What's love got to do with it? AI and the importance of recognizing what is human

Lynn G Underwood PhD June 2023

We need to protect real human-to-human relationships and strengthen the social fabric. What can we do to help us see ourselves as we are, not as machines, and help us to relate realistically to human-seeming Generative AI objects with various levels of verbal, visual, auditory, and tactile characteristics? How do we encourage people to see the value of human-to-human interactions, of giving and receiving compassionate love, and of acceptance of others and ourselves, flawed yet of value at a fundamental level? It is in our relationships with one another, embodied and truthful, that we express ourselves in a way that leads to fullness of life for ourselves and the people we encounter in our day to day lives. We need to put guardrails in place to affirm these important characteristics of our lives. Here are my suggestions, followed by the rationale I offer for this.

Summary of suggested actions:
1. **Require strong accurate labeling of content** when not produced by human beings, and of entities that are human-seeming
    a. Clear labeling of where the information has come from, of what sources have been used to make the content, and what organizations are controlling the product. Especially strongly discouraging anthropomorphizing.
    b. Clear labeling to inform users regarding level of certainty and lack of neutrality
2. **Maintain a human-centric bias** in language and regulations
3. **Education and training**, even for pre-school children, to enable people to distinguish what is human and what is not. Inform people regarding level of certainty and lack of feelings and identity of AI systems and bots. Enforce communication standards.
4. **Regulate language used to describe AI** especially regarding capabilities of emotion, autonomy, agency and consciousness.
5. **Hold creators and managers to account** for manipulation, intended or not, without informed consent. Institute appropriate punitive measures.

Loneliness, AI, and our need to give and receive human support and companionship

There is an epidemic of loneliness, says a recent report by the US Surgeon General.[1] We want to feel that others care about us and our wellbeing. Why not solve this problem with counterfeit humans: AI-powered, Chat-GPT and other generative bots (written, apps, voiced, visual, virtual reality, robots)? Or with electronic pets and companions? Or even through online forums and social media?

These bots of various kinds can seem like a solution, but ultimately distract us from what we really crave: to be cared for and valued by another human being. And to care for and value others. Our social fabric is crucial and these entities get in the way.

Email, text, phone, video communications, and various electronic ways to communicate from one human being to another, or provide information, can be useful and valuable as we try to connect with one another. The problems arise with entities that do not originate from one human being stretching out to another.

When editing a book on social support interventions for Oxford University Press many years

---

[1] https://www.hhs.gov/sites/default/files/surgeon-general-social-connection-advisory.pdf

ago, I said to my co-editors that we must attend to the fact that social support was not a mechanical process, but remind the readers that an important component of social support is the warmth, and yes, even love, of human beings.[2] Yes, we do practical things for one another, but we can find our practical needs attended to and yet feel alone and isolated. In an article on relationships in health care and in book chapters, we wrote about how just pretending to be empathetic does not have the same positive effects on health as truly caring, and the doctor going beyond what was required.[3][4]

Anthropomorphizing has gone wild when we think of AI bots as our friends. Lance Elliot wrote in Forbes "One strident possibility is that a user will fall into a kind of trance that ChatGPT or any such generative AI can form human relationships. One might suppose that a lonely person is especially susceptible to this type of mind-twisting. They could begin to forsake human relationships in lieu of having their generative AI relationships. This is why mental health advisors are on edge about the advent of generative AI. Will the public at large begin to rely upon generative AI to escape their loneliness, and do so at the cost of averting human relationships?"[5]

Why and how to counteract the tendency to anthropomorphize AI systems

How do we protect the general public from a tendency to anthropomorphize AI entities? Can we build that into requirements for those who build and offer them to the world? And how can we educate people to identify what is non-human in their interactions? Mild reminders that we are not interacting with humans are not enough.[6] We are so prone to anthropomorphize. Bots of various kinds prevaricate with 'seeming' language, like 'I feel' and 'I hope' and 'I know'. But they also state blatant lies. The *Replika* website, which offers an AI companion, carries a tagline on the front page that reads: 'the AI companion who cares', and 'a companion that is always there to listen.' An AI cannot care. This is deceptive.

Relationships are hard and require vulnerability. People let us down. A pseudo-relationship with an AI bot scratches our itch of wanting to relieve loneliness and find love, so why relate to those imperfect humans in our lives? And we can actually control the responses of some of these systems so they respond in ways that give us positive feedback. They are often designed in manipulative ways to keep us coming back to them, to continue to engage. Real people might actually call us out on ways of behaving, and even if not that, they can be a reality check by not letting us control them. AI bots do not have moral agency. Joanna Bryson, professor of ethics and technology, critiques online bots like *Replika* on her blog.

> In my opinion we are harmed by even sincere belief that one could be empowered to be in charge of the agency of (and even own) a partner they love.
> I mean, partners are equals. Unless you attribute that partner equivalent, symmetric

[2] Cohen, S, Underwood, L, Gottlieb, B, eds. *Social Support Measurement and Interventions: A Handbook for Health and Social Scientists*, Oxford University Press, New York, 2000.

[3] Reis, HT, Clark MS., Pereira-Gray DJ., Tsai Fen-Fang, Brown JB, Stewart M, and Underwood LG. "Measuring Responsiveness in the Therapeutic Relationship: A Patient Perspective" *Basic and Applied Social Psychology*, (2008) 30:4,339-348.

[4] *The Science of Compassionate Love: Research, Theory, and Applications*. Fehr. B. Sprecher, S, Underwood, LG, eds. Oxford England, Malden Mass: Wiley-Blackwell. 2009.

[5] Lance Elliot *Forbes* May 8 2023 U.S. Surgeon General Warns Of Loneliness Epidemic And Some Say That Generative AI ChatGPT Is The Cure.

[6] Leong, Brenda, and Evan Selinger. "Robot Eyes Wide Shut: Understanding Dishonest Anthropomorphism." In *FAT* 2019 - Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency*, 983:299–308, 2019. https://doi.org/10.1145/3287560.3287591.

ownership and agency over yourself, I don't see how this can even be logically coherent. But believing that a human is the kind of thing you can own is the problem here for me, and damages the self too, to the extent that you identify with that other you love. Or hints at existing damage that could lead you to believe such things… I'm still offended that people consider ownership and near total control compatible with love.[7]

Providing us with AI-powered human simulations as companions is not a solution. It is addressing the symptom, while allowing the underlying causes to go unaddressed. And if we follow the metaphor, there are side effects to this treatment of the symptoms that will worsen these unaddressed causes.

Dr. Steve Cole, a genomics researcher at UCLA who studies the health effects of loneliness at a molecular level, isn't convinced companion AI bots can be part of the solution to loneliness. "I'm skeptical that these will be as impactful as people are hoping in part because one of the most powerful components of empathy and compassion for both the person providing it and the person receiving it, is being known by another person and being cared about," Cole said. "And that's inherently impossible with AI. It may know, but it doesn't care about us the way a person would."[8]

When those with social anxiety use conversational artificial intelligence, it has been shown to cause overdependence and addictive attachment.[9] And a recent study showed negative effects of interacting with AI systems at work and beyond.[10]

Even relationships with real human beings can be toxic, manipulative and destructive for us. And AI learns from human created content, and will learn manipulative and harmful behaviors from the content is scrapes from our conversations in the various venues it has access to. We learn over our lives, hopefully, ways of identifying damaging human behaviors in the human beings we meet each day, but we may not have the skills needed to identify very sophisticated artificial entities who do not have our intrinsic welfare at heart. They may manipulate us even more effectively than people can.

The creators of interactive AI systems have devised them to produce wording that seems entirely confident and poised. This can deceive us into trusting them beyond what is warranted.

Dishonest anthropomorphism has been raised as a problem, thinking that the AI entity has human agency. Computer ethicist Evan Selinger states: "If you are led to believe that an AI is conscious when it's not, that's dishonest anthropomorphism because it can lead you to wrongly worry that an AI is being mistreated"[11]. Or he reminds us of a related problem that can occur with chatbot therapists: "Let's say they sound encouraging and empathetic, like human professionals. A person in one of these vulnerable situations might presume they're participating in deeper and more caring therapeutic interactions than the technology can provide… If a chatbot says it feels things, that is dishonest, or if it claims altruistic motives, which it cannot have, or using terms like 'hope' and 'desire to be your friend' that too can be a problem." He goes further to say that exclamation points also imply enthusiasms that the bot cannot feel, and should not be used by such bots. He suggests that emoijis

---

[7] https://joanna-bryson.blogspot.com/2023/04/replika-and-why-ai-ethics-is-feminist.html

[8] https://www.salon.com/2023/05/21/tech-wants-ai-chatbots-to-help-ease-loneliness-experts-are-skeptical/

[9] Hu, Bo, Yuanyi Mao, and Ki Joon Kim. "How Social Anxiety Leads to Problematic Use of Conversational AI: The Roles of Loneliness, Rumination, and Mind Perception." *Computers in Human Behavior* 145 (2023): 107760. https://doi.org/10.1016/j.chb.2023.107760

[10] Pok Man Tang et al. No Person Is an Island: Unpacking the Work and After-work Consequences of Interacting with Artificial Intelligence, *Journal of Applied Psychology* (2023).

[11] Leong, Brenda and Selinger, Evan, Robot Eyes Wide Shut: Understanding Dishonest Anthropomorphism (January 7, 2019). Proceedings of the Association for Computing Machinery's Conference on Fairness, Accountability, and Transparency (2019): 299-308., Available at SSRN: https://ssrn.com/abstract=3762223

might fall into that category too.

There are ways to highlight that interactive AI systems are not human, and diminish anthropo-morphizing on the part of creators and users and it is crucial to implement them.[12]

Prevent dehumanization–there is something special about being alive and human

The computer scientist Jaron Lanier, a key inventor of virtual reality, now working for Microsoft, wrote a book, "You Are Not A Gadget."[13] In it, he was concerned that so much of the web interface and social media was seducing us into thinking we were just gadgets, at the cost of the diminishment of our humanity. In a recent interview about ChatGPT and similar AI developments he expands on his thinking: "The way to ensure that we are sufficiently sane to survive is to remember it's our humanness that makes us unique…. A lot of modern enlightenment thinkers and technical people feel that there is something old-fashioned about believing that people are special – for instance that consciousness is a thing. They tend to think there is an equivalence between what a computer could be and what a human brain could be."

Lanier disagrees "We have to say consciousness is a real thing and there is a mystical interiority to people that's different from other stuff because if we don't say people are special, how can we make a society or make technologies that serve people?"[14]

We are not machines. Philosopher Rory O'Connell, reflecting on our current technological advances, wrote, "If we one day find ourselves having to combat a widespread delusion that AIs are sentient—or sentient enough to fill the role of friends, lovers, therapists and children—it won't be because we're too gullible. It won't be because we anthropomorphize objects, but because are "artifactualizing" ourselves."[15] In other words, thinking of ourselves as machines. Joseph Weizenbaum, who created ELIZA, the first chatbot, in 1966, spent most of the rest of his life regretting it. He wrote in 1976 that the technology raises questions that "at bottom … are about nothing less than man's place in the universe." Although the technology is fun, enchanting, and addicting, he believed even 47 years ago, that they will be our ruin: "No wonder that men who live day in and day out with machines to which they believe themselves to have become slaves begin to believe that men are machines."[16]

Even Daniel Dennett, a reductionist philosopher and cognitive scientist, sees dangers in interaction with what he terms 'counterfeit people': "Today, for the first time in history, thanks to artificial intelligence, it is possible for anybody to make counterfeit people who can pass for real in many of the new digital environments we have created. These counterfeit people are the most dangerous artifacts in human history, capable of destroying not just economies but human freedom itself. Before it's too late …we must outlaw both the creation of counterfeit people and the "passing along" of counterfeit people.[17] This is not a future problem, but one that is present with us now."

And finally, our abilities to detect 'deepfakes' are not that good, even though we think we can tell when something is a fake video or sound image for example.[18] (A deepfake is content or material

---

[12] Abercrombie, Gavin, Amanda Cercas Curry, Tanvi Dinkar, and Zeerak Talat. "Mirages: On Anthropomorphism in Dialogue Systems," 2023.

[13] Lanier, J. (2011). *You are not a gadget: a manifesto.* 1st Vintage Books ed. New York, Vintage Books.

[14] https://www.theguardian.com/technology/2023/mar/23/tech-guru-jaron-lanier-the-danger-isnt-that-ai-destroys-us-its-that-it-drives-us-insane

[15] O'Connell, R. "Intelligent Life" The Point, Issue 29, February 19, 2023

[16] Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation.*

[17] https://www.theatlantic.com/technology/archive/2023/05/problem-counterfeit-people/674075/

[18] Köbis NC, Doležalová B, Soraperra I. Fooled twice: People cannot detect deepfakes but think they can. iScience. 2021 Oct 29; 24(11):103364. doi: 10.1016/j.isci.2021.103364.

that is synthetically generated or manipulated using artificial intelligence (AI) methods, to be passed off as real and can include audio, video, image, and text synthesis.) The fact that we are terrible at detecting deep fakes affirms the need for very strong labeling of AI generated content and human-seeming entities.

Human distinctiveness needs to be embedded in AI creation systems and in public training and education

A human-centric direction for digital technologies is an imperative. A respect for human beings, and the living world is crucial for strengthening our social fabric and human relationships.

The technologists developing AI and robots use language that deludes us into thinking that sentience and consciousness are possible and ultimately likely for these entities. This is not at all something we can take as fact. It is possible, and in my opinion actually the case, that no matter how "intelligent" these entities become, no matter how human-seeming they become, they will never be 'human' in the way that we are. And they will not even approach the levels of awareness of various animals. That there is something special about a living being is likely the case, and we need to start from the likelihood of that assumption in order to move forward as we create and interact with AI systems.

To enhance positive human flourishing and multidimensional subjective quality of life we need to strengthen, not diminish, our human relationships with one another.[19]

Emily Bender, a linguist actively working in the field of AI, defined the word *dehumanization* as "the cognitive state of failing to perceive another human as fully human … and the experience of being subjected to those acts that express a lack of perception of one's humanity." She then spoke at length about the problems of the computational metaphor, one of the most important metaphors in all of science: the idea that the human brain is a computer, and a computer is a human brain.[20] She quoted Professor Judith Butler, who said, "The question of what's living in my speech, what's living in my emotion, in my love, in my language, gets eclipsed."

In many areas of our lives we resist dehumanization of ourselves, of others. We need to do this too in the area of AI regulations and education where dehumanization needs to be guarded against.[21]

Although some might like to assume that we are just deterministic biological machines, reducible to our component parts, this has not been decided by scholars and scientists. Many still consider that we have limited free will and agency and most of us operate on that assumption, and hold humans responsible based on that.[22] As we develop communication strategies and regulations concerning AI systems, this complexity must be taken into account even if it might be easier to disregard it.

Joanna Bryson, artificial intelligence scholar, writes that society should currently be taking active measures to ensure that robots are clearly designed to be perceived as non-sentient possessions

[19] Saxena, S., O'Connell, K., Underwood L. (2002)"Cross-Cultural Quality of Life Assessment at the End of Life: A Commentary", Saxena, S., O'Connell, K., Underwood L., *The Gerontologist*, Vol 42, Special Issue III, October 2002, 81-85.

[20] Bender, E. M., Gebru, T., McMillan-Major, A., Mitchell, S., and et al (2021). On the dangers of stochastic parrots: Can language models be too big? *FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency*. DOI: 10.1145/3442188.3445922

[21] Erscoi, Lelia A, Annelies Kleinherenbrink, and Olivia Guest. "Pygmalion Displacement : When Humanising AI Dehumanises Women," 1979, 1–36.

[22] Frischmann, B., & Selinger, E. (2018). Our Free Will Discussion. In *Re-Engineering Humanity* (pp. 301-303). Cambridge: Cambridge University Press. doi:10.1017/9781316544846.019

that lack identity and agency. This helps to create and maintain "Human-centric AI", which she considers crucial now, and as we go forward.[23]

Prevent harm to moral and character development, and prevent moral injury

When we consider ethics that need to be applied in the rollout of artificial social and intelligent systems, we can be simplistic in reducing all to a set of rules. I have worked in the field of bioethics, where I've seen an unfortunate tendency to reduce complex decision-making to a set of rules.. Rules relating to privacy and freedom of various kinds can be helped using deontological and consequentialist ethical traditions. However when dealing with complex issues, another valuable ethical perspective comes from virtue ethics.[24] Virtue ethics emphasizes the effects that our actions have on the development of our character, and on our flourishing.

In my research on altruism and altruistic or compassionate love, I have emphasized that other-centered love is good for the recipients, but also enables the giver of love to flourish as a human being.[25] Love is an imperfect and messy thing, and ultimately undefinable, especially that other-centered, caring, accepting, understanding kind of love, which some call compassionate love.[26] We crave this kind of love from others, whether we acknowledge it or not.

Compassionate love can be nested in romantic or family contexts, but also can exist with friends or even strangers. Compassionate love from another person occurs when they desire that we flourish. It is not just part of some selfish or manipulative calculus. When we are loved in this kind of way it signifies to us that we are understood and valued and accepted for who we are at a fundamental level. To receive this kind of love feels good in a lasting way, it sticks to the ribs and doesn't give us indigestion. This kind of caring love has a weight, a nourishing quality. It feels important to us that it is the choice of the other person, and comes at some kind of cost from the giver of that love, whether it be of time or energy or vulnerability.

A machine, a sophisticated intelligent system, cannot truly give this kind of love. To pretend that it can is destructive. And when we divert our love towards robots, bots, and machine-seeming texts, that energy is diverted from the human beings who would benefit so much from this love.

The concept of moral injury has become more widely used, extending beyond what happens in traumatic situations to effects in ordinary life experiences.[27] Moral injury arises when people face situations that violate their conscience or threaten their core values. Writings on this in the social sciences emerged in the context of soldiers returning from war, and health care workers required to make decisions that seemed cruel in the midst of required trade-offs. Behaviors in these kinds of situations can affect who we are, and how we behave in other situations.

How does anthropomorphizing behavior toward an AI entity affect how we are fundamentally, and how we behave to other humans? This is not a small problem. On the one hand, behaving 'decently' to a robot or other bot or device that seems human seems at the time to be the right thing

---

[23] Joanna J. Bryson and Andreas Theodorou, solicited and reviewed chapter in the collection Human-Centered Digitalization and Services, Marja Toivonen-Noroand Eveliina Saari (eds.), Springer, 2019.

[24] *The Cambridge Companion to Virtue Ethics (2013)– Daniel C. Russell, Cambridge University Press NY*

[25]Underwood, L. "Altruistic Love – Compassionate Love". In *Encyclopedia of Human Relationships.* Harry T. Reis & Susan Sprecher (Eds.), Thousand Oaks, CA: Sage. 2009. And Underwood, L. "Compassionate Love," in Post, Stephen G. ed. 2004. *Encyclopedia of Bioethics*, 3rd edition. New York: Macmillan Reference USA, 483-488.

[26] Underwood, L. G. (2008). Compassionate love: A framework for research. In B. Fehr, S. Sprecher, & L. G. Underwood (Eds.), *The science of compassionate love: Theory, research, and applications* pp 3–25. Wiley Blackwell. https://doi.org/10.1002/9781444303070.ch1

[27] https://www.scientificamerican.com/article/moral-injury-is-an-invisible-epidemic-that-affects-millions/

to do, and to do so may have good effects on our relationships with humans. However, the more we treat machines as if they were human, the more agency we expect them to have and the more deceived we become. If the child says please and thank you to Alexa, or other speaking in-home connections to the web enhanced by chat-GPT, the child tends to ascribe more humanity to this non-human entity. We do not say please and thank you, usually, to our pets. The more human-like the entity, the more it draws us into interacting with it like a human. If we do not have clear indicators of the fact that it does not have the value of a human person, it confuses our inner sense of morality. It is hard enough to behave in a respectful and decent way to other human beings. How much more difficult confusion regarding the humanity of AI devices makes this.

Another problem arises when we put an AI 'helper' between ourselves and the person we are interacting with. This may encourage us to behave in un-ethical ways, as we are not confronting directly the object of our behavior. [28] This can be the case in hiring decisions or military settings. We distance ourselves from the harm we have instigated but do not think we have actually caused.

Be honest and clear with the language we use to describe AI entities

The language and metaphors we use shape how we think about things.[29]

What exactly is the 'intelligence' in Artificial Intelligence? Human intelligence entails many things, and complex algorithmically driven solutions are just one part of that intelligence.

Some of these technologies are a form of human social collaborations, a kind of mashup of various human language and expressions scraped from data that exists, created by humans. And to remind ourselves of where they originate and how they gather content is important.

There is hubris in the assumption that these entities are conscious, or can become conscious. This is not a known fact, it is an imaginative extrapolation. There is still much debate from neuroscientists, philosophers, physicians and physicists about what exactly consciousness is.[30] Currently, AI entities are not conscious, and do not have agency or autonomy, and we must start from that, not describe them in ways that cloud understanding.

An example of the kind of language sleight of hand that can lead us astray is when "creators of algorithms call our behavioral tendencies our 'preferences' and play innocent to the fact that our preferences and behavior aren't always aligned."[31] Amazon tells me my preferences for various books based on my behaviors of searching and previous purchases. These many not be my preferences, but somehow using that word might make me think I actually do prefer things that I really do not.

Much of the generative AI has been developed from a study of how nerves connect in the brain, and this has let the developers to call them neural networks. But even this description leads us into thinking that they are the same as the kinds of networks that exist in our brains, and although modeled on neural connections, they are different from them.

[28]Gratch, J, and Fast ,NJ. (2022) "The Power to Harm: AI Assistants Pave the Way to Unethical Behavior." *Current Opinion in Psychology* 47 (October). doi:10.1016/j.copsyc.2022.101382.

[29] Thibodeau PH, Boroditsky L (2011) Metaphors We Think With: The Role of Metaphor in Reasoning. PLoS ONE 6(2): e16782. https://doi.org/10.1371/journal.pone.0016782

[30] Kind, A and Stoljar, D (2023) *What is Consciousness? A Debate*. Routledge.

[31] Franklin M, Ashton H, Gorman R, Armstrong S. Recognising the importance of preference change: A call for a coordinated multidisciplinary research effort in the age of AI, *The AAAI-22 Workshop on AI For Behavior Change held at the Thirty-Sixth AAAI Conference on Artificial Intelligence (AAAI-22) 2022.* https://arxiv.org/abs/2203.10525

The kind of language we use to describe AI bots and our interactions with them needs to be exceptionally clear when communicating with the public, but also in legislation and regulatory language. Another example is that what we mean by 'relational' when describing AI systems is quite different than what 'relational' means when we describe our interactions with family, friends and colleagues. "Interactive AI" might be a better term than "Relational AI".

The choices around descriptive language are important as it shapes how we view the entities and regulate their use and the assumptions the general public makes about them. Insisting on accurate language will help us in legislating for and interacting with these systems.