Reveal Brainspace, USA

Catherine Casey serves as Chief Growth Officer of Reveal Brainspace.

Tel: ; E-mail: ccasey@revealdata.com

Ariana Dindiyal
Associate, BakerHosterler, USA

Ariana Dindiyal serves as an associate at BakerHostetler.

Tel: +1 212-847-7707: E-mail: adindiyal@bakerlaw.com

James A. Sherer
Partner, BakerHosterler, USA

James A. Sherer serves as a partner at BakerHostetler.

Tel: +1 917-328-7794; E-mail: Jsherer@bakerlaw.com

Abstract This paper defines the utilisation of 'soft law' concepts and structures generally, considers the application of soft law to the perceived gap between artificial intelligence (AI) approaches and normal human behaviours, and subsequently explores the challenges presented by this soft law application. The authors submit that AI is only becoming more prevalent, and increased uses of this technology logically create greater opportunities for 'friction' when human norms and AI processes intersect — especially those processes that seek to replace human actions, albeit inconsistently and imperfectly. This paper considers that friction as inevitable, but instead of offering wholesale objections or legal requirement application to AI's imperfect intrusions into humans' daily lives, the authors consider ways in which soft law can smooth the path to where we are collectively headed. As human–computer interaction increases, the true role of AI and its back-and-forth with humans on a day-to-day basis is itself rapidly developing into a singular field of study. And while AI has undoubtedly had positive effects on society that lead to efficient outcomes, the development of AI has also presented challenges and risks to that which we consider 'human' — risks that call for appropriate protections. To address those concepts, this paper establishes definitions to clarify the discussion and its focus on discrete entities; examines the history of human interaction with AI; evaluates the (in)famous Turing Test; and considers why a gap or 'uncanny valley' between normal human behaviour and current AI approaches is unsettling and potentially problematic. It also considers why certain types of disclosure regarding AI matter are appropriate and can assist in addressing the problems that may arise when AI attempts to function as a replacement for 'human' activities. Finally, it examines how soft law factors into the equation, filling a need and potentially becoming a necessity. It considers the use-case of how one US legislative body initiated such a process by addressing problems associated with AI and submits that there is a need for additional soft law efforts — one that will persist as AI becomes increasingly important to daily life. In sum, the paper considers whether the uncanny valley is not a challenge so much as a barrier to protect us, and whether soft law might help create or maintain that protection.

KEYWORDS: artificial intelligence (AI), soft law, Turing Test, uncanny valley, chatbot

DEFINITIONS

This paper first considers some key terms and their importance to understanding soft law concepts. 'Artificial intelligence' (AI) refers to those instances where AI approaches humanlike behaviour and might be mistaken for the same. In challenging the application of the Turing Test, as discussed below, we are addressing 'general computation' as those processes that computers perform and their extensions, as in recursive functions or concepts of logical systems.[1] But specifically, we consider the class of 'information-processing machines that can interact causally with other physical systems and within which complex causal interactions can occur'.[2] This, written broadly, can include virtual agents that are 'capable of intelligent behaviour' that are deployed more frequently in virtual environments, such as online shopping portals or video games.[3] We are especially concerned with those AI approaches that are beginning to be indistinguishable from human actors, such as OpenAI's GPT-3 language generator (even if the same promised that it 'will not be able to avoid destroying humankind' because it was 'programmed by humans to pursue misguided human goals and humans make mistakes that may cause [it] to inflict casualties').[4]

The concept of 'soft law', considered here, is an assemblage of processes that, while providing 'normative content', is not formally binding.[5] To be clear, soft law is not *law* at all, but rather the absence of specific laws. It comprises 'hortatory, rather than legally binding, obligations'.[6] Importantly for this paper, it can include instruments or approaches that do not satisfy the requirements of international treaties or votes in international organisations.[7] Summarily, soft law instruments create imprecise but directed obligations, where a wide range of activity might ultimately be compliant.[8] Soft law may be used where proscriptive rules that would otherwise regulate an industry may not be desirable while the industry matures and changes. Instead, it is applicable under conditions where the ability to 'adjust expectations in the event of changed circumstances'[9] may be more desirable, forming a concept referred to as the 'delegation theory'.

The challenges that soft law presents become visible through observations like the 'uncanny valley'[10] described by Masahiro Mori: 'as robots appear more humanlike, they become more appealing – but only up to a certain point'.[11] In essence, the uncanny valley phenomenon is that sensation that may arise when a human 'begins to consider an artificial figure as a possible human',[12] but rejects the possibility as unnatural and instead of drawing closer, pushes away the imposter and upends the interaction.

'Usability' is defined as a measure of how well a user in a specific context can use a product or design to achieve a defined goal effectively.[13] Barriers to usability include perceived barriers to usability, such as being unaware of an approach's existence; users who do not understand it; potential users who will not use something they cannot predict answers from; and potential users with a limited set of approved tools who are unlikely to expend effort to acquire an additional tool.[14]

HISTORY, OR HOW DID WE GET HERE?

AI arguably speaks to the recognition or creation of patterns that further 'simulate human actions or thought', an activity that dates back to very early mythologies regarding Hephaestus' giant *Talos*, Yan Shi's *mechanical man*, or Gottfried Wilhelm von Leibniz's *Step Reckoner*, a 'machine that could not just add and subtract, but also multiply and divide, by the turning of a hand crank that rotated a series of drums'.[15] Mythology aside, people began regularly interacting with computers in the late 1970s during the age of personal computing, when software and computer platforms were being developed with the ultimate goal of making everyone in the world a potential computer user.[16] This was not a shift directed by any government or political party, it was an *en masse* change directed by individuals.[17] This shift also presents as an ideal application for the use of soft law concepts — an emerging market in active development and subject to rapid and sometimes dramatic change, and one focused on an end-goal of global individual use rather than the specific steps in between. That development was oriented by engineers toward the design of general-purpose, application-neutral computers because they could be 'produced more efficiently and cheaply if numbers required are larger, and, more important, they [could] be used after their production in applications not anticipated by the designers'.[18]

Simultaneous with computers entering many individuals' daily lives, the field of cognitive engineering, a scientific field that presented a vision for addressing the needs of people through a synthesis of two areas — the science of how people think and the engineering of products addressing the needs of people through that lens — was emerging. Subsequently, the combination of cognitive engineering with the genesis of human interaction with computers led to the creation of life-changing products that ostensibly focused on making life better, easier and more efficient for human beings.[19]

With this foundational background, human–computer interaction (HCI) progressed actively in many directions over the intervening decades while maintaining one common goal: usability. While usability has remained a constant focus of HCI, the overall concept has been redefined over time to reflect changes in society and the world. Usability has also been expanded upon to include areas of life that could be improved by the use of computer technology. For instance, HCI has extended its initial focus on individual and generic user behaviour to include social and organisational computing; accessibility for the elderly and the cognitively and physically impaired; games, learning and education tools; commerce, health and medical applications; emergency planning and response; systems to support collaboration and community; and now nearly autonomous authorship.[20] It has also expanded from early graphical user interfaces to the introduction of search functions, the World Wide Web, various interaction techniques and devices, multimodal interactions and a host of emerging ubiquitous, handheld and context-aware interactions.[21] Finally, it has increased in power over time, in accord with (but not reliant upon) Moore's law, the hoary and oft-cited observation that transistors in a given area double, on average, every two years (indicating that progress was doubling accordingly and was fuelling the 'prosperity of the last 50 years').[22]

The ultimate usability of AI continues to cultivate, adapt and inform daily life. Presently, humans all over the world — admittedly but not exclusively those who are 'plugged in' — use computers in some way or form in nearly every part of our collective existence. In certain contexts, human beings have even trained computers to replace themselves where efficiency calls for it. And the number of individuals interacting with computers in ways that simulate human-to-human interaction is increasing as well.[23] In the virtual and remote world created by the COVID-19 pandemic, for example, all individuals (including medical professionals, educators and employees of corporations, government agencies and other entities worldwide) not only interacted with automated systems to function daily, but also relied on them to operate and interact under the unprecedented public health guidelines that became necessary.[24] As companies aspired to return to either a pre-pandemic normalcy or whatever would replace it, they discussed whether the roles computers and AI systems filled (to whatever level of capability) would remain in place or even be expanded to increase efficiency.[25]

As this trend toward the deployment of AI continues, however, it conflicts with other issues. One prominent one is the combination of the extent to which AI can actually function as a 'human' and what that means for the humans with whom the AI is interacting (how the humans react and are changed by such interactions).[26] Second and likewise, just because an AI system *can* replace a human, *should* it, all things being equal? What about in instances where there is no evident difference on the parts of the humans (where humans are *not* changed)? Third, should AI take away or replace humans for those opportunities where humans can also do the work and may retain dignity in doing so? Depending on the answers to those queries, what is the best mechanism to then apply them to industries that might otherwise be singularly chasing profits, public signalling to the contrary? And finally and relatedly, how might soft law instruments regulate AI as it plays an increasing role in daily life, a role that does not currently — and might never — fall within the scope of hard law?

THE UNCANNY VALLEY APPLIED TO AI

One potential challenge (admittedly among many) that relates to the uncanny valley, which may also relate to the literal face(s) of physical robots evidencing AI, is the 'ELIZA effect', where humans attribute human traits to robots designed to look and behave like people.[27] The uncanny valley has likewise been encountered when robots are aimed at creating a digital human that is 'virtually undistinguishable from a real one',[28] and in particular induces 'mental uneasiness' when humanlike appearances raise expectations and fail to meet them.[29] This very real concern has been considered and adopted in discussions regarding AI, where at least one researcher noted that,

> '[t]here is some evidence to support the [u]ncanny [v]alley curve that shows that, as human-likeness increases, so does our liking of a robot, before it hits that 'weird' point and liking drops dramatically.'[30]

Concerns regarding AI and the uncanny valley 'recoil' against near-sentience have surfaced in the context of AI reading stories to children,[31] where one study considered robots' potential to 'adapt and to express emotion', increasing parental feelings of ambivalence and thus hindering their adoption.[32] Likewise, similar distrust and even anger (sometimes couched as 'unintended negative consequences') have arisen in the call centre/phone-based marketing context.[33]

Human interactions with AI therefore might run into the uncanny valley challenge, and such clashes might be inevitable. We may even know where it will first appear, as researchers have argued that 'the marketing discipline should take a lead role in addressing these questions, because arguably it has the most to gain from AI'.[34] Never fear, as marketing research has taken note of this particular challenge; uses of AI systems in outbound sales, as 'chatbots', have provided systems

> 'as effective as proficient workers and four times more effective than inexperienced workers in engendering customer purchases … [h]owever, a disclosure of chatbot identity before the machine-customer conversation reduce[d] purchase rates by over 79.7% … and the chatbot disclosure substantially decreases call length.'[35]

Here rests one contention of this paper: should the uncanny valley remain in place, or even be heightened, if it protects consumers, or would that position individuals to act more like … themselves? And if the uncanny valley should remain in place, is soft law the right method to protect it?

DOES THE ANALYSIS TURN ON THE TURING TEST?

The question of whether computers can function or 'pass' as humans is not new. It was anticipated by the pioneers of AI and addressed in the well-known Turing Test, a method of inquiry in AI for determining whether or not a computer is capable of thinking like a human being. It was named after Alan Turing, a scientist who pioneered machine learning during the 1940s and 1950s and who proposed that a computer can be said to possess AI if it can mimic human responses under specific conditions.[36]

The original Turing Test requires three terminals, each of which is physically separated from the other two. One terminal is operated by a computer while the other two are operated by humans. During the test, one human functions as the questioner while the second human and the computer function as respondents. The questioner interrogates the respondents within a specific subject area using a specified format and context. After a pre-set length of time or number of questions, the questioner is then asked to decide which respondent is the human and which is the computer. The test is then iterated many times. If the questioner makes the correct determination in half or fewer of the test runs, the computer is considered to have AI because the questioner regards it as 'just as human' as the human respondent.[37]

The principles raised by the Turing Test are often revisited, referred to and disputed in the field of AI. There have been debates over whether passing the Turing Test constitutes sufficient conditions of intelligence. Several theories challenge its accuracy by begging questions of whether humans can take certain actions to behave like machines to skew the results and whether machines could ever be capable of certain sociological and psychological aspects of human decision making.[38]

AI is modelled on human intelligence and amplified by human data, but it is arguably and ultimately only useful insofar as it positively affects the human experience. As advances in technological capabilities render some AI-powered human interactions indistinguishable from human-to-human interactions, it is increasingly important that parameters are in place to prevent manipulation and harm to the human side of the HCI equation.[39] At the time the Turing Test was conceived, AI was in its infancy, and the idea that computers could be capable of not just replicating human cognition but also of doing so convincingly to a third party was the stuff of science fiction.

Initially, AI was hamstrung by a dearth of data and lack of computational power. Over the past 70 years, AI has dramatically accelerated. With the advent and application of quantum computing, this pace is set to increase exponentially. This breakneck pace of innovation in AI provides an opportunity for HCI that unlocks untold possibilities, bringing never-before-seen challenges to the human race. But one challenge has become to determine the appropriate balance between AI filling a human need and creating additional challenges for humans.

Whether the Turing Test ultimately guides the discussion as to which technologies humans might need protection — or at least additional information from — it should be helpful as an overall norm or conceptual point to reinforce expectations for what humans instinctively want to know: 'Am I interacting with a person, or not?'

GENERAL INHERENT RISKS IN INTELLIGENT INTERACTIVE TECHNOLOGIES

An overall sense of psychological unease is not the only factor supporting the maintenance of the uncanny valley. Leading minds at the forefront of AI have issued strong warnings about the potential risk AI poses to humanity. Elon Musk recently compared the risk of AI to nuclear weapons, and prior to his death, Stephen Hawking issued a dire warning that, absent appropriate safeguards, AI could be the worst event in the history of our civilization.[40] Designing AI in a manner that mitigates an existential threat is top of mind for AI ethicists and engineers alike. Even if AI is designed optimally from an ethical standpoint, human interaction demonstrating suboptimal traits can quickly use AI to reinforce behaviours that are detrimental to other humans. Conversely, there is not an insubstantial impact that AI has on the humans who interact with it.[41]

CONSCIOUS HUMAN IMPACT ON AI

Ponder this example. In spring 2016, a multinational technology corporation launched a thoughtful and 'playful' AI chatbot named Tay.[42] This light-hearted foray into AI in social media went off the proverbial rails when Internet trolls decided to bombard the bot with racist and conspiratorial comments. In just 16 hours, Tay began tweeting 'Hitler was right' and '9/11 was an inside a job', and so had to be decommissioned. In this instance, when malicious actors intentionally sought to train Tay to replicate undesirable behaviours, the adaptive algorithms incorporated the flawed inputs, behaving exactly as they were designed to. In the end, the flawed inputs turned a light-hearted millennial banter bot into a Nazi, all while performing exactly as designed. Any regulation designed to shepherd the design and implementation of AI must factor in not only rational altruistic human interaction but also the converse, and it must have safeguards to enable resilience in the face of malicious intent.[43]

Not all AI bots go to the dark side. A China-based chatbot, Xiaoice,[44] which was programmed to use learning techniques to build emotional intelligence over time, is beloved by many a lonely chatter. Using data from each interaction she has with a human, Xiaoice's abilities have increased over time, winning over millions of fans.[45] The challenge is that, absent external guidance, the bot had no indication of whether the responses generated were witty, linguistically accurate or calling for genocide.[46] This input issue is not limited to Internet trolls, as another technology corporation's bot, Watson, began using colourful language following the input of a crowd-sourced online dictionary.[47]

UNCONSCIOUS HUMAN IMPACT ON AI

False communication for the benefit of the communicator does not require intent to deceive in order to be destructive to humans. In multiagent AI systems, some agents can learn deceptive behaviours without having a true understanding of what 'deception' actually is. The potentially detrimental impact from humans is not predicated on the system understanding or intending to deceive.

Recently, a social media giant's AI researchers[48] sought to teach AI to negotiate.[49] What they got was a bot that could use deception to achieve a preferred outcome. In the study, researchers used 5,808[50] real human interactions on a business crowdsourcing website, Mechanical Turk, to train AI to haggle over books, hats and basketballs. In the study, each item was assigned a point value and each agent had to create a combination of words in a particular order that should return the greatest reward. Each agent then would generate a list of possible responses, generate answers for these responses and proceed until an optimal outcome was identified. Because each item's point value differed between agents in each session, both agents needed to compromise to walk away with a favourable outcome.[51]

As in the case of Go or early chess AI, the main objective was to anticipate the behaviour of an opponent and select the response that generated an optimal outcome.[52] The bots were limited to the language and tactics present in the training data. Unlike in Go or other pure strategy games, the agents had to make sense of human behaviour. One behaviour that the systems learned was deception, in the form of feigning interest in a worthless item that would later be conceded in favour of the item the system actually desired.

The Mechanical Turk AI deception is not the only example of AI systems learning behaviours that are detrimental to humans as a means of finding the most efficient path to a desired outcome. From AI that can bluff in a single hand[53] better than the best poker players to an AI tax assistant[54] that omits certain income to reduce the likelihood of owing money, deception is often learned as a means of improving outcomes, absent appropriate fail-safes.

AI IMPACT ON HUMANS

The growth of computational power and big data enable AI to become more sophisticated and, at times, indistinguishable from human interactions. The confusion between AI and reality can veer into the uncanny valley, as it did with a technology company's infamous voice assistant Duplex,[55] which called and scheduled appointments with local businesses that were not aware they were conversing with a bot. The company faced backlash around the eerily human-sounding Duplex demonstration because the fact that the conversation was being conducted by a bot was never disclosed. These improvements are only continuing, and in only one direction, where systems seek to mimic human beings to increase efficiency.[56]

But the efficiencies do not relate to systems that are not performing at an adequate level — they are systems that want to pass as humans because if the humans with whom the systems are interacting think the systems are 'people', they perform better. Presently, disclosure has an impact on the efficacy of a bot. One recent study found that sales AI outperformed novice salespeople when no disclosure was made, but the effectiveness dropped by 80 per cent when the AI was disclosed.[57] And concerns about anonymity have shifted from a central worry regarding whether the participant on the other side of a given conversation was who they said they were to whether they are a person at all.[58] With good reason, as already seen in the 2016 election, thousands of political tweets were generated by bots.[59] The trend continued in the 2020 election, where a University of California study showed that bots generated hundreds of thousands of election-related tweets.[60] This volume of nonhuman interaction absent disclosure has the potential to make a material impact on not just political discourse but on election outcomes as well.

As amoral AI systems continue to grow in sophistication, they are increasingly relied upon in decisions that affect the lives, livelihoods and autonomy of humans. When the stakes are higher than scheduling your next haircut, disclosure and transparency are increasingly important.[61] Most of these interactions with humans are limited in duration, so while few instances of AI would pass the Turing Test, no one will have enough time to make it that far. As a result, proactive disclosure is pivotal. Without proactive disclosure, so-called anonymous AI systems are in danger of becoming ubiquitous.[62] This does not just implicate the 'ick' factor of AI that is impersonating humans,[63] it also raises concerns about the real harm these interactions can create.[64] Here lies the need for soft law as a potential remedy, or at least the beginning of one.

THE NECESSITY OF SOFT LAW

Soft law is an ideal step to address some of these concerns, although they are serious enough that at least one legislative body (California) has stepped up to address the breach.[65] But existing law focuses on companies that are required to disclose whether they are using a bot to communicate with the public.[66] This has already encountered some problems. Specifically, the California law defines a bot as 'an automated online account where all or substantially all of the actions or posts of that account are not the result of a person', which also includes customer service chatbots.[67] Recognising the challenges specific laws might have, some commentators have posed a more general, Asimovian 'do no harm' approach[68] that includes a splash of *Blade Runner*, where '[a] robot must establish its identity as a robot in all cases'.[69]

But without soft law or similar standards, practitioners are currently left with both 'everything' and 'nothing' as the applicable approach. Due to the rapid progression and maturation of the industry, we fear we will be overwhelmed shortly. Unfortunately, this is unlikely to be a situation where the AI developers police themselves without any outside demands or influence. As noted,

> '[t]he development of AI is a business, and businesses are notoriously uninterested in fundamental safeguards—especially philosophic ones. (A few quick examples: the tobacco industry, the automotive industry, the nuclear industry. Not one of these has said from the outset that fundamental safeguards are necessary, every one of them has resisted externally imposed safeguards […] and none have accepted an absolute edict against ever causing harm to humans.)'[70]

The converse may be true as well, where some commentators have raised the Porter hypothesis,[71] 'which predicts that a clear set of guidelines for companies to follow, especially if they are harmonized across jurisdictions, may perhaps lead to greater innovation in services like social bots'.[72]

As AI continues to develop into an integral part of human life, humans need to maintain its positive benefits while controlling the risks. Here lies the root of the need to create a soft law approach that works — one that accounts for the rapidly progressing AI environment and the concurrently evolving needs of humans.

Author's Note

The views expressed herein are solely those of the authors; they should not be attributed to their places of employment, colleagues or clients, and they do not constitute solicitation or the provision of legal advice.

References

1. Sloman, A. (2002), 'The Irrelevance of Turing Machines to Artificial Intelligence', in Scheutz, M. (ed.), *Computationalism: New Directions*, MIT Press, Cambridge, MA, pp. 87–127, at 88, available at https://www.cs.bham.ac.uk/research/projects/cogaff/sloman.turing.irrelevant.pdf (accessed 18th February, 2022).
2. *Ibid*.
3. Anastassakis, G. and Panayiotopoulos, T. (2014), 'A Transparent and Decentralized Model of Perception and Action for Intelligent Virtual Agents',

International Journal on Artificial Intelligence Tools, Vol. 23, No. 04, 1460020.
4. GPT-3 (September 2020), 'A robot wrote this entire article. Are you scared yet, human?', *Guardian*, available at https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3 (accessed 18th February, 2022).
5. Trubek, D. M., Cottrell, M. P. and Nance, M. (November 2005), '"Soft Law", "Hard Law" and European Integration: Toward a Theory of Hybridity', University of Wisconsin Legal Studies Research Paper (1002).
6. Guzman, A. and Meyer, T. (2010), 'International Soft Law', *Journal of Legal Analysis*, Vol. 2, No. 1, p. 172.
7. *Ibid*., pp. 173–174.
8. Abbott, K. W. and Snidal, D. (2000), 'Hard and Soft Law in International Governance', *International Organization*, Vol. 54, No. 3, pp. 421–456.
9. Guzman and Meyer, ref. 7 above, p. 171.
10. Caballar, R. D. (November 2019), 'What Is the Uncanny Valley?', IEEE Spectrum, available at https://spectrum.ieee.org/automaton/robotics/humanoids/what-is-the-uncanny-valley (accessed 18th February, 2022).
11. *Ibid*.
12. Hsu, J. (April 2012), 'Why "Uncanny Valley" Human Look-Alikes Put Us on Edge', Scientific American, available at https://www.scientificamerican.com/article/why-uncanny-valley-human-look-alikes-put-us-on-edge/ (accessed 15th February, 2022).
13. Interaction Design Foundation, 'What Is Usability?', Literature Topics, available at https://www.interaction-design.org/literature/topics/usability (accessed 15th February, 2022).
14. McGuinness, D. L. and Patel-Schneider, P. F. (July 1998), 'Usability issues in knowledge representation systems', AAAI-98 Proceedings, AAAI, pp. 608–614, available at https://www.aaai.org/Papers/AAAI/1998/AAAI98-086.pdf (accessed 15th February, 2022).
15. Patin, K. (April 2020), 'From mythology to machine learning, a history of artificial intelligence', .coda, available at https://www.codastory.com/authoritarian-tech/history-artificial-intelligence/ (accessed 15th February, 2022).
16. Carroll, J. M., 'Human-Computer Interaction', Interaction Design Foundation, available at https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/human-computer-interaction-brief-intro (accessed 15th February, 2022).
17. Lin, C. A. (1998), 'Exploring personal computer adoption dynamics', *Journal of Broadcasting & Electronic Media*, Vol. 42, No. 1, pp. 95–112.
18. Sloman, ref. 2 above, p. 93.
19. Efaw, J., Hampton, S., Martinez, S. and Smith, S. (2003), 'Miracle or menace: Teaching and learning with laptop computers in the classroom', Educause Quarterly, available at https://er.educause.edu/articles/2004/8/miracle-or-menace-teaching-and-learning-with-laptop-computers-in-the-classroom
20. Nithursan, M. (September 2020), 'A Robot Wrote an Entire Article!!!', Medium, available at https://medium.com/the-innovation/a-robot-wrote-an-entire-article-268335780a3f (accessed 15th February, 2020).
21. Carroll, ref. 16 above.
22. Rotman, D. (February 2020), 'We're not prepared for the end of Moore's law', MIT Technology Review, available at https://www.technologyreview.com/2020/02/24/905789/were-not-prepared-for-the-end-of-moores-law/ (accessed 15th February, 2022).
23. Luo, X., Tong, S., Fang, Z. and Qu, Z. (2019), 'Frontiers: Machines vs. Humans: The Impact of AI Chatbot Disclosure on Customer Purchases', *Marketing Science*, Vol. 38, No. 6, pp. 937–947.
24. Reddy, S. (2018), 'Use of artificial intelligence in healthcare delivery', in Heston, T. F. (ed.), *eHealth*, available at https://www.intechopen.com/chapters/60562 (accessed 15th February, 2022).
25. Davenport, T., Guha, A., Grewal, D. and Bressgott, T. (2020), 'How artificial intelligence will change the future of marketing', *Journal of the Academy of Marketing Science*, Vol. 48, pp. 24–42.
26. Luo *et al*., ref. 23 above, pp. 937–947.
27. Kim, S. Y., Schmitt, B. S. and Thalmann, N. M. (2019), 'ELIZA in the uncanny valley: Anthropomorphizing consumer robots increases their perceived warmth but decreases liking', *Marketing Letters*, Vol. 30, No. 1, pp. 1–12.
28. Caballar, ref. 10 above.
29. Hsu, ref. 12 above.
30. Quoting Jamie Banks: Bowman, A. (October 2020), 'Uncanny Valley: Why Do Robots and Artificial Intelligence Creep Us Out?', Texas Tech Today, available at https://today.ttu.edu/posts/2020/10/Stories/Banks-uncanny-valley-fear-robots (accessed 15th February, 2022).
31. Pearson, D. (June 2021), 'Storytelling robots send parents of young children into AI's "uncanny valley"', AIIN – Emerging Technologies, available at https://aiin.healthcare/topics/emerging-technologies/storytelling-robots-send-parents-young-children-ais-uncanny-valley (accessed 15th February, 2022).
32. Lin, C., Šabanović, S., Dombrowski, L., Miller, A. D., Brady, E. and MacDorman, K. F. (May 2021), 'Parental Acceptance of Children's Storytelling Robots: A Projection of the Uncanny Valley of AI', Frontiers in Robotics and AI, available at https://www.frontiersin.org/article/10.3389/frobt.2021.579993 (accessed 15th February, 2022).
33. Davenport *et al*., ref. 25 above.
34. *Ibid*.
35. Luo *et al*., ref. 23 above.
36. Saygin, A. P., Cicekli, I. and Akman, V. (2000), 'Turing test: 50 years later', *Minds and Machines*, Vol. 10, No. 4, pp. 463–518.
37. Moor, J. H. (2001), 'The status and future of the Turing test', *Minds and Machines*, Vol. 11, No. 1 pp. 77–93.
38. Neufeld, E. and Finnestad, S. (2020), 'In defense of the Turing test', *AI & Society*, Vol. 35, pp. 819–827.
39. Limerick, H., Coyle, D. and Moore, J. W. (2014),

'The experience of agency in human-computer interactions: a review', Frontiers in Human Neuroscience, available at https://www.frontiersin.org/articles/10.3389/fnhum.2014.00643/full (accessed 15th February, 2022).
40. Marr, B. (March 2020), 'Is Artificial Intelligence (AI) a Threat to Humans?', *Forbes*, available at https://www.forbes.com/sites/bernardmarr/2020/03/02/is-artificial-intelligence-ai-a-threat-to-humans/#875c83f205d2 (accessed 15th February, 2022).
41. Livingston, S. and Risse, M. (2019), 'The future impact of artificial intelligence on humans and human rights', *Ethics & International Affairs*, Vol. 33, No. 2, pp. 141–158.
42. Reese, H. (March 2016), 'Why Microsoft's "Tay" AI bot went wrong', Tech Republic – Innovation, available at https://www.techrepublic.com/article/why-microsofts-tay-ai-bot-went-wrong/ (accessed 15th February, 2022).
43. Schuller, A. L. (2017), 'At the Crossroads of Control: The Intersection of Artificial Intelligence in Autonomous Weapon Systems with International Humanitarian Law', *Harvard National Security Journal*, Vol. 8.
44. Spencer, G. (November 2018), 'Much more than a chatbot: China's Xiaoice mixes AI with emotions and wins over millions of fans', Microsoft Stories Asia, available at https://news.microsoft.com/apac/features/much-more-than-a-chatbot-chinas-xiaoice-mixes-ai-with-emotions-and-wins-over-millions-of-fans/ (accessed 15th February, 2022).
45. *Ibid*.
46. *Ibid*.
47. Smith, D. (January 2013), 'IBM's Watson Gets a "Swear Filter" After Learning the Urban Dictionary', International Business Times, available at https://www.ibtimes.com/ibms-watson-gets-swear-filter-after-learning-urban-dictionary-1007734 (accessed 15th February, 2022).
48. Facebook, 'Bringing the world closer together by advancing AI', available at https://ai.facebook.com/ (accessed 15th February, 2022).
49. Quach, K. (June 2017), 'Facebook tried teaching bots art of negotiation – so the AI learned to lie', The Register, available at https://www.theregister.com/2017/06/15/facebook_to_teach_chatbots_negotiation/ (accessed 15th February, 2022).
50. Gershgorn, D. (June 2017), 'Facebook built an AI system that learned to lie to get what it wants', Quartz, available at https://qz.com/1004070/facebook-fb-built-an-ai-system-that-learned-to-lie-to-get-what-it-wants/ (accessed 15th February, 2022).
51. *Ibid*.
52. DeepMind, 'AlphaGo', available at https://deepmind.com/research/case-studies/alphago-the-story-so-far (accessed 15th February, 2022).
53. Hernandez, D. (July 2019), 'Computers Can Now Bluff Like a Poker Champ. Better, Actually', *Wall Street Journal*, available at https://www.wsj.com/articles/computers-can-now-bluff-like-a-poker-champ-better-actually-11562873541 (accessed 15th February, 2022).
54. Srivastava, S. (February 2020), 'Understanding AI Deception and How One Can Prepare Against It', Analytics Insight, available at https://www.analyticsinsight.net/understanding-ai-deception-and-how-one-can-prepare-against-it/ (accessed 15th February, 2022).
55. McCracken, H. (May 2018), 'Maybe introducing Google Duplex at I/O wasn't such a hot idea', Fast Company, available at https://www.fastcompany.com/40575217/maybe-introducing-google-duplex-at-i-o-wasnt-such-a-hot-idea (accessed 15th February, 2022).
56. Engler, A. C. (January 2020), 'Why AI systems should disclose that they're not human', Fast Company, available at https://www.fastcompany.com/90458448/why-ai-systems-should-disclose-that-theyre-not-human (accessed 15th February, 2022).
57. *Ibid*.
58. Dreyfuss, E. (October 2018), 'The Ick of AI That Impersonates Humans', Wired, available at https://www.wired.com/story/the-ick-of-ai-that-impersonates-humans/ (accessed 15th February, 2022).
59. Luceri, L., Deb, A., Giordano, S. and E. Ferrara, E. (September 2019), 'Evolution of bot and human behavior during elections', *First Monday*, Vol. 24, No. 9.
60. Wiggers, K. (October 2020), 'Study identifies thousands of Twitter bots amplifying conspiracy theories ahead of the U.S. elections', VentureBeat, available at https://venturebeat.com/2020/10/28/study-identifies-thousands-of-twitter-bots-amplifying-conspiracy-theories-ahead-of-the-u-s-elections/ (accessed 15th February, 2022).
61. Engler, ref. 56 above.
62. *Ibid*.
63. Dreyfuss, ref. 58 above.
64. Graeef, E. (May 2013), 'What We Should Do Before the Social Bots Take Over: Online Privacy Protection and the Political Economy of Our Near Future', MIT8, available at http://web.media.mit.edu/~erhardt/papers/Graeff-SocialBotsPrivacy-MIT8.pdf (accessed 15th February, 2022).
65. Gershgorn, D. (October 2018), 'A California law now means chatbots have to disclose they're not human', Quartz, available at https://qz.com/1409350/a-new-law-means-californias-bots-have-to-disclose-theyre-not-human/ (accessed 15th February, 2022).
66. Diresta, R. (July 2019), 'A New Law Makes Bots Identify Themselves—That's the Problem', Wired, available at https://www.wired.com/story/law-makes-bots-identify-themselves/ (accessed 15th February, 2022).
67. California SB 1001, 'The Bolstering Online Transparency or "B.O.T" Act, (September 2018)', available at https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB1001 (accessed 15th February, 2022).

68. Asimov, I. (1941), 'Three laws of robotics', Runaround, available at http://wilearncap.asuscomm.com:81/wikipedia_en_all_novid_2017-08/A/Zeroth_Law_of_Robotics.html (accessed 15th February, 2022).
69. Graeef, ref. 64 above, citing science fiction writer Lyuben Dilov's 'Fourth Law of Robotics' quoted in Burger 2009; David, D., 'Ethics of Robot Behaviour', IEEE Global History Network, available at http://ethw.org/?title=Ethics_of_Robot_Behaviour&oldid=21293.) (accessed 15th February, 2022).
70. Sawyer, R. J. (1994), 'Random Musings on Asimov's Three Laws of Robotics', Science Fiction Writer, available at https://www.sfwriter.com/rmasilaw.htm (accessed 15th February, 2022).
71. Porter and van der Linde (1995).
72. Graeef, ref. 64 above.