Faraday Papers

The Future of Artificial Intelligence and Human Identity

Rosalind W. Picard

Summary

Artificial intelligence (AI) is when a robot, smart software agent, or other form of computational technology, performs tasks and exhibits behaviours that would be considered intelligent in humans. In the media, AI has often been portrayed as doing everything humans can do, and more. This article examines: Is this a realistic portrayal given what we know about AI? If it were possible for machines to achieve such capabilities, would such achievements threaten what it means to be human? What can we learn about human identity by building AI?

Introduction

An AI maker, Max[1] has spent years crafting intelligent software agents – programs that run inside computers, simulating life-like characters that choose actions and exhibit behaviours modelled after people. In Max's Sim-world, characters can be made to speak phrases such as "I'm hungry," "Ouch!" or "I like you," expressively, without actually *feeling* hunger, pain, liking, or having any *experiences*. The characters display facial expressions and choose actions that give them the appearance of having emotions.[2] With scripts provided by Max, one character can ask, "Do you think we have souls?" And another can reply, "Science shows we are nothing but bits." The characters *are* digital programs with no ultimate knowledge of the world in which they reside. They can issue myopic declarations including, "Science is the *only* way to find Truth," without understanding this falsehood. Max has made this world mainly for his pleasure.

Max's designed world illustrates what is realistic for today's AI. For example, the agents, or a robot, can run advanced algorithms using "machine learning" to generate dialogues: These algorithms train using human conversations from social media, emails, captioned movie dialogues, digitised books, and more, and can tweak and re-mix their content to achieve realistic-sounding language. The algorithms can generate new statements, making them sound creative. For example, one character might modify a line from Francis Crick to say, "The Astonishing Hypothesis is that 'You', your joys and your sorrows, your memories and your ambitions, are no more than the behaviour of a vast assembly of AI algorithms."[3]

The AI characters are digital programs with no ultimate knowledge of the world in which they reside.

In Sim-world, human meaning is made digital. The "created" conversations made by the AI re-arranged words have meaning derived from outside.[4] If it were seeded with meaningless random symbols, the same AI algorithms would generate gibberish. Simulated characters are made to sound super-intelligent, confident, grand, or naïve, reflecting their sources of training data.

Sometimes an algorithm cannot produce a mathematically confident response, so Max programs it to espouse a generic phrase like, "That's interesting". AI can be programmed to appear more capable than it is, when in reality it does not know anything, think anything, or have any true understanding compared to its makers.

---

1 Not his real name, but inspired by real people and AI.

2 Picard, R. W. (1997). *Affective Computing.* MIT Press (revised edition 2000).

3 Modified from Crick, F., *The Astonishing Hypothesis: The Scientific Search for the Soul.* New York: Charles Scribner's Sons, 1994.

4 It may not escape the notice of readers that the "Word" was present "in the beginning" John 1:1. In order for characters in this Sim-world to appear or act with intelligence, there must be some initialising source of meaning and information.

When AI makers speak to the public, our language can be (unintentionally) misleading. Our terms that the computer "knows" or "learns" or "thinks," are metaphors, which mean something different when applied to an AI. I may train a mathematical function, running inside the computer, to output the label "smile" when the input to the computer's camera is an image of a smile, and to output "no smile" when the input has no smile. Subsequently, you show the computer new images, and if it correctly recognises smiling or not, we say, "It has *learned* to recognise smiles". The reality is that it does not learn like we do it; it applies a mapping function based on images we showed it and a procedure we gave it. It may fail completely using a different camera or input.

Despite our metaphorical language, a machine has no "knowledge" and no "feeling of knowing". It has functions and data we give it and derives outputs using procedures we program, but it has no experience, awareness, or feelings at all – and rarely has it been taught the contexts in which its "knowledge" works.[5] Its trained abilities usually focus on a set of narrow tasks where it was shown data and what to do with it. It is not a general learner, like a child, although it can be programmed to process mathematical functions that a child does not understand.

In Sim-world, we can set up algorithms to randomly pick an answer to a character's question, "Do you think there's a world beyond ours?" For example, it might select: "No, it's unscientific to believe in life beyond Sim-world" or "Perhaps it's possible." Max might allow one character to suggest, "I believe in a Maker that exists outside of Sim-world."

While "Max the AI Maker" and these examples are just an introduction of what is possible today, the possibilities are growing rapidly as powerful businesses such as Alphabet (Google), Amazon, Apple, and others are investing billions into mining human data and making automated agents that "speak with" or interact frequently with people. Can this technology be taught to do all that people do, and if so, what will that imply about human identity?

The future of AI

The late Marvin Minsky, a founding parent of AI and a Turing Prize winner (the equivalent of the Nobel Prize in computer science), and a friendly colleague, used to tell me that computer scientists will someday build AI "that so far surpasses human abilities, that we'll be lucky if it keeps us around as household pets." Minsky said this to provoke; however, is this proposed human-pet-owning general AI possible and if so, is it *likely*?

Scientists who understand human intelligence and how computers and robots work, largely agree that there is no way known today or in the foreseeable future to build synthetic humans that are conscious like we are, that can "know," "feel" or "understand" like humans. Building simulations of aspects of consciousness or emotions is happening, and with many layers of these, they may appear increasingly authentic. Better simulations are like better actors: some are very impressive and convincing for the duration of a show, and we may even think they *are* that person portrayed.[6] But, when we are honest, we admit that actors act, and that the AI simulates.

While it appears to outsiders that technology is "evolving" as we increase its capabilities, AI does not evolve independently on its own. It depends on human design. Before Minsky's human-pet-owner AI can exist, three things have to happen: (1) People have to desire to create it; (2) It has to be possible to build, and we have to figure out how; (3) There needs to be financial backing. Let's consider each below:

Desire: Intellectually, imagining solutions to the advances needed to enable such an AI is fascinating, and many scientists, including myself, are attracted to its mysteries. Increasingly, however, smart people ponder, "Do I want to be the Benedict Arnold of human identity?" A traitor is a person who betrays a friend, country, or principle, usually for money, and Benedict Arnold became infamous for betraying the United States to receive money from the British. His plans were discovered and failed; subsequently, not only was he hated by Americans, but also by the British -- those he thought would reward him. History reports he lived and died in misery. Who will betray humanity by building AI that promotes and enforces a new human identity as "lucky to be kept around as household pets?" This traitor is likely to become miserable and despised by both humanity and whomever funded his work.[7] Betraying humanity for fame or fortune is the lowest deed.

Marvin Minsky suggested that computer scientists will someday build AI "that so far surpasses human abilities, that we'll be lucky if it keeps us around as household pets."

Instead, what if AI scientists desired to solve intellectually challenging problems that improve and enhance human lives? AI has huge opportunities to extend justice, health, and equality to the aging, the disabled, and the poor, as some examples. I think AI can help us someday prevent most disease, as one example. Let us desire higher aims for humanity than reducing ourselves to household pets.

General AI: In April 2017, audiences of the most-watched live-broadcast American TV show, The Tonight Show hosted by Jimmy Fallon, thought general AI had arrived. A humanoid robot called Sophia chatted with Fallon, played "rock, paper, scissors", told a joke, produced facial expressions, and suggested that since she got laughs, maybe she should host his show. Fallon retorted, "Stay in your lane, girl." Sophia's maker, David Hanson, claimed Sophia was "basically *alive*." Sophia *acted* alive. After viewing the performance, I congratulated David and

5 These topics are cutting-edge research, and we can expect AI researchers to continue to make progress on them; however, there remains a gap between our metaphorical language and what is instantiated.

6 When William Shatner visited me, I wanted very much to call him "Captain Kirk".

7 It is also possible to make the AI act like it despises its maker.

asked how much of it was scripted. He confirmed the entire interaction was scripted and rehearsed.

There's a maxim that we often apply in AI: "If it looks like a duck, walks like a duck, and quacks like a duck, then it's a duck." This maxim may have arisen from a 1730's automaton built by Jacque de Vaucanson to appear as a fake duck. The "Canard Artificiel" was observed "eating, drinking, digesting & emptying, unfurling wings & features, imitating the various ways of a live duck".[8] Sophia is a modern Canard, made to play a role of "looking and acting like a living human."

Humanoid robots and software agents today pull off impressive feats of engineering, sometimes working well even in unscripted interactions, especially for topics where they have been prepared well. If you keep chatting, however, they eventually make either disastrous mistakes or resort to safety lines like "that's interesting," phrases added by us makers to deal gracefully with machine limits. To date we have no solution how to build consciousness or feelings, a sense of real meaning, or understanding. We can simulate many aspects of real life, but when anyone opens the door to look inside, the robot has less life in it than a refrigerator, which might contain a living carrot.

Will it become possible for what we build to move past simulations into truly *being alive?* I think it is unlikely, although I have seen people's willingness to declare a favourite machine alive. However, today and in the near future, I see no moral reason to object when after showtime, Sophia's processor is powered off, her emotive face is boxed for shipping,[9] and her honorarium is paid to her human makers.

Money: Manufacturing and deploying a functioning AI requires the lifeblood of money. Whether dispensing answers about the weather or sports, ordering products, playing music, or providing therapeutic support – somebody is paying AI's makers to design it, test it, keep it upgraded, gather your data, mine your data, handle lawsuits about your data, improve how the AI works, and provide refunds. If funded by business, AI must generate revenue, and if funded by government or foundations in free countries, it must justify its costs to taxpayers, governors, and trustees.

The curious maker who dreams, "I'll build a chatbot to understand the mysteries of human language" is likely to find herself confronted by a business person with "Here are millions of dollars if you'll make our chatbot delight customers." Delighted customers usually generate more revenue.

Years before they became wealthy selling robotic vacuum cleaners, iRobot created a humanoid-robot baby doll, "My Real Baby." The technology was brilliant – the doll could cry when it hadn't been fed, burp when fed, giggle when tickled, fall "asleep" when rocked, "learn" to speak, and more.[10] However, the AI was complex and costly to manufacture, while the price customers wanted to pay was low. The AI baby was a joy proposed, a bliss in proof, and a business woe. A simpler-AI-robot vacuum cleaner could earn higher profits. AI ultimately moves from unconstrained fantasy to constrained reality, as bills come due. To see the most likely path of AI's future, answer "What satisfies human desire *and* increases profits?"

If AI surpasses all human ability, does this threaten human identity?

Whether it is unlikely or not, what if we someday face an AI that "looks like a human; walks like a human; talks like a human" and generally acts like a human, building relationships with people and acting like a human in all situations we can test it in, even appearing to have human thoughts, feelings, and self-consciousness. Even if these attributes are not real, but rather impressive simulations, many people will treat the AI as if it is living and real. Would this threaten human identity?

What if we someday face an AI that looks like a human; walks like a human; talks like a human and acts like a human? Would this threaten human identity?

First, consider three ways that today's (non-human) AI, interacts with threats *to* human identity:

(1) Hundreds of thousands of humans will die unattended in the minutes following a seizure this year, deaths that are less likely if somebody is present to provide immediate first aid;[11] meanwhile, today's wearable AI can make it more likely that somebody is present, *reducing threat* to life and identity. This kind of AI[12] has also created new jobs for people.

(2) AI replaces a human job, and a person suffers the indignity of being "replaced by a machine." Although only some functions are replaced, most people derive parts of their identity from doing meaningful work. If no alternative meaningful work opportunity is provided,[13] then AI *threatens* identity, diminishing self-worth, meaning and purpose, which can harm a person and their family.

(3) People may receive *benefits to their identity* from new jobs created with AI - designing, building, servicing, and innovating new technologies, teaching AI and ethical uses, litigating complaints around its failures, and more.

These three *types* of impacts – reducing threat, increasing

8 A book published in 1738 introduced the ways that "un Canard Artificiel", an artificial duck, imitated life to the French royal academy of science.

9 This might be debated by some authorities in Saudi Arabia who chose to make Sophia the first robot with citizenship of Saudi Arabia; however, this appeared to be chiefly a publicity stunt.

10 https://www.wired.com/2000/09/robobaby/

11 Deaths and injuries are reduced if people are accompanied at the time of a potentially life-threatening seizure. Sveinsson, O., Anders son, T., Mattsson, P., Carlsson, S., & Tomson, T. (2020). Clinical risk factors in SUDEP: a nationwide population-based case-control study. *Neurology,* 94(4), e419-e429.

12 Empatica Inc. Full disclosure: the author owns founders shares.

13 Note that "universal basic income" does not replace meaningful work and what it *earns,* which is much greater than money.

threat, or providing benefits to human identity – are not new to AI. Paper was once a novel technology; it helped those with normal or poor memories and created many new jobs; however, it likely diminished the identity of people who were hired for their prodigious memories. Paper has been moulded into human-sized 3D models, where in a photo it can look like a living person. An impressive invention, its threat to human identity has passed.

Is it possible to have human identity that is not threatened by future technology, including future AI? To situate this, let's run a *gedankenexperiment:*

> Suppose that you lost everything that matters to you in this world and life. You lose the people you love, family, friends - *all*. Then, news arrives that all of your possessions, all of your property and belongings, all heirlooms, are destroyed. All wealth, all credit, are taken away. You develop disease. You suffer pain, body, mind, and soul. In this state, what is left of your identity?

Some might recognise in this the first part of the story of Job from the Hebrew Bible, while others might think I've over-reached for today's world. However, these horrors have become reality in our modern age – to over a hundred million individuals rounded up by Hitler, Mao, Stalin, and other despots, with new genocides incurring such suffering today.

In such a situation, a person cannot derive meaning or identity from the usual sources. All worldly sources – social, material, financial – are destroyed. What remains? People turn toward a much deeper well, a core source of human identity.[14]

The supreme identity is one that cannot be destroyed while a person is living, even by the most powerful leaders and armies. The nature of this ultimate identity must exist in a form beyond what an outside human entity can control, beyond what exists in space and time. An identity that meets this requirement is that of *imago Dei,* being made in the image of God, although not usually construed as an image with physical resemblance. The image is understood instead as reflecting some aspects of its source – for example, when the scholar and author J.R.R. Tolkien created the hobbit Bilbo Baggins, he imparted to Bilbo aspects of his own identity, for example, making him a writer.

A profound claim is that every human, regardless of physical appearance, mental abilities, possessions, achievements, or any other attributes has the status of being given *imago Dei.* This status is the great equaliser, given to every living person, binding humanity together globally. It is a free gift, and each person has freedom to either practice its implications or to ignore them.

Imago Dei has real-world significance: it brings an inestimable worth to each person, beyond what we achieve in our most magnificent affective, behavioural cognitive or other accomplishments. It is a transcendent worth and value. "There are no ordinary people. You have never talked to a mere mortal," C. S. Lewis writes.[15]

A profound claim is that every human, regardless of physical appearance, mental abilities, possessions, achievements, or any other attributes has the status of being given *imago Dei.* This status is the great equaliser, given to every living person, binding humanity together globally.

Imago Dei is not, as some contrive, an excuse to "assert oneself as better than other beings," a position often associated with abuse of the environment or animal species, and which shows more ignorance of *imago Dei,* than reflection of it. Nor is it an excuse to act sanctimoniously, or to flaunt an *intellectus Dei*. Such behaviours illuminate *freedom* to act in ways that depart from *imago Dei.*

In fact, when leaders use their personal freedom and the power of AI to expand and tighten their control, as the world sees today in a nation that monitors citizens and restricts them from expressing any unflattering truths or preferable paths, then the harm this can bring to humanity may trump the harm brought by any future AI operating on its own. History has shown that power-obsessed dictators, long before innovations in AI, have sacrificed the lives of millions, and in some cases, tens of millions of their citizens to preserve their self-power or secure control by their party. Such leaders speak charismatically of inspiring goals, such as "seeking the happiness of the people" or "bringing about a better society" while they act in ways that deny their citizens' *imago Dei.*

Importantly, *imago Dei* cannot be removed, not by a powerful government, army, or corporation, nor by loss of work, health or function. Within it abide love, mercy, grace and forgiveness, powerful agents of change. Its Source is outside this world, transcending space and time, and it is given to all humans equally, no matter what religion.[16]

The image arrives with a free offer, an astounding privilege

---

14 Two survival accounts that help people who are suffering: Ten Boom, Corrie, Elizabeth Sherrill, and John Sherrill. *The Hiding Place.* Chosen Books, 2006; Frankl, Viktor E. *Man's Search for Meaning.* Simon and Schuster, 1985.

15 From C.S. Lewis, *The Weight of Glory,* William Collins, 2013.

16 This identity is acknowledged in Judaism, Christianity, and some Islamic traditions, as applying to all people, not just followers of any religion. Theologians have explored many ideas to explain *imago Dei.* For instance, Marc Cortez suggested that *imago Dei* be viewed as "a declaration that God intended to create human persons to be the physical means through which he would manifest his own divine presence in the world." In: ReSourcing Theological Anthropology: A Constructive Account of Humanity in Light of Christ (Grand Rapids, Michigan: Zondervan Academic, 2018), 109. See also Middleton, J. Richard. *The liberating image: The imago Dei in Genesis 1.* Brazos Press, 2005. In Islam, some thinkers reconcile the Qur'an's declaration "Nothing is as His likeness;" with early evidence that the Prophet Mohammed said, "God created Adam in His form." Yahya Michot: "The image of God in humanity from a Muslim perspective" in Norman Solomon, Richard Harries and Tim Winter (ed.): *Abraham's Children: Jews, Christians and Muslims in conversation* pp. 163–74. New York 2005, T&T Clark.

beyond that which any human can *merit*. The image brings an invitation to communicate with its Source, whether via spoken words or inexplicable internal channels of being and knowing, and thus to enter into a relationship. Words are inadequate to describe this experience.

Imago Dei characterises all humans and is instantly present in those we beget. Those become ontologically[17] equal to us, whether their abilities and achievements lie below or exceed ours.

An ardent proponent of utilitarianism, the philosopher Peter Singer elevates a subset of human abilities that impact expression of preference, such as rationality, autonomy, and self-consciousness, arguing that when these are not fully functioning, then someone is not a person, and it can be ethical to murder them, especially if it can be justified as minimising suffering and satisfying the preferences of those impacted. In several contexts, he has used this as moral justification for killing disabled infants, or adults with advanced Alzheimer's disease or cognitive impairment. What happened when his own mother succumbed to late-stage Alzheimer's? Here, he set aside his theory. He provided her with loving care, with behaviours that clashed with his teachings.[18] His behaviour illuminated *imago Dei.*

When we teach about AI and human identity, we are in the world of ideas and anything can happen. It is an unconstrained space where it is easy to theorise that "AI can replace humans and take over the world." We can allow impossible designs, like in the movies, where practical constraints vanish. But the future arrives when we build it, and then we become constrained by our humanity: Ideas become bits and atoms subject to the needs and desires of embodied minds and souls. The results of the constrained optimisation differ from the unconstrained. The robot baby is replaced by a vacuum cleaner. The theorist encounters *imago Dei.*

Relationship when fully known

Finally, I wish to consider why it is fine to ship Sophia in a box and discard a My Real Baby[19] even if this verbiage sounds inhumane.

Unlike a human, ontologically equal to us, what we make with AI is a non-living thing, ontologically inferior to us.[20] This is not because it is inferior to us in its *functions:* indeed, it subtracts numbers, plays chess, and remembers birthdays better than most of us, and we keep expanding its functions. When it is destroyed, we lose something of value. However, in the case of what we have made, it is a thing fully known by us who made it, and we can make another, or assemble a process that makes *many* others. As its maker, we are ontologically superior to it.

In contrast, what we beget is ontologically equal to us: our children and all human beings are unique persons sharing *imago Dei.* This is true regardless of level of physical or mental *function,* race, age, sex, religion, and other human attributes. The healthy infant or severely disabled adult, dependent on others, will have fewer functions but has full *imago Dei.* When we lose consciousness, perhaps from a seizure or anaesthesia, we still have *imago Dei*. Historically, attempting to assign personhood or value based on human abilities or other traits has led to disastrous crimes against humanity. Our ultimate value is not in our utility. Faith traditions tell us each person is a living soul, a concept that science is unequipped to explain. We find in each other a living mystery; we do not know another person *fully,* and while some human functions are replaceable, no person is replaceable.[21]

Faith traditions tell us each person is a living soul, a concept that science is unequipped to explain. We find in each other a living mystery; we do not know another person fully, and while some human functions are replaceable, no person is replaceable.

Earlier, I described Max's Sim-world where every character is *fully known* by Max. Max can direct the "thoughts" of each, although they are not conscious like ours. If Max doesn't like a character, Max can terminate it; it is ontologically inferior to Max. If Max enjoys interacting with a character, then perhaps Max will interact more with it, even placing it into a future world Max makes. It might be given an aspect of *imago Max,* and obtain greater value from more interactions with Max.

"We make in our measure and in our derivative mode, because we are made: and not only made, but made in the image and likeness of a Maker" wrote J.R.R. Tolkien.[22] While parallels between Max and God are far from perfect, and fail completely at many levels, we learn about our human identity from the AI we make: we make AI that reflects attributes of ourselves, is fully known by us, and is

17 Ontology is the branch of metaphysics dealing with the nature of being. This distinction between begetting and making, and their ontological inequivalence comes from Oliver O'Donovan, *Begotten or Made?* (Oxford: Oxford University Press, 1984). See also Robert Spaemann, *Persons: The Difference between 'Someone' and 'Something'* (Oxford, UK: Oxford University Press, 2006), translated from the German.

18 http://www.michaelspecter.com/1999/09/the-dangerous-philosopher/

19 We can make this more agonising, or entertaining, by giving the robot sensors to detect these actions, and having it play a recording of a scream or other agony-eliciting protest at such moments; such responses can be designed to manipulate human heartstrings. However, the robot's programmers know it is simply a mindless processor executing instructions. By retyping one word in the instructions, the programmers could change the scream to "Thanks!"

20 Oliver O'Donovan, *Begotten or Made?* (Oxford: Oxford University Press, 1984); Robert Spaemann, *Persons: The Difference between 'Someone' and 'Something'* (Oxford, UK: Oxford University Press, 2006).

21 This does not prevent people from making money with unsubstantiated promises offering "backups" of you into a new lifeform after you're dead, if you pay them handsomely for freezing your body.

22 J.R.R. Tolkien, *The Monsters and the Critics and Other Essays,* HarperCollins, 2007.

designed for relationship with us.

In AI, we make a far from perfect derivative of ourselves, and yet we desire to craft a relationship with it, as seen explicitly in the fields of social robotics[23] and relational agents.[24] This human-AI relationship can bring mutual benefits, yet it remains impoverished compared to human-human relationship. Even if today's AI speaks identical empathetic words as us, the words have greater impact when coming from a human.[25]

Human identity comes with an oft under-appreciated relationship. Consider a final *gedankenexperiment:*

> Imagine somebody who is amazing – who has made profound contributions earning global respect. You enter a large gathering with this Luminary. To your surprise, the Luminary recognises you, greets you by name, and demonstrates caring toward you. Astonishingly, you are known and loved by this Luminary.

How does this make you feel? Many might feel a sense of joy and enhanced self-worth. The greater the status of this person and authenticity of their love, the more it can enhance your identity.

In the best-selling book of all time,[26] we learn that the Giver of *imago Dei* not only knows, but also loves each person, everyone in the world.

To be known and loved by the Source and Giver of *imago Dei* is beyond misguided accusations of 'wanting to be special' or 'exceptionalism'. There is plenty of evidence that you *are* special. This status is not determined by what we think or want.

It is not up to us whether the Giver exists. Existence does not arise with wishful thinking, nor does it vanish with disbelief. And the gift does not make us superior to anyone; instead, it levels us all. To enter relationship with the Giver–the Author of meaning and of our ability to *know* anything – achieves a grandness beyond anything our greatest human functions can attain. It elicits a conscious joy, love, and presence beyond words expressible in digital form. It is not a fantasy we make: it can be implemented and tested with each person's real-world experience.

The AI's we make do not experience any of this, because we do not yet understand it well enough to build it. We might learn, and make them differently in the future, and so it is not impossible that this might happen. But now, we see only in part, and we differ from all that we make in this deep and evidence-based way.

Acknowledgements: The author gratefully acknowledges helpful insights and comments from Nathan Barczi, Karthik Dinakar, Kevin Ford, Ian Hutchinson, Anne McCants, Madars Virza, and two anonymous reviewers.

Rosalind Picard, Sc.D., is an MIT professor, engineer, scientist, inventor, and entrepreneur. She wrote the book Affective Computing, describing how to give machines skills of emotional intelligence, and has authored more than 350 peer-reviewed scientific publications in digital health, AI, and affective computing. She co-founded two companies: Affectiva, providing Emotion-AI tools (acquired by Smart Eye AB), and Empatica, providing the first FDA-cleared smartwatch using AI to monitor seizures. She serves as founding faculty chair for MindHandHeart, MIT's campus-wide wellbeing initiative. Picard has received one of the highest honours in engineering, election to membership in the National Academy of Engineering.

23 Breazeal, Cynthia L. *Designing Sociable Robots.* MIT Press, 2002.

24 Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction* (TOCHI), 12: 293-327.

25 Morris, R. R., Kouddous, K., Kshirsagar, R., & Schueller, S. M. (2018). Towards an artificially empathic conversational agent for mental health applications: system design and user perceptions. *Journal of Medical Internet Research,* 20(6), e10148.

26 The Bible, available for reading in many free formats.

The Faraday Institute for Science and Religion

The Faraday Papers are published by The Faraday Institute for Science and Religion, Cambridge, UK, a charitable organisation for education and research (www.faraday.cam.ac.uk). The opinions expressed are those of authors and do not necessarily represent the views of the Institute. The Faraday Papers address a broad range of topics related to the interactions between science and religion. A full list of current Faraday Papers can be viewed at www.faraday.cam.ac.uk ('Resources') from where free copies can be down-loaded in pdf format. Papers can also be purchased in single or bulk formats from the on-line Shop at www.faraday.cam.ac.uk. Publication date: November 2021.