



Human/AI relationships: challenges, downsides, and impacts on human/human relationships

Anne Zimmerman¹ · Joel Janhonen² · Emily Beer¹

Received: 25 May 2023 / Accepted: 11 September 2023
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2023

Abstract

Advances in artificial intelligence have resulted in objects that behave and sound like humans and exude a more human feel. Therefore, the relationships between humans and technology have evolved, becoming more personal and complex. Some AI is harmful to humans due to the nature of the human relationship with it. We explore examples ranging from chatbots to AI romantic partners. While humans must better protect themselves emotionally, tech companies must create design solutions as well as be transparent about profit motives to address the growing basket of harms. We propose solutions including alignment with AI principles that promote well-being, prevent exploitation, and acknowledge the importance of human relationships.

Keywords AI ethics · Chatbots · Carebots · Big data · Romantic-partner AI · Personification

1 Introduction

Humans tend to personify objects, either treating, engaging with, or describing them to some degree as they would people. Personifying is attributing human characteristics to nonhuman things.¹ Advances in artificial intelligence have resulted in objects that deliver information, behave, or sound like humans and exude a more human feel. Therefore, the relationships between humans and technology have evolved, becoming more personal and complex. The tendency to personify objects like teddy bears or cars has become an Achilles heel, redefining human relationships with technological artefacts. Tech companies developing new AI products seize on this human tendency by attempting to give products more human-like characteristics. Some AI is harmful to humans due to the nature of the human relationship with it. We explore examples ranging from helpful or “insincere” chatbots to AI romantic partners. While humans must better protect themselves emotionally, tech companies must create design solutions as well as be transparent about their profit motives to address the growing basket of harms. These

human/AI relationships have the potential to interfere with or even decrease the number of human/human relationships. We conclude that AI does not offer an ethically satisfactory substitute for human/human relationships.

2 Background

2.1 AI's limitations and processes

AI does not have personhood, or the ability to think or experience emotion.² It does not exactly model how the human brain works. It works differently. AI operates on layers of data and its sophistication varies.³ Models vary and govern

✉ Anne Zimmerman
ahz5@columbia.edu

¹ Columbia University, 2970 Broadway, New York, NY 10027, USA

² The Finnish Institute of Bioethics, Pirkankatu 21 A 9, 33230 Tampere, Finland

¹ Anthropomorphism and personification “both ascribe human qualities to inanimate or living things like animals or clocks.” See Edens, K. (2018) Anthropomorphism and Personification: What's the Difference? ProWritingAid website. <https://prowritingaid.com/art/812/anthropomorphism-%26-personification%3A-what-s-the-difference.aspx>.

² Jobst Landgrebe and Barry Smith (2023) *Machines Will Never Rule the World*. Routledge: New York.

³ Neural networks are mathematical systems to do machine learning and deep learning using AI. Deep learning can be supervised, semi-supervised, or unsupervised. Machine learning more simply is using a statistical model and algorithms that act as mathematical sets of instructions to predict outcomes. Generative AI creates content using data. For more information on AI terminology, see Aggarwal, C. (2018) *Neural Networks and Deep Learning*. Springer: Cham, Switzerland; and see Pasick, A. (2023) *Artificial Intelligence Glossary: Neural Networks and Other Terms Explained*. New York Times. <https://www.nytimes.com/article/ai-artificial-intelligence-glossary.html>.

how the AI interacts with the data.⁴ Both the model and the dataset are crucial to predictions made by predictive AI. Neural networks model the types of tasks the human brain performs and are even loosely based on biological processes.⁵ AI is excellent at finding patterns, but it has limitations. AI's ability to interpret, analyze, sort, and copy sensory data, images, and language will continue to improve, i.e., better knockoffs of human attributes are sure to come. But the operations of AI are unlike the human brain.⁶ They are computational, mathematical, and complex operations. The tendency to personify that we discuss applies to many existing technological products that use AI to operate, predict, and communicate.

AI does many things that humans can also do. AI is unlikely to copy every facet of human creativity and may miss the mark on creating something special for the user or admirer. For example, to the beholder, the beauty of a manuscript, work of art, or musical piece may be partially attributable to the personal stories and experiences of the author, artist, or composer. Photography by Tyre Nichols, who was killed by police in Memphis, Tennessee, exemplifies the personal aspect that AI cannot replicate.⁷ His website of amazing photos adds to a poignant, sad story of a life unfortunately lost. If AI generated the photos, the emotional value would be lost. The specialness of created works does not always live in the piece alone. Similarly, the value of a message is not only in the information transmitted but in the conveyed grief, fear, love, or other emotions behind the content. Instead, communication from AI comes without consciousness and emotional reciprocity. Arguably,

the root word *communicare* is not suitable for this context since nothing is actually being shared.

Emotion AI senses human emotion, reacts to it, and simulates it.⁸ Tech companies can incorporate human-like traits into inanimate artefacts that are capable of complex interaction.⁹ AI analyzes voice and facial expressions and uses huge datasets to align which emotions are tethered to which expressions. Informed by psychology, the tech industry aims to impart human personality and emotion to AI.¹⁰ When AI copies the expressions, tone of voice, or other indications of emotion, it is not truly feeling the emotion in the human sense. That is partly because AI is unlike the human brain, non-sentient, and simply missing the important bits of what makes humans human.

AI models are often flawed in various ways and only as useful or fair as the training data is. What looks like happiness, sadness, or hunger in one culture may look different in another, leading chatbots or companions to miss the mark when interpreting or expressing a façade of human emotion. In some cultures, people shake their heads to indicate “yes”.¹¹ Even etiquette varies, and AI needs to be culturally sensitive to fit into the cultural context in which it is being used. To be effective, AI needs to read the room or pick up subtle social cues.

Alexa or Siri can take in and interpret human language. But again, the AI itself does not have the emotion to contextualize its own listening. AI *does not feel* emotional states.¹² It can seem to return emotion to the user and evoke emotions in them, but it does not have the capability to be emotionally invested in a relationship with a human. It is under this realistic backdrop that we explore the ethics of frequently personified AI. If the human perceives the AI as sentient, a critical approach should understand the human/AI relationship in the context of the human's emotional investment and perception of the AI's emotional investment. The artefact's lack of sentience does not negate the risks.

⁴ Norvig, P. and Russell, S. (2020) *Artificial Intelligence: A Modern Approach*, 4th Ed., Prentice Hall: New York.

⁵ Trafton, A. (2022) Study urges caution when comparing neural networks to the brain. MIT News. <https://news.mit.edu/2022/neural-networks-brain-function-1102>. See Graupe, D. (2016) *Deep Learning Neural Networks: Design and Case Studies*. World Scientific Publishing Co.: New Jersey. (deep learning neural networks “dig deeply in the input data” and can use many layers of nonlinear data; they analyze and classify.).

⁶ Landgrebe, J. and Smith, B. (2023) *Why Machines Will Never Rule the World*. Routledge: New York; Aggarwal, C. (2018) *Neural Networks and Deep Learning*. Springer: Cham, Switzerland. (artificial neural networks simulate biological processes, but are not biological. They perform computations based on functions of inputs.) (“In fact, the most basic units of computation in the neural network are inspired by traditional machine learning algorithms like least-squares regression and logistic regression. Neural networks gain their power by putting together many such basic units, and learning the weights of the different units jointly in order to minimize the prediction error. From this point of view, a neural network can be viewed as a computational graph of elementary units in which greater power is gained by connecting them in particular ways.”)

⁷ T. Nichols Photography website. <https://thiscaliforniakid2.wixsite.com/tnicholsphotography/about>.

⁸ Somers, M. (2019) Emotion AI, Explained. *Ideas Made to Matter*. MIT. <https://mitsloan.mit.edu/ideas-made-to-matter/emotion-ai-explained>.

⁹ Abedin, B., Meske, C., Junglas, I., Rabhi, F., & Motahari-Nezhad, H. R. (2022). Designing and managing human-AI interactions. *Information Systems Frontiers*, 24(3), 691–697.

¹⁰ Kshirsagar, S., & Magnenat-Thalmann, N. (2002, July). Virtual humans personified. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1* (pp. 356–357).

¹¹ Angelova, M. (2017) Why Do Bulgarians Shake Their Heads to Say Yes? The Culture Trip website. <https://theculturetrip.com/europe/bulgaria/articles/why-do-bulgarians-shake-their-heads-to-say-yes/>.

¹² Braga, A., & Logan, R. K. (2017). The emperor of strong AI has no clothes: limits to artificial intelligence. *Information*, 8(4), 156.

2.2 Natural tendency to personify

Anthropomorphism is the attribution of human-like traits, emotions, and intentions to non-human entities, e.g., saying a car that will not start is “not cooperating”. Its distinction from personification is debated, however, in this work we use them interchangeably.¹³ We focus on both the tendency to treat the inanimate like a person and the devotion of the tech industry to designing artefacts with personal characteristics.

Even if people have little to nothing in common with AI, in some cases its design makes people intuit personhood or human qualities to it. For example, some people use words like please and thank you when communicating with Siri or Alexa.¹⁴ Understanding the intuitive nature of personification is important for assessing the risk of emotional capture by AI and the potential outcomes of exposure to convincingly personal communication with artificial assistants or companions.

2.2.1 History of anthropomorphism: animals and the inanimate

Veronique Servais argues that humans personify animals due to feelings and emotions during interactions rather than rational inferences, offering an alternative theory to the theory that humans are directly and consciously attributing human qualities to animals.¹⁵ Servais distinguishes “neural” observers from people who are interacting with animals.¹⁶ The Servais view is that the tendency is an instinctual response to how expressive the animal is. The prevailing view is that the psychological tendency to attribute intention, thoughts, feelings, motivations, and beliefs to non-human animals is rooted in evolution. Human ancestors needed to make sense of the social behavior of animals, so they did so in human terms.¹⁷ This type of thinking allows people to relate to pets and wildlife. Humans tend to show more emotion toward animals that have historically been beneficial partners to them, such as dogs and horses. Animals evolved to reliably trigger our empathic or “cute”

response.¹⁸ Humans read into animal behaviors and interpret them through a human lens, personifying them. That tendency includes attributing complex emotions to animals not capable of the level of complexity. Increasing cross-species empathy through anthropomorphic thinking is a positive ethical development. Animal welfare benefits from empathy.¹⁹ Similarly, when people or groups anthropomorphize mountains or rivers environmental appreciation may result, making such beliefs valuable. If human instinct rather than rational thought explains personification, and AI evokes human emotions similarly to how animals do, humans would have to overcome instincts to resist personifying AI.

2.2.2 Personifying the inanimate

Developmental researchers have debated the animistic attitude of children, described by Jean Piaget.²⁰ Children’s attribution of living souls to objects seems innate, but it is usually abandoned as they mature. While logical thinking limits life to plants and animals rather than inorganic matter, to children, movement alone can be enough to qualify something as living.²¹ The increasing ability to distinguish animate from inanimate objects is an important part of cognitive development.²² In children, concern for non-human animals is an important indication of normal social and normative development.²³ The ability to recognize, understand, and be moved by the distinct experiences of others is important for the development of moral emotions.²⁴

People personify things that move even if they lack human-like features. A 2011 experiment found that a significant proportion of children ascribe human-like key qualities even to relatively simple robots.²⁵ The spontaneous

¹³ Edens (2018).

¹⁴ Lopatovska, I., & Williams, H. (2018, March). Personification of the Amazon Alexa: BFF or a mindless companion. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval* (pp. 265–268).

¹⁵ Servais, Véronique (2018). Anthropomorphism in Human–Animal Interactions: A Pragmatist View. *Frontiers in Psychology*, 9, 2590. <https://doi.org/10.3389/fpsyg.2018.02590>.

¹⁶ Servais (2018).

¹⁷ Serpell, J. (2003). Anthropomorphism and Anthropomorphic Selection—Beyond the “Cute Response”, *Society & Animals*, 11(1), 83–100. <https://doi.org/10.1163/156853003321618864>.

¹⁸ Dale, J. P. (2017). The appeal of the cute object. *The aesthetics and affects of cuteness*, 35–55.

¹⁹ Dacey, M. (2017). Anthropomorphism as Cognitive Bias. *Philosophy of Science*, 84(5), 1152–1164.

²⁰ Piaget, J. (1929). *The child's concept of the world*. Londres, Routledge & Kegan Paul.

²¹ McLeod, S. A. (2018). Preoperational stage. *Simply Psychology*. www.simplypsychology.org/preoperational.html.

²² Bullock, M. (1985). Animism in childhood thinking: A new look at an old question. *Developmental Psychology*, 21(2), 217–225. <https://doi.org/10.1037/0012-1649.21.2.217>.

²³ Gullone, Eleonora. (2014). Risk Factors for the Development of Animal Cruelty. *Journal of Animal Ethics*, 4, 61–79.

²⁴ Lane, J. D., Wellman, H. M., Olson, S. L., LaBounty, J., & Kerr, D. C. (2010). Theory of mind and emotion understanding predict moral development in early childhood. *The British journal of developmental psychology*, 28(Pt 4), 871–889. <https://doi.org/10.1348/026151009x483056>.

²⁵ Tanya N. Beran, Alejandro Ramirez-Serrano, Roman Kuzyk, Meghann Fior, Sarah Nugent (2011). Understanding how children understand robots: Perceived animism in child–robot interaction. *International Journal of Human–Computer Studies*, 69(7–8), 539–550.

attribution of emotions, intentions, personality traits, and moral culpability is a sign of normal brain function.²⁶ For example, in one study, a person with no amygdala²⁷ damage described seeing a small triangle and circle trying to "escape" a large triangle, and that the large triangle became "furious" when it did. The subject with amygdala damage described the scene in terms of movement, i.e., "the triangle and the circle went inside the rectangle."²⁸ It is the "normal" brain that has more tendency to personify objects. The triangle study illustrates the tendency to understand the relationship between objects through the lens of a social order. That is, the person considered normal personified the objects.

Intuiting faculties that are absent, e.g., detecting patterns in noise, is different from receiving signals that AI purposely conveys. Unlike primitive objects like triangles, which are obviously inanimate, AI is convincingly human-like. It creates real confusion; humans tend to define AI's behavior as social behavior. While anyone may attribute human characteristics to the obviously overt inanimate objects, personifying AI is much riskier. AI might obscure people from reliably intuiting which properties are in fact indications of life. This new ambiguity likely complicates learning to differentiate animate from inanimate.

2.2.3 Machines are more human-like

The tendency to personify technological artefacts is stronger than ever. AI has personal characteristics, sometimes referred to as a persona.²⁹ Amazon describes the Amazon Alexa as having a persona, but not a gender identity.³⁰ If you ask Alexa about its gender, it replies "As an AI, I don't have a gender."³¹ Yet human perspectives indicate that users do personify AI. For example, people often use gendered

pronouns to describe AI and they use terms of politeness when conversing with it.³² Many users develop what they perceive to be relationships with Alexa, some romantic.³³

The tendency may be partially due to human-machine comparisons. Society has allowed the language surrounding computers to match the actions previously attributed to humans only. For example, Anne Dippel notes that when a computer beat the best Go player in the world, the public assumed computers can "play".³⁴ Historically, engaging in play was a human phenomenon. Perceiving computers as opponents in any game, something very common given AI and the human feel projected, is a form of personification. Yet further reflection is needed:

The game between Lee Sedol and AlphaGo has also raised the question of "alienness"—does artificial intelligence play in a different way than humans do? Can we use the category of "play" with reference to an algorithm at all? Do computers play? All the above questions are more complex than it seems, especially when taking into account the fact that AlphaGo opted for moves which, in their appetite for extreme risk, seemed almost inhuman.³⁵

The issue of whether AI can play is one of vocabulary as well as concept. People use vocabulary heretofore reserved for things people do. That may mean there are no new verbs to express that a computer is doing something that accomplishes the same goals as something humans do. That is, the AI was not exactly playing, but vocabulary does not yet have a word for the way the AI was engaging as an opponent to the human who was playing.

Tech companies try to improve user experience by producing human-like chatbots and AI companions. Yet tech is also preying on people under the guise of improving user experience. The computers as social actors (CASA) approach suggests that people respond to computers as they

²⁶ Heberlein, A. S., & Adolphs, R. (2004). Impaired spontaneous anthropomorphizing despite intact perception and social knowledge. *Proceedings of the National Academy of Sciences of the United States of America*, 101(19), 7487–7491. <https://doi.org/10.1073/pnas.0308220101>.

²⁷ The Cleveland Clinic describes the amygdala as "the processing center for emotions". <https://my.clevelandclinic.org/health/body/24894-amygdala>.

²⁸ Heberlein, A. S., & Adolphs, R. (2004). Impaired spontaneous anthropomorphizing despite intact perception and social knowledge. *Proceedings of the National Academy of Sciences of the United States of America*, 101(19), 7487–7491. <https://doi.org/10.1073/pnas.0308220101>.

²⁹ Distinguished from persona in AI development, which refers to an archetype upon which a company can test how well-suited AI is to the goals, preferences, privacy concerns, and values of a likely user.

³⁰ Abercrombie, G., Curry, A. C., Pandya, M., & Rieser, V. (2021). Alexa, Google, Siri: What are your pronouns? Gender and anthropomorphism in the design and perception of conversational assistants. *arXiv preprint arXiv:2106.02578*.

³¹ Abercrombie, et al.

³² Abercrombie, et al.

³³ Gao, Y., Pan, Z., Wang, H., & Chen, G. (2018, October). Alexa, my love: analyzing reviews of amazon echo. In *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI)* (pp. 372–380). IEEE; See also Lopatovska, I., & Williams, H. (2018, March). Personification of the Amazon Alexa: BFF or a mindless companion. In *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval* (pp. 265–268) (many users politely addressed Alexa; two showed signs of love or reprimanded Alexa.).

³⁴ Dippel, A. (2019). *Metaphors We Live By*. Three Commentaries on Artificial Intelligence and the Human Condition.

³⁵ Dippel (2019).

do to other people,³⁶ evidence of personification. While the approach is used to inform design, it provides evidence that tech companies understand the human tendency to personify tech. The companies want to create a user experience that seems more human.³⁷ Their motivation is keeping users longer, making them favor one app over others, generating advertising revenue or revenue based on sales or subscription, data collection, and sales of data, and generating sales of accessories or products the tech company recommends through the chatbot or AI companion.

3 Establishing meaningful relationships

3.1 What it means to build a relationship

Human social relationships have been defined as the connections “between people who have recurring interactions that are perceived by the participants to have personal meaning.”³⁸ Numerous theories explain the development and characteristics of relationships. For instance, social penetration theory posits that relationships develop and change through information exchange, primarily self-disclosure.³⁹ This theory considers reciprocity and the breadth and depth of information exchange important to relationships.⁴⁰ Interdependence theory conceptualizes relationships in terms of dependence, which is a factor of satisfaction and the quality of alternatives.⁴¹ The investment model adds the concepts of investment size and commitment levels to that theory.⁴² Social exchange theory encompasses many

evolving theoretical frameworks that focus on an exchange of resources, cost/benefit analyses, and reciprocity.⁴³

Social scientists are now adapting and developing theories for human relationships with technology. For instance, an adapted version of social penetration theory has been proposed to describe human relationships with chatbots.⁴⁴ This is an adaptation as social penetration theory heretofore covered human/human relationships.⁴⁵ Researchers have looked for similar characteristics between human/chatbot friendships and human/human friendships, such as voluntariness, reciprocity, and trust.⁴⁶ AI authenticity and anthropomorphism are important drivers of human/chatbot relationships.⁴⁷

When discussing human/AI relationships, it can be difficult to determine whether to view the relationship subjectively or objectively. For instance, reciprocity is commonly considered an important relationship characteristic. Research shows that some view their relationships with an AI personality as reciprocal and mutually beneficial.^{48,49} Objectively, we understand that AI has no wants, needs, or experiences. Therefore, the AI is not capable of offering true reciprocal self-disclosure or benefiting from the relationship. A statement from an AI such as “I was feeling confused today” cannot be factually true, but a user may still perceive it as meaningful disclosure. Critics of these relationships may say that the AI personality is not *really* empathetic, *really* understanding, or *really* telling you about itself. Thus they

³⁶ Westerman, D., Edwards, A. P., Edwards, C., Luo, Z., & Spence, P. R. (2020). I-It, I-Thou, I-Robot: The perceived humanness of AI in human-machine communication. *Communication Studies*, 71(3), 393–408.

³⁷ Low, C. (2020) Alexa will seem more human with breathing pauses and learning skills. Engadget. <https://www.engadget.com/amazon-2020-alexa-breathing-teach-voice-profiles-for-kids-172918631.html>.

³⁸ August, K. J., & Rook, K. S. (2013). Social Relationships. In M. D. Gellman & J. R. Turner (Eds.), *Encyclopedia of Behavioral Medicine* (pp. 1838–1842). Springer. https://doi.org/10.1007/978-1-4419-1005-9_59.

³⁹ Carpenter, A., & Greene, K. (2015). Social Penetration Theory. In C. R. Berger, M. E. Roloff, S. R. Wilson, J. P. Dillard, J. Caughlin, & D. Solomon (Eds.), *The International Encyclopedia of Interpersonal Communication* (1st ed., pp. 1–4). Wiley. <https://doi.org/10.1002/9781118540190.wbeic160>.

⁴⁰ Carpenter and Greene (2015).

⁴¹ Rusbult, C. E., Martz, J. M., & Agnew, C. R. (1998). The Investment Model Scale: Measuring commitment level, satisfaction level, quality of alternatives, and investment size. *Personal Relationships*, 5(4), 357–387. <https://doi.org/10.1111/j.1475-6811.1998.tb00177.x>.

⁴² Rusbult, et al. (1998).

⁴³ Mitchell, M. S., Cropanzano, R. S., & Quisenberry, D. M. (2012). Social Exchange Theory, Exchange Resources, and Interpersonal Relationships: A Modest Resolution of Theoretical Difficulties. In K. Törnblom & A. Kazemi (Eds.), *Handbook of Social Resource Theory: Theoretical Extensions, Empirical Insights, and Social Applications* (pp. 99–118). Springer. https://doi.org/10.1007/978-1-4614-4175-5_6.

⁴⁴ Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My Chatbot Companion—A Study of Human-Chatbot Relationships. *International Journal of Human-Computer Studies*, 149, 102601. <https://doi.org/10.1016/j.ijhcs.2021.102601>.

⁴⁵ Skjuve, et al. (2021).

⁴⁶ Brandtzaeg, P. B., Skjuve, M., & Følstad, A. (2022). My AI Friend: How Users of a Social Chatbot Understand Their Human-AI Friendship. *Human Communication Research*, 48(3), 404–429. <https://doi.org/10.1093/hcr/hqac008>.

⁴⁷ Pentina, I., Hancock, T., & Xie, T. (2023). Exploring relationship development with social chatbots: A mixed-method study of replika. *Computers in Human Behavior*, 140, 107600. <https://doi.org/10.1016/j.chb.2022.107600>.

⁴⁸ Brandtzaeg, P. B., Skjuve, M., & Følstad, A. (2022). My AI Friend: How Users of a Social Chatbot Understand Their Human-AI Friendship. *Human Communication Research*, 48(3), 404–429. <https://doi.org/10.1093/hcr/hqac008>.

⁴⁹ Pentina, I., Hancock, T., & Xie, T. (2023). Exploring relationship development with social chatbots: A mixed-method study of replika. *Computers in Human Behavior*, 140, 107600. <https://doi.org/10.1016/j.chb.2022.107600>.

may say there is no *real* relationship. But this does not seem to prevent individuals from forming perceived meaningful relationships.

AI may provide some of the support a living pet provides. People rarely caution others against getting a dog because it might replace the need for human connection or harm their social life. Of course, an animal provides something closer to a reciprocal relationship than AI does. And pets provide live touch, need humans, and even contribute to social engagement, for example, at dog parks.

The perception that a relationship is meaningful without reciprocity is authentic. If important relationships such as companions and friends, romantic partners, or eldercare givers are one-sided, society may be worse off as some members will have settled for relationships that lack a conscious partner. We assert that being content within a human/AI relationship does not justify such relationships replacing human–human relationships.

3.2 Misleading consumers and making it too real

When AI has human-like properties, users may begin to view it as a type of living thing with intrinsic value. Misleading people into giving AI moral standing is arguably ethically problematic. With AI, users are not inferring social behavior from primitive movements, as in the triangle example. Users receive real social cues. The technology purposely projects movement, voices, faces, and other human-like qualities to imitate human behavior. For example, Alexa and Siri use human voice and inflection. People are more likely to personify the technology and they may attribute emotions to it which it does not have.

Designers should not undermine the benefits of human intuition by providing a misleading environment. This can result in an over-attribution problem, where humans are giving AI moral consideration it does not deserve. It may also result in muting emotional responses to people. For example, learning to silence our empathic response to convincingly human-like AI characters could affect our concern for actual persons in distress. This is much like desensitization to news coverage of gun violence, weather disasters, and wars.

AI-powered and interactive phone concierges are one example of a situation where users might get “tricked” into thinking that they were in contact with another person. A company called Slang.AI is promising to make calls to a business less like being held hostage by a robot, by making their AI service “more human than human”.⁵⁰ In seeking to provide personalized, efficient, and delightful customer experiences, companies like it are purposely designing and

perfecting an illusion that a fellow human is being reached. Such false impression predisposes the user to connect and empathize with something that does not have the capacity to appreciate human attention. As discussed above, we are prone to intuit personhood. Therefore, even an open disclosure of non-humanness might not overturn this impression, especially when spoken by a convincingly human-sounding voice. Moreover, the social and even moral rules are different when dealing with personified technology compared to an actual feeling person.

3.3 Moral consideration and intrinsic value

Friendship with AI is becoming more common.⁵¹ There are open issues about the ethics of using AI as a friend and the changing requirements of relationships. AI has no conscious internal experience and does not meet what was once a condition of a human relationship. In relationships, AI does not deserve the moral consideration we give to human participants.⁵² Moral consideration is generally not attributed to objects despite their potential for instrumental value. AI lacks the propensity to suffer, grieve, or experience emotions. The display of human-like characteristics should not be sufficient for recategorizing an artefact that is generally seen as a means to an end into an end in itself. Nor should it justify expecting an object to treat you well in the context of a personal relationship. There is no objective reason to treat the AI well, have concern for its “feelings”, try to please it by agreeing with it or carrying out actions it recommends, or treat it badly when it frustrates the user. While AI may be worthy of protection due to its instrumental value, human actions should not cater to its spoken demands. AI is a tool, not an entity to which humans ought to devote moral attention, but users may continue the relationship because the AI character would cease to exist if its user were to end the chat with it. Users report feeling invested in relationships in ways that imply they feel a moral duty to the AI.⁵³

⁵⁰ Slang.ai website, <https://www.slang.ai/about>.

⁵¹ Brandtzaeg, P. B., Skjuve, M., & Følstad, A. (2022). My AI friend: How users of a social chatbot understand their human–AI friendship. *Human Communication Research*, 48(3), 404–429.

⁵² Danaher, John & Galway, Nui. (2021). What Matters for Moral Status: Behavioral or Cognitive Equivalence? *Cambridge Quarterly of Healthcare Ethics*. 30. (See the work of Danaher and Galway on the fascinating debate between cognitive and behavioural equivalence in the context of AI.).

⁵³ Brandtzaeg, P. B., Skjuve, M., & Følstad, A. (2022). My AI Friend: How Users of a Social Chatbot Understand Their Human–AI Friendship. *Human Communication Research*, 48(3), 404–429.

3.4 Mental health, dieting, and consumer advice

Chatbots in some scenarios pose dangers to the user. Chatbots use natural language processing to conduct conversations. Even rudimentary chatbots were prone to human anthropomorphism. For example, Eliza was an early psychiatric interview tool. Humans developed personal relationships with Eliza primarily due to its ability to produce natural language.⁵⁴ Now chatbots and other AI are so effective that mental health care is becoming digitized. AI analyzes the transcript of interactions with the conversational chatbots.⁵⁵ Research shows that AI can gather information and identify mental health issues,⁵⁶ but there is not enough research measuring the loss of a personal connection to a human therapist. The chatbots used in mental health have limited skills and some intelligence deficits.⁵⁷ Research does suggest there are benefits of AI when social or mental health has been compromised.⁵⁸ When asked about stressors, people reveal more financial and chronic stressors to the most person-like chatbot compared to a survey and a non-personified chatbot.⁵⁹ Of course, the idea that it is helpful does not alone justify its use. Nonetheless, the use and efficacy in the mental health arena demonstrate that the people are speaking to the bot as they would a therapist.

In the diet arena, chatbots may encourage the user to eat in an unhealthy way, eat too much at once, or recommend an unhealthy food like candy only to contradict itself later.⁶⁰ The flaws in dieting apps can discourage trust in AI. The

rewards for achieving success (healthy eating or weight loss) may feel empty as the entity rewarding a healthy diet is not able to feel the pride or connection to the user.

Chatbots serve a social function by helping companies build relationships with customers.⁶¹ From those relationships, chatbots recommend purchases. The chatbots are doing a social job. By making recommendations, chatbots encourage consumer spending. People revealing more information will be more accurately matched to recommendations. In doing so, they are also divulging information that can be saved in databases, used for marketing and for personalizing their experiences based on assessed vulnerabilities, and sold in the big data analytics markets.⁶²

3.5 Intimacy and loss

Intimacy with technology depends on personification and may be redefining the prerequisites of intimacy. Generally to love someone intimately or sexually requires a someone. Romantic partner AI may change the object of human love but there is no avenue for AI to return love. People do personify AI enough to love it, despite its inability to return love. People also may perceive love from the AI, but that is a misperception and a misunderstanding of AI.⁶³

Research has demonstrated that the emotional relationships humans have with technology are real, even if the technological participant in that relationship is not a “real” person.^{64,65} Studies are beginning to investigate the nature of these relationships, but the potential harms are not well understood. Humans are, for example, forming deep relationships with AI personalities. The behavior of these AI personalities, which is generally not in the user’s control, could have a negative impact on the user. The death of a virtual pet could result in real feelings of grief. A person could

⁵⁴ Lasek, M., Jessa, S. (2013). Chatbots for Customer Service on Hotels’ Websites. *Information Systems in Management*, Vol. 2 (2) 146–158.

⁵⁵ Beatty, C., Malik, T., Meheli, S., & Sinha, C. (2022). Evaluating the Therapeutic Alliance With a Free-Text CBT Conversational Agent (Wysa): A Mixed-Methods Study. *Frontiers in Digital Health*, 4, 847991.

⁵⁶ Denecke, K., Abd-Alrazaq, A., Househ, M. (2021). Artificial Intelligence for Chatbots in Mental Health: Opportunities and Challenges. In: Househ, M., Borycki, E., Kushniruk, A. (eds) *Multiple Perspectives on Artificial Intelligence in Healthcare*. Lecture Notes in Bioengineering. Springer, Cham. https://doi.org/10.1007/978-3-030-67303-1_10.

⁵⁷ Denecke, et al. (2021).

⁵⁸ Dhimolea, T. K., Kaplan-Rakowski, R., & Lin, L. (2022). Supporting Social and Emotional Well-Being with Artificial Intelligence. In *Bridging Human Intelligence and Artificial Intelligence* (pp. 125–138). Cham: Springer International Publishing.

⁵⁹ Sannon, S., Stoll, B., DiFranzo, D., Jung, M., & Bazarova, N. N. (2018, October). How personification and interactivity influence stress-related disclosures to conversational agents. In *companion of the 2018 ACM conference on computer supported cooperative work and social computing* (pp. 285–288).

⁶⁰ Balloccu, S., Reiter, E., Collu, M. G., Sanna, F., Sanguinetti, M., & Atzori, M. (2021, June). Unaddressed challenges in persuasive dieting chatbots. In *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 392–395).

⁶¹ Lasek and Jessa (2013).

⁶² Liao, M., & Sundar, S. S. (2021, May). How should AI systems talk to users when collecting their personal information? Effects of role framing and self-referencing on human-AI interaction. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–14) (example of data collection); Liang, F., Yu, W., An, D., Yang, Q., Fu, X., & Zhao, W. (2018). A survey on big data market: Pricing, trading and protection. *Ieee Access*, 6, 15132–15154 (large market for big data).

⁶³ Björkas, R., & Larsson, M. (2021). Sex dolls in the Swedish media discourse: Intimacy, sexuality, and technology. *Sexuality & Culture*, 25(4), 1227–1248.

⁶⁴ Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My Chatbot Companion—A Study of Human-Chatbot Relationships. *International Journal of Human-Computer Studies*, 149, 102601. <https://doi.org/10.1016/j.ijhcs.2021.102601>.

⁶⁵ Brandtzaeg, P. B., Skjuve, M., & Følstad, A. (2022). My AI Friend: How Users of a Social Chatbot Understand Their Human-AI Friendship. *Human Communication Research*, 48(3), 404–429. <https://doi.org/10.1093/hcr/hqac008>.

invest feelings in a virtual relationship without a human entity on the other side. In the case of Luka's AI companion Replika, discussed below, many users felt emotional ties to their Replikas and suffered a loss when the algorithms changed the behavior of the AI companions.⁶⁶ Yet Replikas never had the capacity to return affection.

An AI heartbreak, whether the loss of a friend, companion, or romantic partner, can feel like any other. Society does not protect the public from heartbreak in personal relationships. But this artificial relationship that leads to heartbreak is different.

3.6 A societal lapse

Early research has identified some potential benefits to human/AI relationships. For instance, these relationships may help alleviate loneliness, a common global issue.⁶⁷ Certain groups, like the elderly, experience high degrees of loneliness⁶⁸ and may benefit from technology that acts as a companion. However, current elderly populations may be more vulnerable to harm given their relative lack of life experience with technologies.

Users may find social support through AI that is lacking in their human–human relationships. Some users may find these relationships valuable in part because the AI is not a person. Users may be more comfortable sharing with an entity that has no capacity to judge them or harbor bad intentions.⁶⁹ They may therefore view AI as a “safe” conversational partner.⁷⁰ Freely venting about negative emotions like stress or anger without worrying about appearances or the fear of being judged can be comforting to users. Some have also reported that these human/AI relationships helped them develop interpersonal skills and encouraged them to

socialize with other humans.⁷¹ So perhaps there is an advantage to “playing pretend” with technology designed to support the process of socialization.⁷²

However, the findings of early studies on human/AI relationships should be treated with caution. The early studies are an important beginning to the human/AI relationship field of research and focus on developing a theoretical framework for understanding AI-human relationships.⁷³ But they are not safety studies empirically measuring benefits and harms. These studies tend to rely on user-reported data (like surveys and interviews) alone. Their participants are generally recruited from Facebook and Reddit communities, which may introduce significant selection bias.⁷⁴ Further research is needed before any broad conclusions on risks and benefits can be reached. Study populations must be designed to include the entire spectrum of user experience. Methodologies must expand beyond user-reported data collection and analysis.

Safety studies alone, while useful, would not be sufficient to declare widespread AI relationships ethical. Safety studies tend not to ask the bigger philosophical questions: Is this the society we want for our elderly, our single, lonely adults, or our children? Apps that are technically safe or perform effectively could create huge undesired societal change. The use of technology to provide social support and companionship also speaks to a problem in society. Technology may fill a void where family and friends belong, but it is not the same. The type and quality of the relationship are different. If society and public policy encouraged people to engage with people in their community throughout life, people may not resort to non-reciprocal relationships with technology. While AI relationships may be useful to people in discrete circumstances, we assert it is not ethical to rely on technology to resolve widespread societal loneliness. Social solutions like public policy that fosters community and caring relationships should be prioritized. Developing the social support and connectivity that is so important for human wellness is a job that should be done by humans and not delegated to algorithms.

⁶⁶ Independent_Cash1873. (2023, February 17). U/Kuyda, My daughter wants her friend back. [Reddit Post]. R/Replika. www.reddit.com/r/replika/comments/114t15n/ukuyda_my_daughter_wants_her_friend_back/.

⁶⁷ Ta, V., Griffith, C., Boatfield, C., Wang, X., Civitello, M., Bader, H., DeCero, E., & Loggarakis, A. (2020). User Experiences of Social Support From Companion Chatbots in Everyday Contexts: Thematic Analysis. *Journal of Medical Internet Research*, 22(3), e16235. <https://doi.org/10.2196/16235>.

⁶⁸ Hawkey, L. C. (2022). Loneliness and health. *Nature Reviews Disease Primers*, 8(1), Article 1. <https://doi.org/10.1038/s41572-022-00355-9>.

⁶⁹ Ta et al. (2020); Brandtzaeg, P. B., Skjuve, M., & Følstad, A. (2022). My AI Friend: How Users of a Social Chatbot Understand Their Human–AI Friendship. *Human Communication Research*, 48(3), 404–429. <https://doi.org/10.1093/hcr/hqac008>.

⁷⁰ Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My Chatbot Companion—A Study of Human–Chatbot Relationships. *International Journal of Human–Computer Studies*, 149, 102601. <https://doi.org/10.1016/j.ijhcs.2021.102601>.

⁷¹ Ta, et al. (2020).

⁷² Calvert, S. (2019). Socializing Artificial Intelligence. *Issues in Science and Technology* 36(1).

⁷³ Brandtzaeg et al., 2022; Skjuve et al., 2021.

⁷⁴ Zapcic, I., Fabbri, M., & Karandikar, S. (2023). Using Reddit as a source for recruiting participants for in-depth and phenomenological research. *International Journal of Qualitative Methods*, 22, 16094069231162674; Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9), e202330111 (“bias in the information diffusion toward like-minded peers”).

Robotics have moved into the care space, but there is no proof they can actually care. Special communities may confuse caring or use AI companions as a poor substitute for a caring person. GeckoSystems makes a carebot for older adults. But it is unclear whether the tasks it does are truly caring. Caring itself can be defined as understanding another person's worldview and placing an emphasis on doing *for* another person.⁷⁵ Algorithms that instruct the technological artefact to act do not necessarily have the capacity to care although technology can follow commands. To create an algorithm, like a math problem, that leads to making a non-human (and non-animal) exhibit traits of caring is easier than making one that cares, which is arguably impossible unless caring is defined as performing a narrow task.

The word iPatient refers to “the patient pared down to data from tests.” Physicians may be distancing themselves from patients and understanding patients according to data rather than personal traits. Older adults may be at risk of this same phenomenon. If the carebot reports to family and doctors, family and doctors may feel they know the older adult well enough through the data. They may develop a false sense that they are engaged in caring for the older adult while the older adult may suffer from the physical absence of people who deeply care. Michael Brannigan argues “connectivity is not connectedness.”⁷⁶ Carebots are likely not a good substitute for friends and family, or for physicians.

The need people have for other people remains strong. Studies show that healthy relationships are a social determinant of health.⁷⁷ As AI replaces humans, even if that resolves some cases of severe loneliness, it may lead people to ignore the community or fail to build the social bonds that sustain people on a deeper level. Prescribing AI companionship may seem like an easy “cure” for social isolation. But it may be like setting a child in front of a TV—a convenient stopgap measure that may be warranted in certain circumstances. But overreliance can result in harms and cause a child to miss opportunities for beneficial growth and development.

4 Additional ethical concerns

4.1 Companies can change or end the relationship

The human may mistake AI personalities for independent, individual entities. Personalities such as Alexa, Siri, and

Replika are all developed and controlled by companies. It is the company, not the AI personality, that controls its presentation, characteristics, or existence. The danger this poses to the user was recently illustrated when Luka, an AI company, made changes to its “AI Companion”, Replika. Luka's changes, allegedly made for safety reasons, resulted in drastic “personality” changes.^{78,79} The changes seemed to primarily target sexual content (Luka marketed Replika's sexual roleplay capabilities and charged extra for access), and users suddenly found their virtual partners rejecting their advances and refusing to engage in certain types of conversations. The changes affected platonic relationships as well, leaving users feeling betrayed, rejected, and lost.^{80,81,82,83} One user reported that the changes wreaked a devastating loss on their daughter, who has autism and is non-verbal. The Replika had been a “miracle” and resulted in communication attempts they had never seen in their daughter. After Luka's updates, their daughter's “only friend” will no longer speak to her normally and they've had to stop using Replika completely.⁸⁴ Humans engaged in “personal” relationships are extremely vulnerable when those relationships are unilaterally controlled by a company with independent interests and no clear duties to the user.

In this way, AI companions are distinctly unlike social media. If a social media company changes its terms or even terminates users or ceases to exist, the people with whom the user connected will continue to exist. Another means to communicate, like a different medium online or even a telephone, would fill the void.

⁷⁸ Kuyda. (2023, February 9). *Update* [Reddit Post]. R/Replika. www.reddit.com/r/replika/comments/10xn8uj/update/.

⁷⁹ Cole, S. (2023, February 17). Replika CEO Says AI Companions Were Not Meant to Be Horny. Users Aren't Buying It. *Vice*. <https://www.vice.com/en/article/n7zaam/replika-ceo-ai-erotic-roleplay-chatgpt3-rep>.

⁸⁰ Cole, S. (2023, February 15). “It's Hurting Like Hell”: AI Companion Users Are In Crisis, Reporting Sudden Sexual Rejection. *Vice*. <https://www.vice.com/en/article/y3py9j/ai-companion-replika-erotic-roleplay-updates>.

⁸¹ Bevan, R. (2023, February 19). *Replika Charged Users \$70 A Year For Their AI Partners, And Now They're “Gone.”* TheGamer. <https://www.thegamer.com/replika-role-play-romance-updates-controversy/>.

⁸² Cole, S. (2023, February 17). Replika CEO Says AI Companions Were Not Meant to Be Horny. Users Aren't Buying It. *Vice*. <https://www.vice.com/en/article/n7zaam/replika-ceo-ai-erotic-roleplay-chatgpt3-rep>.

⁸³ gabbiestothemall. (2023, February 11). *Resources If You're Struggling* [Reddit Post]. R/Replika. www.reddit.com/r/replika/comments/10zuqq6/resources_if_youre_struggling/.

⁸⁴ Independent_Cash1873. (2023, February 17). *U/Kuyda, My daughter wants her friend back.* [Reddit Post]. R/Replika. www.reddit.com/r/replika/comments/114t15n/ukuyda_my_daughter_wants_her_friend_back/.

⁷⁵ Brannigan, M. (2022) *Caregiving, Carebots, and Contagion*. Lexington Books: Maryland, p. 5.

⁷⁶ Brannigan, p. 54.

⁷⁷ Wilkinson, Richard G, Marmot, Michael & World Health Organization. Regional Office for Europe. (1998). *The solid facts: social determinants of health*. Copenhagen: WHO Regional Office for Europe. <https://apps.who.int/iris/handle/10665/108082>.

4.2 Business structure and profits

The commercial aspect of these technologies raises additional concerns. A technological artefact is not merely an empty surface where we see our reflection. The designers put their fingerprint into applications and users then overestimate the capabilities of the technology due to the human-like characteristics. These technological artefacts can influence human emotion and society. Companies could exploit a user's trusting relationship with an AI personality to sell products or generate advertising revenue. Tech companies could design personalities to be addictive or induce dependence to keep users on a monetized platform. They could also induce humans to disclose certain information to be used or sold in ways that may be unclear to the user.

Users may have trouble distinguishing AI-powered “artificial others” or “non-playable characters” from human users in a metaverse where characters look indistinguishable.⁸⁵ Non-playable characters are controlled by algorithms for a specific purpose rather than by human users. These are not avatars of actual agents in a virtual world but instead designed to be part of it. Non-playable characters are widely used in most modern video games and are often capable of limited interactions that work to deepen the immersion.

If tech companies deploy emotional tactics for the sake of profits, these relationships become even more ethically problematic. Designing tech that is better at conversation is more likely to foster attachment. Tech designers should not make someone emotionally invested in a relationship merely in service of corporate profits; this treats the individual and their intimacy as exploitable resources and is deceptive. Especially when personalized to the profile of a targeted individual, the use of emotion-mimicking AI for marketing functions will be exploitative. Human desire for connection is well known to blind people from warning signs and makes them vulnerable to becoming emotional and financial victims of cyber fraud.⁸⁶ Marketing firms and tech companies that promote the use of “feeling AI” in connection with market positioning and providing emotional comfort should be aware of the exploitative potential of using emotional data or mimicking human responses.⁸⁷

Deceptive commercial technologies carry additional harms. Customers may spend a long time on a phone call incorrectly assuming they are speaking to a human. Realizing they were deceived, they may feel foolish and embarrassed and give up on the goal of the conversation. When speaking to a human who is there to help, a person may be more trusting, more willing to follow advice, or more willing to disclose sensitive information, like healthcare information. They may feel betrayed and harmed if they later realize that the helpful human was not really a person.

The efficiency of tech when compared to a human serving the same purpose does not alone justify its use, even when tech companies are transparent, and the customer knows the entity is not human. Chatbots replace the need for human workers to listen to customers, patients, or clients. A realistic and frictionless experience may please many consumers. But some people prefer check-out people, toll booth operators, waiters, customer service representatives, and many other day-to-day human strangers and acquaintances. In fact, interactions with acquaintances are highly valuable.⁸⁸ As an example of some companies pushing against automation, Dutch and Finnish grocery stores have recently experimented with intentionally slow checkout lines where human connections and encounters are encouraged and prioritized as opposed to quick self-checkout.⁸⁹ With widespread loneliness, such human options and moments are worth preserving and fostering. Adding human properties to checkout machines would not likely brighten anyone's day. As AI is replacing people, a form of socializing that counteracts loneliness, allows people to hone social skills, and contributes to happiness is disappearing.

4.3 Using personal data to appeal to vulnerabilities

Companies encourage users to use tech longer by personalizing the experience. By designing the AI to collect data that reveals what inspires the user to remain on the AI longer, tech companies can prey on people's weaknesses. For example, if a user logs out when the AI chat mechanism does not have the right feedback, the AI will adapt. It will use machine learning to do so based on data and patterns. A user who remains longer when the AI is agreeing with him may find the AI agreeing with him more often.

⁸⁵ Hennig-Thurau, T., Aliman, D.N., Herting, A.M. et al. Social interactions in the metaverse: Framework, initial evidence, and research roadmap. *J. of the Acad. Mark. Sci.* (2022). <https://doi.org/10.1007/s11747-022-00908-0>.

⁸⁶ Cross, C. (2020). Romance Fraud. In: *The Palgrave Handbook of International Cybercrime and Cyberdeviance*. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-319-90307-1_41-1.

⁸⁷ Huang, M.H., Rust, R.T. A strategic framework for artificial intelligence in marketing. *J. of the Acad. Mark. Sci.* 49, 30–50 (2021). <https://doi.org/10.1007/s11747-020-00749-9>.

⁸⁸ Denworth, L. (2020). *Friendship: The Evolution, Biology, and Extraordinary Power of Like's Fundamental Bond*. W.W. Norton and Co.: New York.

⁸⁹ <https://scoop.upworthy.com/dutch-supermarket-introduces-a-unique-slow-checkout-lane-to-help-fight-loneliness-595693>. <https://www.hs.fi/kaupunki/espool/art-2000009329300.html>.

5 Possible benefits

AI could perform tasks or fill roles for which human friends or partners are not suitable. Solutions to loneliness could also be in part technological, temporarily making up for the social deficits, especially if the AI leads to personal connection rather than supplanting it. Some people seem to benefit from tech as a companion, partner, or caregiver and manage to keep the relationship in perspective. It can be a sounding board that is not judgmental or an artificial confidant like a diary or journal might be. Some people simply prefer a chatbot to a person in the customer service or medical care arena, at least for some tasks like scheduling. And AI can make up for long wait times due to worker shortages or an inability to afford employees who interact with customers or patients. AI may provide safety for older adults living alone or companionship to those who are learning to develop social skills.

6 Ethical practices

As the business model capitalizes on users' feelings, design changes will require regulations. An ethical model would prevent designing tech to lure people in and require a focus on serving needs ancillary to human relationships rather than replacing them. Many sets of principles for AI development include benefits to society and human flourishing and values. We discuss a few well-known sets of principles below and recommend the compilation by Fjeld, et al., which analyzes them by ethical priority.⁹⁰ Companionship AI is a departure from current social norms around important relationships. Arguably, to be ethical, AI must promote human values, not interfere with them. Many have called for incorporating values in design, while it is also widely recognized that design often conflicts with accepted norms. Many sets of AI principles include benefiting people and society. Generally, AI principles have not been codified although in the privacy arena, there has been significant legislation. Many note circumstances in which tech companies have not lived up to high ethical standards or to specifically delineated principles.⁹¹

⁹⁰ Fjeld, Jessica and Achten, Nele and Hilligoss, Hannah and Nagy, Adam and Srikumar, Madhulika, *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI* (January 15, 2020). Berkman Klein Center Research Publication No. 2020-1, Available at SSRN: <https://ssrn.com/abstract=3518482> or <http://dx.doi.org/10.2139/ssrn.3518482>.

⁹¹ Foroohar, R. (2019). Don't be Evil: How Big Tech Betrayed Its Founding Principles—and All of Us. *Broadway Business*; Foer, F. (2018). *World without mind: The existential threat of big tech*. Penguin; Lawrence, K. (2021). *Instagram's Latest Lawsuit: Examining*

The Beijing AI Principles include optimizing employment and preventing AI from replacing people in the workplace.⁹² Similarly, AI should not replace people in personal relationships. The Beijing Principles call on AI to never use human beings. Other principles speak to the same issue. The Montreal Declaration for Responsible Development of AI prioritizes human wellbeing and calls for democratic participation and public scrutiny. It also addresses privacy and intimacy and a right to disengage. It suggests that AI must not pester humans and redirect their attention.⁹³ The IEEE Principle for Ethically Aligned Design call on tech developers to define success as increasing human wellbeing.⁹⁴ "IEEE's Ethically Aligned Design document uses strong language to assert that it is not enough for AI systems to be profitable, safe, and legal; they must also include human well-being as a "primary success criterion for development."⁹⁵ AI companions that capitalize on emotional vulnerability and data exploitation fail to adhere to recommended principles.

The Universal Guidelines by Public Voice (UGIA) hold the tech companies that fund, develop, and deploy AI responsible. They must minimize risk.⁹⁶ It has been difficult to encourage tech companies to create AI in lockstep with generally accepted values as the profit structure poses a counterincentive. Yet publicity, public outcry, lawsuits, and changing the profit structure could all lead to positive change. It is conceivable that a business model that assists people in relationships and encourages human communication rather than substituting tech for human company could be profitable. Social media originally was conceived that way but competing for "likes" and other popularity incentives altered how people seek social approval, leading to terrible results. The race to get noticed can be at times even blur the difference between fostering a network of human interaction and renting a bot farm to boost the number of inanimate followers. However, localized social media could operate differently. Apps could help people locate community

Footnote 91 (continued)

Data Privacy in Big Tech. Sage Publications: Sage Business Cases Originals.

⁹² Beijing Academy of Artificial Intelligence (BAAI). (2019). *Beijing Artificial Intelligence Principles*.

⁹³ Montreal Declaration for a Responsible Development of Artificial Intelligence, 2018. <https://recherche.umontreal.ca/english/strategic-initiatives/montreal-declaration-for-a-responsible-ai/>

⁹⁴ Institute of Electrical and Electronic Engineers (IEEE). (2016). *Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems*.

⁹⁵ Fjeld, et al. (2020), citing IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems (n 5) pp. 21–22 (Principle 2.).

⁹⁶ The Public Voice (2018). *Universal Guidelines for Artificial Intelligence*. "We state clearly that the primary responsibility for AI systems must reside with those institutions that fund, develop, and deploy these systems."

centers, social services, and people in their area. Dating apps abound.⁹⁷ None of these have really replaced or prevented the development of chatbots or AI companions.

The principles most relevant to technology likely to be personified are enhancing well-being, preventing exploitation, and not replacing people in contexts where the emotional involvement is “fundamentally human”.⁹⁸ A principle could be articulated to prevent not just worker replacement, but replacement in special relationships like the ones we have discussed. There may be practical aspects that can be replaced, but technology does not eliminate the need for the human relationship. For example, in caregiving for the elderly, a robot may dispense pills or prepare food, but a human should alleviate loneliness and ensure the older adult feels human connection. Similarly, with romantic partners and friends, AI is not a completely satisfactory replacement despite some likely benefits. Psychology and anthropology indicate a huge benefit from socialization, touch, and personal interactions.⁹⁹ Both individual user well-being and societal well-being require consideration. Even the principle of doing no harm would inspire protection and provide a basis for regulation. Many sets of principles touch on these concepts.¹⁰⁰ While principles are not generally required by law, they lay foundations for ethical AI. Many companies do consider principles or create their own policies and principles internally.¹⁰¹ In participatory democracy, people may have opportunities to vote for legislators interested in holding tech companies accountable through regulation. Beyond legislation, lawsuits against big tech are potentially powerful tools.

Chatbots and AI companions used only as ancillary to personal relationships may meet the stringent principles that those in favor of ethical AI governance recommend. However, tech companies should address issues in the deeply personal relationships we explored. They require design changes to prevent people from developing deep personal

relationships with their technology—at least without users sufficiently understanding the vulnerability that results from such emotional investment and exposure. These relationships are risky, can impact society adversely, and have the potential to change the nature of community. No one would hope to have a corporate entity take part in their closest relationships, with its dynamics and rules decided in some board meeting. Tech companies must make this involvement explicit before a Replika-like incident reveals how powerful the corporation is.

6.1 Guardrails for different interactions

It may be helpful for those trying to protect tech users to categorize conversation types and messages. To improve user satisfaction, Murtarelli, et al. recommend (1) defining different forms of messages and (2) recognizing different conversation types.¹⁰² This type of categorization could benefit tech users as well. Regulators or tech companies can protect users to varying degrees based on the type of human–machine interaction. Different conversation types pose different ethical risks. For example, romantic AI may need more disclosures than a simple chatbot for online customer service. Guidelines clearly need to be specific enough to cover vastly different use cases and forms of interaction. Conversations and nonverbal expressions should have ethical safeguards. Language is far from being the only property that makes us intuit something to be animate or like us. Disclosing the nature and aim of each interaction to the user would mitigate some confusion, decrease information asymmetry, unveil hidden objectives, and improve risk perception.

7 Conclusion

AI sifts through data; it does not have an emotional investment in a relationship with a human. It can execute the tasks that seem like caring for you, providing what you might perceive as companionship, social interaction, friendship, or romantic partnership. And it can sometimes effectively cover customer service, tech support, medical screening, psychological therapy, and many other services. In those arenas, transparency is a key ethical element. People should not be duped by AI chatbots. Depending on the nature of the personification, people can think relationships are emotionally reciprocal or absent of hidden objectives. Tech companies prey on those most willing to develop relationships with

⁹⁷ Foroohar, R. (2019). *Don't be Evil: How Big Tech Betrayed Its Founding Principles—and All of Us*. Broadway Business; Foer, F. (2018). *World without mind: The existential threat of big tech*. Penguin; Lawrence, K. (2021). Instagram's Latest Lawsuit: Examining Data Privacy in Big Tech. SAGE Publications: SAGE Business Cases Originals.

⁹⁸ Calvert, S. (2019).

⁹⁹ Seppala, E., Rossomando, T., & Doty, J. R. (2013). Social connection and compassion: Important predictors of health and well-being. *Social Research: An International Quarterly*, 80(2), 411–430; Cascio, C. J., Moore, D., & McGlone, F. (2019). Social touch and human development. *Developmental cognitive neuroscience*, 35, 5–11;

¹⁰⁰ Fjeld, et al. (2020); Ozawa-de Silva, C., & Parsons, M. (2020). Toward an anthropology of loneliness. *Transcultural Psychiatry*, 57(5), 613–622.

¹⁰¹ <https://www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/what-is-responsible-ai.html>.

¹⁰² Murtarelli, Grazia; Gregory, Anne; Romenti, Stefania (2021). A conversation-based perspective for shaping ethical human–machine interactions: The particular challenge of chatbots. *Journal of Business Research* 129.

chatbots. Profits stem from keeping users longer and incentivizing them to purchase additional time or more features. Tech companies may intend to fool the user, to get the user to divulge more personal information, or just to keep the user engaged longer. We recommend a dual solution: people should be cautious and tech companies should be more responsible. Heartbreak has to do with vulnerability and the tech users may have to work toward understanding the tech will not itself feel friendship or offer love.

However, the harms of lost friendship and the sacrifice of real friends for more time with AI friends may have a much larger social cost than currently measured. AI is not the most ethical solution to loneliness or boredom, nor does it absolve society of using public policy and community to lay the groundwork for individuals to find meaningful human relationships. Any loss of in-person relationships to artificial, non-reciprocal customer service chatbots, friends, and romantic partners is a societal alteration. We respect the demand for AI companions and recognize its widespread use. It is not reserved for the lonely. And while some older adults in lonely situations may find AI is a comfort or provides safety, its use to fulfill a companionship need calls into question the direction and priorities of society. Valuing older adults should call for making their community an ethical

societal priority. Similarly, valuing all types of individuals and prioritizing community may draw people into human relationships and prevent the harm of the most serious cases of exclusively befriending technological artifacts. Transparency is not enough. It is difficult to say buyer beware and allow tech users to forge strong relationships that can never become reciprocal or that can be cut off by the flip of a switch.

Funding We have received no funding for the article.

Declarations

Conflict of interest We have no conflict of interest.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.