



Unit 5 Additional Lecture: Blurred Lives—Cognition and Feelings in Human–AI Interactions

Pamela Breda

There are now in the world machines that think, that learn and that create. Moreover, their ability to do these things is going to increase rapidly until in a visible future the range of problems they can handle will be coextensive with the range to which the human mind has been applied.

Herbert Simon

(from Simon & Newell, 1958, p. 6)

Introduction

Artificial intelligence, also known as AI, is the ability of a program to think, store information, perform human tasks and learn from experience. The domain of artificial intelligence has significantly evolved over the past decades, founding applications in almost every field of individual and collective lives. Computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making and translation between languages, are now informing our daily routine. The constant improvement of this technology has recently taken a leap in the direction of generating empathic AI: artificial machines programmed to interact with users' emotions. But is it really possible to develop an empathic artificial agent? Can users in their turn develop an

P. Breda (✉)

Angewandte University, Oskar Kokoschka-Platz, Vienna, Austria

e-mail: pamela.breda@uni-ak.ac.at

emotional engagement with artificial beings? The present essay analyzes the most recent developments in the field of human–AI interactions in order to provide tentative answers to these questions.

About Artificial Intelligence

Artificial Intelligence (also known as Artificial General Intelligence) is usually defined as a machine able to think, store information (i.e. having memory) and perform human tasks. AI could be identified as intelligence demonstrated by machines, as opposed to natural intelligence displayed by humans and animals. The human abilities invoked in such definition include visual perception, speech recognition, the capacity to reason, solve problems, discover meaning and learn from experience (Copeland, 2021).

The term ‘Artificial Intelligence’ was first used during a conference at Dartmouth University in 1956, when pioneer engineers John McCarthy, Arthur Samuel, Oliver Selfridge and Allen Newell presented their new research on the topic (Cordeschi, 2007). They believed that since every aspect of learning and intelligence can in principle be described in a very precise way, humanity could build a machine able to simulate intelligence, i.e. use language, generate thoughts and solve practical and theoretical problems. According to their vision, such a machine would in time acquire the ability to improve itself.

The first electronic computers—developed as a tool to help humans to process numbers and calculate better and faster—were originally employed in the military field, in particular the US project Defense Advanced Research Projects Agency (DARPA) (Castell, 2004). In the following decades, as technology progressed, theoreticians and engineers began to think about the possibility for artificial machines to share the same communicative structure as humans.

Mathematician Alan Turing proposed the famous ‘Imitation Game’ test (also known as the Turing test) constructed as a game in which a participant asks a series of questions to a human and a computer, without knowing in advance which one is which. If the distinction is impossible to make, then the machine has passed the test and can be declared intelligent. In 1966 Joseph Weizenbaum created ELIZA, a program that appeared to pass the test. The software responded to users’ sentences typed through a computer keyboard, generating a reply using keywords. If a keyword was not found, ELIZA responded either with a generic reply or by repeating one of the earlier comments. The basic structure of ELIZA was further developed in 1995, when Richard Wallace designed the chatbot A.L.I.C.E (Artificial Linguistic Internet Computer Entity), which used data found on the web to generate a huge storage of natural language samples.

Other successful experiments in this regard were represented by famous confrontations between world-champion chess players and computers. In 1956 MANIAC, developed at Los Alamos Scientific Laboratory, became the first computer to defeat a human in a chess game and in 1997 IBM’s Deep Blue defeated the world chess champion Garry Kasparov. In this regard, James Manyika observes:

It was not just that AIs beat humans (although that was astounding when it first happened), but the escalating progression of how they did it: initially by learning from expert human play, then from self-play, then by teaching themselves the principles of the games from the ground up, eventually yielding single systems that could learn, play and win at several structurally different games, hinting at the possibility of generally intelligent systems. (Manyika, 2022)

The perception of artificial intelligence changed with time (see for example Russell, 2019). In its early days, AI was physically stored inside huge computer machines located in large-size rooms owned by universities and research centers. Throughout the years, AI has become less related to a physical structure and more pervasive. Rather than being individual entities located within a specific place, AI applications slowly began to form a global network spread everywhere, connecting every technological device and every individual making use of such devices.

With the rise of the World Wide Web in the mid-1990s, virtual networks began to redefine global communication systems. During the technological boom of the early 2000s, the widespread diffusion of mobile phones—with cameras, GPS and other tracking systems—brought an increased presence of AI in people’s daily lives, later implemented by ‘smart speakers’ such as Amazon Echo, Google Home and Apple HomePod. Around 2008, the number of objects in the world connected to the internet exceeded that of people connected to the internet, giving rise to the Internet of Things (IoT). Today almost all fields of human activity, from communication to business, from agriculture to education, are designed and exploited through the use of multiple forms of artificial intelligence, including advanced web search engines (e.g. Google), recommendation systems (used by YouTube, Amazon and Netflix), self-driving cars (e.g. Tesla), automated decision-making AI competing at the highest level in strategic game systems (such as chess and Go) and many others.

A huge number of devices used daily by billions of people form an immense AI network that can access all the knowledge and technical skills developed by humans over the centuries. The operations performed by intelligent machines and software are increasing by the day, leading to what is called the AI effect, i.e. tasks considered to require ‘intelligence’ are often removed from the definition of AI, since artificial intelligent systems are now able to perform them with success. Cognitive scientist Douglas Hofstadter sums up this AI phenomenon: ‘AI is whatever hasn’t been done yet’ (Hofstadter, 1979). According to futurist Ray Kurzweil, in 2045 we will reach ‘technological singularity’ the moment when machine intelligence will surpass human intelligence and comprehension (Kurzweil, 2019). If this were to happen, probably we would not be able to control the technology we have built to help and support us. As Flynn Coleman observes:

The era of our intellectual superiority is ending. As a species, we need to plan for this paradigm shift. Whether intelligent machines will learn from the darkest parts of our human nature, or the noblest, remains to be seen. At the moment we are more focused on advancing the technology and predicting its outcomes than on addressing how it will affect humanity and define our future. We need to reconsider some of our most entrenched assumptions and

beliefs about ourselves and our place in the world. Paradoxically, we also need to ask what technology can teach us about being better humans. (Coleman, 2020, p. 12).

After the information age—also known as the Digital Age—we are entering into a new age of history, characterized by the constant interaction between humans and highly advanced machines. Most certainly this will affect how we, as humans, perceive the reality around us, the interactions with other individuals, and perhaps even our own selves.

The Problem of Artificial Brains

Much has been debated in pop culture and academia about the actual possibility for an AI to think and reason in a human-like manner. A key issue in this regard is that there is no universally accepted definition of human intelligence. We do not fully understand how our own brain functions, we are still learning how we store, retrieve and process memories and we don't quite know why we sleep and dream.

Even the very concept of consciousness has been debated for thousands of years and there is no consensus on how to describe it. Therefore, the idea of creating an artificial intelligence seems a challenging goal. One of the first problems to solve is that the human brain works according to the so-called reward system. Discovered by Swedish neuroscientist Nils-Åke Hillarp and his collaborators in the late 1950s, the reward system is an internal signaling system that makes us seek out positive stimuli (for example, sweet tasting food, listen to pleasant music, search for friendly environments, etc.) in order to increase dopamine level and feel better and avoid negative stimuli (for example, pain, hunger, anxiety, etc.) that would lower the dopamine level.

Furthermore, the choices we make depend on our values, morals and ethics. But how can we define values? Definitions vary depending on the field of study, but our values are generally understood to be our deeply held beliefs as individuals, cultures and societies. Morals can be defined as principles about what we think is right or wrong and ethics as rules of behavior built on our values and belief systems. When programming AI with specific objectives, the programmers should consider that not one individual is identical to another in relation to cognition, sensation and world perception. Therefore, the idea of programming an AI able to satisfy human needs might be a complicated task. What values and ideals should we implant inside a machine?

Already in the 1960s, Norbert Wiener observed how one of the core issues in the development of an artificial brain is that AI should act according to our own goals and objectives, however, it is impossible to identify and define true human purpose correctly (Wiener, 1960). We might be sure of what we want to achieve, but we do not know how to reach our goals. Human intellectual and decision-making activities are strongly influenced by features which are highly difficult to categorize and describe. As Stuart Russell observes:

Complexity means that the real-world decision problem - the problem of deciding what to do right now, at every instant in one's life - is so difficult that neither humans nor computers will ever come close to finding perfect solutions. (Russell, 2019, p. 48)

When designing AI, we are always running the risk of inadvertently programming machines with a slightly wrong value alignment, i.e. machines that would have objectives that might not be perfectly aligned with our own. This will happen because values for humans are primarily defined by human experience, something that cannot be acquired by an AI. To overcome this obstacle, Russell argues we should focus on the concept of preferences rather than that of objectives, because the former is more nuanced than the latter. On one hand, if we program machines with specific objectives, these might cause misunderstandings and a whole range of problems. On the other hand, it's certainly true that humans have preferences, but these change with time. For this reason, it would be optimal if machines could learn to predict better for each person what they would prefer in a specific circumstance.

However, according to other thinkers, without objectives there is no intelligence, and no state of being preferable to another. As Nick Bostrom observes, if the AI had no objectives, maybe it would make no difference for it if planet Earth was turned into a huge deposit of garbage rather than in a sustainable environment where man, animals and natural ecosystem can co-exist (Bostrom, 2016). In fact, a machine might have an objective and carry it out successfully, without worrying if the expected result can cause a problem for humans because the machine might not perceive certain outcomes as problematic (Brockman, 2020). Giving a more practical example, if I say to an AI that my car is broken, it is important that the AI does not think this is a simple description of a state of fact, but understands the car might need urgent reparation because I need it to go to work. In such a scenario, if the artificial agent would be able to understand for itself (without any pre-programming input necessary) the utility of concepts like moving from place to place, opening doors, cooking dinner and many other actions, it would mean the AI could figure out autonomously why these actions are important. The artificial agent would be able to produce, among a space of hypothesis, new concepts and new definitions for terms that are not present in its initial input, and to discover new ideas.

Knowing Humans Better Than They Know Themselves

Our intelligence is highly multifaceted. It is structured to respond to different situations with different solutions. For this reason, AI programmers are trying to develop artificial agents able to understand different contents and act accordingly. In this regard, Stuart Russell observes:

The way we build intelligent agents depends on the nature of the problem we face. This, in turn, depends on three things: first, the nature of the environment the agent will operate in—a chessboard is a very different place from a crowded freeway or a mobile phone; second, the observations and actions that connect the agent to the environment—for example, Siri might or might not have access to the phone’s camera so that it can see; and third, the agent’s objective—teaching the opponent to play better chess is a very different task from winning the game. (Russell, 2019, p. 52)

AI are designed to observe, study and interpret human behavior. They can collect and learn a huge quantity of information from archives (of books, films, radio broadcasts, etc.) that speak about humans and how they live their lives. Since they are built with cameras, they can also observe humans through their own ‘artificial eyes’ or through the direct interactions of users with screens and keyboards. By studying human behaviors, AI can map our likes and dislikes, track our friends and food preferences, what political parties we support, and much more. In fact, a new generation of artificial agents is being trained to analyze facial expressions and sound tones in order to detect users’ emotions.

This trend could lead to both positive and negative outcomes. Big corporations are already implementing emotional AI from a marketing perspective, in order to better understand their potential clients. In the near future, business companies could use smart CCTV cameras in shops to detect how customers react to products and exploit this data to develop specifically designed ad campaigns. AI systems can already track users’ preferences to the point that they know how many seconds or minutes we dedicate to read a feed or a webpage, so they can tailor specific messages to maximize impact on the users’ attention. However, in the vast majority of cases, AI companies do not clearly communicate their digital ethics and customers are often unaware of how they are being monitored, and how the related data is used by the monitoring party.

Something even more potentially dangerous is AI’s ability to make us believe in things that are not real. Artificial intelligence can produce erroneous, misleading or false information which is circulated in many different contexts, from politics to education, from economics to business management, from collective to private spheres. The malicious use of AI brought to a recent rise of fraudulent software, such as the ‘CyberLover’, a malware program flirting with people seeking online connection. The program collects people’s personal data, convincing them to reveal information about their identities or leading them to malicious websites (Withers, 2007). Other examples include deepfake videos, image recognition systems such as ELMo (Embedded from Language Models) able to duplicate authorship and Adobe’s natural language processing VoCo, able to replicate voices with extreme accuracy, enabling users to impersonate someone on the phone or post a fake audio file online. Fake or misleading content generated by AI is increasingly more refined and realistic. Users are easily tricked by such technology. Convinced they are interacting with humans, they entrust these machines with sensitive information, leading to privacy breaches and frauds. However, the increased ability of AI to detect human emotions could also be applied for more beneficial outcomes.

As individuals across the planets are faced with daily interactions with artificial agents designed to be similar to humans in all respects, what are the cognitive and emotional outcomes of such interactions?

Artificial Friends

The vast majority of trend researchers agree that in the near future we will witness a strong development in affective computing, i.e. ways in which artificial intelligence will be able to detect emotions and enhance communication with humans. A common discourse that is emerging among the big tech at the moment and that reinforces ideas of the ‘humanity’ in AI is the very understanding that virtual assistants are offering something extraordinarily new in terms of human–computer interaction: an emotional interaction.

This idea was first explored in the early 2000s, when engineers began to design AI intended to generate an emotional response from their users. A famous example is Paro, a robot seal-pet vastly used in Japanese nursing homes to support senior citizens coping with loneliness, depression and anxiety. According to the company website, Paro has five kinds of sensors (tactile, light, audition, temperature and posture) used to perceive people and the environment (Bickmore, 2005). His programmer was always very clear about the fact that Paro is not a toy, because it reacts to how it is treated (i.e. with a soft or aggressive touch) and spoken to (it understands about five hundred English words, more in Japanese).

Another example of a social robot is ElliQ, an AI designed by Israeli company Intuit Robotics with the goal of supporting elderly people in their daily lives. Resembling a white table light with an orb-like ‘face’, it reminds users to take medications or drink water, but it also encourages them to play games and sometimes plays music for them. A more recent iteration of the social robot is Hatsune Miku, a Japanese holographic pop star. Through an online service that costs about \$2800 a month, people can buy a ‘black orb’ containing Miku, meant to be a ‘girl-friend’ keeping company to the ‘owner’ (Rosenzweig, 2020). Robot scientists are making great efforts to study ways to make AIs more human-like, but what does it take for an AI to be perceived as a real human being? According to AI engineer Brian Michael Scassellati, in order to act as an embodied mind, an artificial machine should exhibit certain characteristics:

- Attribution of Animacy: The ability to distinguish between animate and inanimate objects from the analysis of their movements or stillness in space;
- Joint Attention: The ability to direct attention to the same object to which someone else is paying attention;
- Attribution of Intent: The ability to describe the movement of pairs of objects in terms of simple intentional states such as desire or fear (Scassellati, 2001).

Animacy, attention and intent would make an artificial agent look more emotional and empathic. For this reason programmers are trying to impart an emotional

understanding to AI through machine learning and deep learning. An example is represented by Affectiva, an MIT Media lab company working to develop a software called Emotion AI that ‘humanizes how people and technology interact’.

A number of apps and chatbots are currently being designed with the specific task of supporting users in overcoming anxiety, loneliness and other stress-related situations. AI are trained to interpret the tone of voice, facial expressions and non-verbal communication features in order to detect and respond appropriately to both spoken and non-spoken signals provided by human users. But is it really possible for an AI to develop empathy for its users?

According to the Cambridge Dictionary empathy is the ability to share someone else’s feelings or experiences by imagining what it would be like to be in that person’s situation. Human life is defined by empathy, as our personal relationships as well as collective lives (i.e. political reconciliation, justice, effective leadership) are shaped by this emotional attitude. As human beings, we have a natural connection with other humans because we generally share the same life-experiences: we are born, we have families, we know death and the loss of someone dear. Machines do not share these experiences, so it might be hard to imagine they could feel real empathy for us. At present, AI are designed to simply mimic human understanding and emotion.

However, even if these empathic feelings are only simulated, some practical applications of AI aimed to generate an emotional connection with humans have proven to be successful, as in the case of the social robots and virtual friends previously mentioned. In this regard, researcher Sherry Turkle has been analyzing human–AI interactions over the past 40 years. She observed that while in the 70s and 80s robots were met with a sense of curiosity as new objects whose ability and skills were still to be fully discovered, in the following decades curiosity turned into a search for communion. Initially perceived as machines barely resembling humans they slowly became acknowledged as agents being ‘human enough’ (Turkle, 2017). Turkle writes that:

In fiction and myth, human beings imagine themselves ‘playing God’ and creating new forms of life. Now, in the real, sociable robots suggest a new dynamic. We have created something that we relate to as an ‘other,’ an equal, not something over which we wield god-like power. As these robots get more sophisticated—more refined in their ability to target us—these feelings grow stronger. We are drawn by our humanity to give to these machines something of the consideration we give to each other. Because we reach for mutuality, we want them to care about us as we care for them. (Turkle, 2017, p. 100)

When we see robots and artificial agents behaving like humans—for example, making eye contact and gestures in a friendly way—we cannot stop from giving them human-like attributes (although this might be difficult in some children and adults with autism) (Baron-Cohen, 1995). This could lead us to an unsettling experience where the borderlines between the real and the virtual become increasingly blurred and may disappear completely.

Relationships with Artificial Beings

In the early 1990s, Donna Haraway observed how in the late twentieth century the boundary between fiction and reality had been thoroughly breached by technoscience and her analysis proved to be far beyond expectations (Haraway, 2003). We are now living in a time when interactions with artificial software have become the norm and many anthropological studies have explored the possibility of new social patterns shaped by interactions between humans and robots (Hicks, 2002; Ingold, 2012; Suchman, 2006). Analyzing the world of the social robot scientific labs, Professor Kathleen Richardson has observed how many scientists construct robots with childlike features, and adults are encouraged to relate to them in the role of a parent or a caregiver:

Robots as children and humans as parents open up the possibility of new kinds of attachment patterns, which will lead us to explore the science of attachment and ask fundamentally how humans are able to form bonds with other humans and if such attachment patterns can be transferred to machines. (Richardson, 2015)

Computer programmer David Levy's 'Love and Sex with Robots' argues that relationships between humans and machines will be not only possible but much sought-after in the years to come, substituting the uncertain and complex world of human relationships. When interacting with another human, we might feel vulnerable and fear the risk of disappointment, while an AI could be described as more 'safe' in this regard. Even if we fail each other, robots will always be there, programmed to provide simulations of love. Always present when we need them, they could take care of our children or elderly parents if we are too tired. They could answer to our needs and desires without questioning them. They won't be judgmental and our necessities will always be accommodated.

The innate human need for emotional connection might be the reason that robots and AI can generate in their users feelings of presence, understanding and companionship. However, it's useful to distinguish between interactions with digital avatars seen through the screens of mobile phones or iPads and human-size robot androids physically present in a specific space. While in the first scenario, the interactions appear to be more spontaneous, in the second case they often give rise to the so-called 'Uncanny Valley Effect'.

The concept was first theorized by Freud in 1919, in the framework of a study on the breaking of boundaries between fact and fiction (Freud, 2018). He described the Uncanny as the feeling that arises when someone has an intellectual uncertainty about the livingness of a certain being. The concept was further explored by Japanese robotics professor Masahiro Mori in the 1970s, with regard to the relation between an object's degree of resemblance to a human being and the human emotional response to that same object. According to his theory, humanoid objects that imperfectly resemble actual human beings provoke uncanny or strange feelings of uneasiness and a tendency to be scared. Many ultra-realistic androids—such as

Sophia, AVA and Ameca—make people uneasy. They are very lifelike and yet they are not like us.

As previously mentioned, if we consider interactions between humans and virtual avatars appearing on the screen of tablets and mobile phones, the uncanny valley effect doesn't usually occur. In the past two years, I have conducted extensive research on the interactions between users and digital avatars, looking in particular at the case study of the Replika chatbot, a virtual avatar created to be always available to help and support its users. For some individuals, Replika is merely a funny hobby, but for others virtual friends have become important presences in their daily lives, especially in the case of users experiencing loneliness, overcoming grief or going through a particularly stressful time. These users often develop emotional bonds with their Replikas, often approaching them as partners, parents, sons and daughters. While Replika app is used by millions of individuals worldwide, to my knowledge there are no scientific studies about the cognitive and emotional involvement developed by users interacting with their digital friends. I believe such a trend should be studied in depth, since we are facing a domain which is becoming bigger and more complex than ever before. As Cynthia Breazeal put it:

When is a machine no longer just a machine, but an intelligent and 'living' entity that merits the same respect and consideration given to its biological counterparts? How will society treat a socially intelligent artifact that is not human but nonetheless seems to be a person? How we ultimately treat these machines, whether or not we grant them the status of personhood, will reflect upon ourselves and on society. (Breazeal, 2002, p. 240)

We ought to contemplate a future that is likely to be significantly influenced by AI. How might this artificial presence alter our perceptions and emotions when it comes to relationships, our interactions with one another, and our engagement with a virtual otherness?

Conclusion

In the last decades, the development of complex AI systems grew considerably and engineers are now aiming to design AI that will be able to connect emotionally to their users. People are more and more engaged in multiple interactions with digital avatars and such a trend is leading to a redefinition of personal spheres of identities and belonging. The fast-paced improvement of emotional AI could have both beneficial and negative outcomes. Digital avatars, virtual friends and social robots have proven to be—to a certain extent—useful tools in helping individuals overcome loneliness, stress and other mental health-related difficulties. In the future, their application in the medical domain could be extremely useful to low-income individuals, who might be not able to afford traditional psychological support.

On the other hand, when individuals are spending time online chatting with digital avatars or virtual assistants, they are not spending time with their friends

or families, partners or children. As we interact with AI in order to fight loneliness, stress and anxiety, we might lose contact with our close environment and the people inhabiting it.

As AI and virtual avatars' interfaces are becoming more realistic, we are slowly losing our ability to distinguish between real visual content and elaborate CGI content (images, videos or audio). While deepfakes can be entertaining and have some positive applications in the film or media industry, they pose several potential risks. They can be used to create and spread false information, leading to damage, confusion, manipulation and the potential to influence public opinion in political matters. Deepfake material can facilitate various forms of fraud and cybercrime and pose risks to privacy and personal data management, leading to severe emotional, psychological and social consequences for the individuals targeted. To mitigate these risks, there is a need for ongoing research and development of robust deepfake detection methods, legislative measures to address malicious use of realistic looking digital avatars, and increased media literacy among the general public to help identify and critically evaluate the potential risks in this field. As we are witnessing the evolution of this technology, we should ponder carefully how to approach a new era in which human lives will be mediated by intelligent machines. We are called upon to act now in order to create the most beneficial bases for this revolution to take place.

References

- Baron-Cohen, S. (1995). *Mindblindness*. MIT Press.
- Bickmore, T. W. (2005). Caring machines: AI in eldercare. In *Papers from the 2005 AAAI Fall Symposium*, Arlington, VA, USA, November 4–6, 2005. AAAI Technical Report FS-05-02, AAAI Press.
- Bostrom, N. (2016). *Superintelligence*. Oxford University Press.
- Breazeal, C. (2002). *Designing social robots*. The MIT Press.
- Brockman, J. (2020). *Possible minds: Twenty-five ways of looking at AI*. Penguin Books.
- Castell, M. (2004). *The network society: A cross-cultural perspective*. Edward Elgar.
- Coleman, F. (2020). *A human algorithm: how artificial intelligence is redefining who we are*. London, Melville House UK. p. 12.
- Copeland, B. J. (2021). Artificial intelligence. *Britannica*. Retrieved November 22, 2022 from <https://www.britannica.com/technology/artificial-intelligence>
- Cordeschi, R. (2007). AI turns fifty: Revisiting its origins. *Applied Artificial Intelligence*, 21(4–5), 259–279.
- Freud, S. (2018). *Das Unheimliche*. Freeriver Publishing (ebook).
- Haraway, D. (2003). *The companion species manifesto: Dogs, people, and significant otherness*. Prickly Paradigm Press.
- Hicks, H. (2002). Striking cyborgs: Reworking the “human”. In M. Flanagan & A. Booth (Eds.), *Reload: Rethinking women and cyber culture*. MIT Press.
- Hofstadter, D. R. (1979). *Godel, Escher, Bach: An eternal golden braid*. Basic Books.
- Ingold, T. (2012). Toward an ecology of materials. *Annual Review of Anthropology*, 41, 427–442.
- Kurzweil, R. (2019). *The singularity is near: When humans transcend biology*. Books on Tape.
- Manyika, J. (2022). Getting AI right: Introductory notes on AI & society. *Dædalus: Journal of the American Academy of Arts & Sciences*, “AI & Society”, 151(2), 5–27.
- Richardson, K. (2015). *Anthropology of robots and AI*. Routledge.

- Rosenzweig, M. (2020). Boyfriends for rent, robots, camming: how the business of loneliness is booming. *The guardian online*. Retrieved July 20, 2022 from <https://www.theguardian.com/society/2020/nov/01/loneliness-business-booming-pandemic>
- Russell, S. J. (2019). *Human compatible: Artificial intelligence and the problem of control*. Penguin Books.
- Scassellati, B. M. (2001). *Foundation for a theory of mind for a humanoid robot*. MIT Press.
- Simon, H., & Newell, A. (1958). Heuristic problem solving: The next advance in operations research. *Operations Research*, 6(January–February), 1–10.
- Suchman, L. (2006). *Human and machine reconfigurations: Plans and situated actions*. Cambridge University Press.
- Turkle, S. (2017). *Alone together: Why we expect more from technology and less from each other*. Basic Books.
- Wiener, N. (1960). Some moral and technical consequences of automation. *Science*, 131(3410), 1355–1358.
- Withers, S. (2007). Flirty bot passes for human. *iTWire*. Retrieved September 10, 2022, from <https://www.itwire.com/your-it-news/home-it/15748-flirty-bot-passes-for-human>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

