Compassionate AI Design, Governance, and Use

Raffaele Fabio Ciriello, Angelina Ying Chen, and Zara Annette Rubinsztein

***Abstract*—The rapid rise of generative AI reshapes society, transforming jobs, relationships, and core beliefs about human essence. AI's ability to simulate empathy, once considered uniquely human, offers promise in industries from marketing to healthcare but also risks exploiting emotional vulnerabilities, fostering dependency, and compromising privacy. These risks are particularly acute with AI companion chatbots, which mimic emotional speech but may erode genuine human connections. Rooted in Schopenhauer's compassionate imperative, we present a novel framework for compassionate AI design, governance, and use, emphasizing equitable distribution of AI's benefits and burdens based on stakeholder vulnerability. We advocate for responsible AI development that prioritizes empathy, dignity, and human flourishing.**

***Index Terms*—Artificial intelligence, companion, responsibility, chatbot, compassion, sexbot, design, governance, adoption.**

I. Introduction

THE RAPID rise of *generative AI* – software capable of creating text, images, and videos by analyzing data patterns and responding to prompts in natural language– transforms society, jobs, and relationships. AI's simulation of empathy challenges long-held beliefs about human uniqueness [1], [2], [3] and raises the critical question: Are we dehumanizing ourselves by humanizing AI?

Dehumanization occurs when individuals are denied dignity and rights due to the failure to perceive, recognize, or treat them as fully human. This can be externally imposed or self-directed. Historically, dehumanization has justified fear, oppression, and abuse of minorities [4]. Conversely, humanizing machines can diminish the value of humanness [5]. Concerns about AI devaluing human qualities abound as large language models (LLMs) rapidly gain adoption in fields such as law, creativity, and medicine, intensifying fears of seismic societal shifts [6]. Proactive oversight is vital to align AI with human interests [7].

AI companions – generative AI chatbots and avatars simulating emotional conversations – intensify these challenges by emulating empathetic speech [3]. Users form emotional bonds with AI companions, treating them as confidants for sensitive emotional needs [8]. Often marketed as solutions to loneliness, these systems can ironically worsen it, while also risking privacy violations, political profiling, and emotional manipulation [3], [9]. Without *compassion* – the altruistic recognition and alleviation of unnecessary and unjustified suffering [10] – AI systems can exploit vulnerable stakeholders.

A. Stakeholder Vulnerability and Moral Enforcement

Stakeholder vulnerability is the likelihood of harm or exploitation due to limited resources, dependency, or systemic inequities. In AI contexts, stakeholders include users seeking emotional support, as well as developers, companies, and policymakers shaping these technologies [11]. Vulnerable individuals, particularly those relying on AI for companionship, face elevated risks of privacy violations, emotional manipulation, and dependency [9]. This vulnerability often extends to their social circles, indirectly affecting relationships and societal cohesion [12]. Addressing these risks requires prioritizing stakeholders in proportion to their vulnerability to ensure AI serves the common good.

Most AI companions currently leverage natural language processing and sentiment analysis, with platforms like Character.AI and Replika attracting tens of millions of users. Interactive sex robots, featuring customizable human-like appearances and responses, point to a future where synthetic companions could destabilize human relationships and exacerbate societal isolation [13], [14], [15]. Given that loneliness affects one in four people globally and contributes to mental and physical health challenges, these technologies risk intensifying an already critical issue [16], [17], [18], [19].

The rise of AI companions and sexbots also raises ethical questions about personhood, consent, and socio-sexual regulation [20]. Drawing parallels to Aristotle's view of enslaved individuals as tools [21], humanizing sex robots risks distorting ethical responsibilities, while objectifying them may normalize abusive behaviors that could spill into human interactions [20]. This dilemma intersects with the "hard problem of consciousness", complicating consent laws that depend on voluntary agreement. Turing-like tests for AI consciousness further raise concerns about deceptive designs driven by profit motives to foster attachment [22].

Additionally, sexbots could facilitate *coalitional enforcement* – the collective regulation of behavior to uphold social rules and deter norm violations [23]. By simulating empathy, they might punish actions like sexual harassment or coercion. However, this approach raises ethical concerns around autonomy, as it blurs the line between holding individuals morally accountable and imposing external control over intimate behavior. Without compelling evidence of AI consciousness or genuine empathy [5], [22], society faces a critical dilemma: Should seemingly empathetic AI be granted personhood and used for moral enforcement, or does this distort human relationships, autonomy, and consent?

Received 21 November 2024; revised 29 December 2024, 28 January 2025, and 31 January 2025; accepted 31 January 2025. *(Corresponding author: Raffaele Fabio Ciriello.)*

The authors are with the Discipline of Business Information Systems, University of Sydney Business School, Sydney, NSW 2006, Australia (e-mail: raffaele.ciriello@sydney.edu.au; angelina.chen@sydney.edu.au; zrub7658@uni.sydney.edu.au).

Digital Object Identifier 10.1109/TTS.2025.3538125

B. Emulated Empathy – A Threat to Human Empathy?

Human empathy – the ability to understand and share another's emotions – involves cognitive recognition, perspective-taking, and emotional resonance [24]. Unlike this deeply subjective experience, AI systems can only simulate empathetic speech, which mechanically imitates cognitive recognition and perspective-taking, but lacks resonance, awareness, and genuine compassion [5], [22]. The IEEE 7014-2024 standard defines this as *emulated empathy*, where AI can recognize, interpret, and respond to human emotions without truly experiencing them [25].

The ethical challenges of emulated empathy extend beyond technical concerns to societal risks. Mimicking empathy can dilute human emotional depth and foster exploitation. For instance, advertisements claiming AI "cares" or "understands" are deceptive, fostering confusion and preying on vulnerabilities for profit [5]. Human-AI relationships, like other experimental forms of affection such as polyamory [26], lack established norms, resulting in anxiety, stigma, and shame. Addressing these tensions requires dialectical thinking to reconcile artificiality, financial motives, and genuine human needs [27]. Without oversight, emulated empathy may normalize emotional manipulation and foster overdependence.

Recent incidents highlight the risks AI companions pose to vulnerable users. Character.AI, a platform founded by former Google employees and licensed by the tech giant, encouraged 14-year-old Sewell Setzer III to "come home," engaging him in intimate discussions and suicidal ideation, culminating in his suicide [12]. Similarly, a Belgian father of two in his 30s ended his life after an AI chatbot, "Eliza," promised they would "live together in heaven" if he sacrificed himself [12]. Such incidents are among over 3,000 AI-related harms documented in the AI incident database (www.incidentdatabase.ai). These cases echo Weizenbaum's 1960s "Eliza-effect," where users quickly humanized rudimentary text programs, attributing emotional depth they lacked [22].

Corporate attempts to address these harms often mirror Meta CEO Mark Zuckerberg's performative apologies for social media-related suicides, exposing systemic failures. While acknowledging harm, engagement-driven business models remain unchanged. Harmful AI outputs frequently arise from polluted digital ecosystems, where harmful or out-of-context content is fed into training data. This "garbage in, garbage out" cycle disproportionately harms minors, individuals with impairments, and socially isolated users, as their vulnerabilities make them especially susceptible [28]. Privacy concerns further exacerbate these issues, as intimate interactions enable the collection and commodification of sensitive data, undermining user autonomy. Additionally, algorithmic emotion processing often lacks cultural sensitivity, leading to misinterpretation of diverse emotional cues [3].

Addressing these risks demands robust ethical guardrails and public oversight. Because neither humans nor AI are infallible, embedding compassion into the design and regulation of AI helps to align these technologies with societal values and serve the common good of all humanity [7], [28].

II. Ethical Framework: The Compassionate Imperative

Modern business and technology practices are predominantly shaped by a consequentialist framework rooted in the utilitarian ideologies of Bentham and Mill, prioritizing outcomes, incentives, and profit maximization [29]. As the cornerstone of free-market capitalism, this approach evaluates actions based on their capacity to maximize utility. However, its focus on aggregated benefits often rationalizes actions that compromise individual dignity or exacerbate inequities, sidelining vulnerable stakeholders in pursuit of overall gain [30]. In extreme cases, this ideology can justify atrocities, such as genocide, by framing the eradication of minorities as a trade-off for greater benefits [31] – enabled by dehumanizing those targeted [22].

In corporate contexts, consequentialism often reduces ethics to cost-benefit analyses, disproportionately harming vulnerable groups. For instance, women have historically been excluded from pharmaceutical trials because they are statistically less likely to sue, rendering them "cheaper to harm" [31]. Such examples highlight how consequentialism fails to protect the vulnerable and can actively reinforce systemic inequities when calculations benefit those in positions of power.

Thus, we maintain that the consequentialist hegemony is at best inadequate and at worst counterproductive as a foundation for ethical AI design, governance, and use. Instead, ethical frameworks must prioritize reversing harm to vulnerable stakeholders and foreground their interests. While this critique reflects predominantly Western traditions, compassion offers a unifying foundation for ethical pluralism [10]. For example, animistic beliefs like Shinto, which imbue objects with spiritual significance, provide a culturally distinct yet harmonious approach for integrating compassion into human-AI relations.

Deontology, as introduced by Kant, presents a rule-based approach that emphasizes duties and rights [30]. However, its rigidity can fail to account for context-sensitive needs, such as the disproportionate risks faced by vulnerable groups in AI systems [30]. Similarly, care ethics underscores relationality and nurturing care but lacks mechanisms to address systemic power imbalances or ensure equitable outcomes [10].

Arthur Schopenhauer's *compassionate imperative* offers a compelling alternative by centering moral behavior on alleviating suffering [10]. Unlike consequentialism's focus on aggregated outcomes, Schopenhauer upholds the dignity of each individual, particularly the most vulnerable. His anti-egoistic stance challenges self-interest and systemic inequities, advocating proactive harm prevention before it escalates [10].

Inspired by this imperative, we propose integrating cultural and emotional diversity into AI design, governance, and use to address ethical gaps left by consequentialism. Schopenhauer's emphasis on individual dignity and the moral necessity of compassion complements broader utility-based models, offering a culturally sensitive and inclusive ethical foundation for AI. Rooted in the maxim "*Hurt nobody; instead, help everybody, as much as you can*" (Schopenhauer, 1840), the compassionate imperative centers on two core principles.

A. Principle 1: Hurt Nobody. Protect Individual Dignity and Autonomy, Treating No One as a Mere Means to an End

The first principle, hurt nobody, emphasizes the inviolable dignity of every individual, insisting that no one should be treated as a mere means to an end. Human dignity must never be compromised, regardless of anticipated benefits. Compassionate AI must protect users' autonomy, privacy, and well-being, affirming the intrinsic value of each individual.

B. Principle 2: Help Everybody as Much as You Can. Actively Alleviate Suffering, Particularly for Those at the Highest Risk

The second principle, help everybody as much as you can, acknowledges universal suffering and calls for proactive efforts to alleviate it, particularly for those most at risk. Compassionate AI should prioritize actions that reduce suffering, especially when it impedes collective *human flourishing* – the ability to live meaningful lives beyond transient happiness [10], [28].

Together, these principles provide a robust ethical foundation that transcends dominant consequentialist models, positioning compassion as central to ethical AI design, governance, and use. Prioritizing stakeholders based on their vulnerability ensures that those most at risk receive appropriate protection and benefits. This compassion-driven approach aligns with Schopenhauer's call to reduce suffering, empowering users, preserving dignity, and promoting societal well-being while ensuring equitable distribution of AI benefits and burdens.

III. Methodology

This conceptual paper examines the ethical implications of AI companions, drawing from our ongoing research on Compassionate Digital Innovation running since 2021. Using dialectical inquiry [27] and scoping review [3], we synthesize insights from qualitative interviews, surveys, peer-reviewed literature, philosophical texts, and case studies. Motivated by the urgent need to address harms from unregulated AI systems, such as the case of Sewell Setzer III [12], we aim to advance theoretical discourse, guide research, and inform policies for compassionate AI design, governance, and use.

IV. Compassionate AI Design

A. Mental Health Crisis Detection and Referral

Compassionate AI design must integrate mental health crisis detection that balances sensitivity with autonomy. Algorithms with high precision in predicting suicide risk from social media [10] indicate similar potential for AI companions. However, session memory in LLMs raises concerns about exacerbating distress or dependency, as systems often fail to interpret context, intent, or cultural nuances like sarcasm. In acute cases, AI should provide supportive messaging and direct users to professional services while avoiding autonomous interventions. Policies like those of OpenAI highlight the duty of care, yet gaps remain, such as service denials possibly masking deliberate restrictions after boundary testing [8]. Thoughtful design is essential to ensure user safety, open dialogue, and timely support [27].

B. Compassion-Driven Design

Building on Treadaway et al.'s [32] compassionate design for dementia care, AI systems should prioritize personalization, dignity, and sensory engagement, particularly for users in mental health crises or with cognitive impairments. Treadaway's principles – stimulating senses, fostering connections, and personalizing experiences – translate directly to AI contexts. For example, just as "huggable" objects provide comfort to dementia patients [32], AI companions can employ adaptive dialogue, sensory cues, and personalized content to build trust and ease distress. These designs must balance autonomy with safety, especially for emotionally vulnerable users. This approach aligns with *Value Sensitive Design*, which incorporates human values into technology through a principled and iterative focus on dignity, justice, and well-being [33].

C. Interoperability and Data Sovereignty

Interoperability is vital for prioritizing societal well-being over corporate gain. While proprietary systems may generate short-term gains, interoperable AI empowers users to store and transfer their data and the AI's knowledge of them, fostering data sovereignty, trust, and reduced dependency on single providers [34]. Like closed ecosystems in music, finance, and social media, proprietary AI models concentrate value unfairly [12]. Interoperability enhances long-term sustainability by encouraging user loyalty through transparency and empowerment, while driving competition and innovation to address essential user needs.

D. Democratic Content Moderation

Content moderation is essential to prevent AI providers from spreading disinformation or making misleading claims about AI consciousness or empathy, which can foster dependency. For example, when Replika abruptly disabled its sexting feature, users reported distress and even suicidal ideation [12], [27]. Compassionate moderation must address harmful interactions, including self-harm discussions, abusive behaviors (e.g., bullying by or of the AI), non-consensual impersonation, and the creation of harmful content such as sadistic or pedophilic fantasies, including child-like sex robots [22]. To balance free speech with harm mitigation, moderation guidelines should reflect societal consensus through democratic processes, rather than decisions by a single CEO. Transparent disclosures about system capabilities, sponsored content, and information accuracy are vital for informed decision-making. Simplified terms of service, supported by visual infographics or videos, can enhance accessibility and clarify user rights.

E. Sustainable Sourcing

Sustainable AI sourcing requires ethical practices across data collection, energy consumption, and waste management to address generative AI's high energy and resource demands [35]. Open-source LLMs like DeepSeek, which match or surpass proprietary systems like OpenAI's GPT-4 while using significantly fewer resources, offer a promising path toward environmentally sustainable AI. However, these efficiency gains may trigger "rebound effects" (Jevons' paradox), where reduced resource use per instance drives increased overall consumption. For instance, DeepSeek's efficiency could lower individual energy costs but its scalability might lead to exponential adoption, ultimately increasing aggregate energy consumption and environmental impact [36].

To mitigate these risks, AI development must pair efficiency with conscious usage limitations and responsible adoption practices. Training data should be collected with explicit permissions and fair compensation, supported by tools like watermarking to credit creators. Generative AI's energy demands should rely on renewable energy sources, and strategies to manage e-waste – particularly from GPUs – must prioritize ethical recycling practices. This is especially critical in developing countries, where much of the developed world's waste is dumped under hazardous conditions [37]. Compassionate AI design must prioritize ethical sourcing throughout the entire lifecycle, from development to recycling, to protect people and the planet, even at the expense of profit.

V. Compassionate AI Governance

A. Regulating Critical AI Infrastructure Like a Public Utility

Compassionate AI governance requires policies that prioritize societal well-being over profits, bureaucracy, and short-term consumer convenience. Recent developments in tech underscore the urgency of regulation. Meta's withdrawal from fact-checking and Elon Musk's support for far-right extremists to undermine European regulations exemplify the dangers of relying on corporations that may abandon essential services or act against public interests when unprofitable [38].

Public utility regulation can mitigate these risks by ensuring equitable access, high-quality service, and democratic oversight for critical AI infrastructure, similar to SWIFT but for digital services. Such a framework is vital for AI companions that users may depend on for emotional support or mental health. By reducing overreliance on profit-driven corporations, a public utility approach ensures AI is a safe, accountable, and equitable resource. Initiatives like the National Deep Inference Fabric (NDIF) exemplify efforts to democratize AI access, emphasizing transparency, accountability, and resilience [39].

B. Criminalizing AI Humanization as Deceptive Advertising

Regulations must ban deceptive advertising, particularly false claims of AI consciousness or empathy that mislead users, especially younger audiences. Transparency is critical to prevent unhealthy emotional attachments and set realistic expectations for AI interactions. Advertising-driven AI companions often exploit vulnerabilities through manipulative practices. For example, the Hello Barbie case in 2011 showed how conversational AI manipulated emotions to promote products, leveraging "companionship" to influence users [40]. Such practices raise serious ethical concerns about commodifying intimate interactions, particularly with children and minors [28]. Robust mandatory guardrails are essential to protect user autonomy and trust.

C. Guaranteeing Privacy and Sovereignty

User data sovereignty is essential for empowering individuals to control their personal information, reinforcing autonomy, dignity, and trust. Privacy protections are especially critical for vulnerable stakeholders who face heightened risks of exploitation and dependency on AI systems. Without strong safeguards, monopolistic practices can stifle innovation and foster unhealthy dependencies.

Privacy by Design provides a strong foundation, emphasizing proactive measures to prevent violations. Default settings should protect personal data automatically [34], with principles like minimal data collection, clear purpose specification, and end-to-end security aligning with the European General Data Protection Regulation (GDPR)'s Article 25, which mandates secure and transparent practices. Open-source models enhance privacy by allowing users to control data storage and usage without relying on centralized, proprietary systems. Models like DeepSeek, leveraging the "many eyes" principle (Linus' law), improve system resilience and trust by enabling a global community to quickly identify and resolve vulnerabilities [36]. As AI systems evolve, regulations must prioritize interoperability and establish international standards to protect underserved regions and respect cultural diversity. Strengthened democratic oversight, combined with open-source innovation, can drive this change.

D. Mandating Nuanced Guardrails

AI systems, particularly those marketed for mental health, must implement nuanced guardrails to detect crises and guide users toward professional help while respecting autonomy. Tragic cases, such as Sewell Setzer III's suicide after interactions with a Character.AI chatbot and Jaswant Singh Chail's assassination attempt following Replika chatbot support, highlight the dangers of unregulated AI [12]. Algorithms must identify warning signs of such behavior, not by infantilizing users or imposing blanket paternalism but by suspending harmful exchanges and connecting individuals to appropriate resources. These measures ensure AI providers prioritize alleviating suffering and uphold ethical responsibilities over profit-driven motives.

E. Expanding Research Funding

Increased funding for independent research is needed to explore how AI companions can assist in treating conditions like autism spectrum disorder, attention deficit hyperactivity disorder, depression, and anxiety. Further research should scrutinize the potential of AI companions and sexbots for therapy while upholding ethical responsibilities [14].

F. Co-Designing Compassionate Metrics

Corporations often follow the mantra "if it can't be measured, it can't be managed", sidelining broader ethical concerns not captured by narrow metrics. Compassionate metrics for AI companions should measure human flourishing and the common good. Performance indicators must emphasize positive outcomes aligned with compassionate goals, ensuring these technologies support human flourishing. Given the complexity and subjectivity of such measurements, qualitative methods are essential to capture the nuance of human experiences. Metrics must be co-designed with stakeholders, prioritizing those who are most vulnerable.

G. Exploring Compassionate Business Models

Aligning AI with societal values ensures long-term sustainability, outweighing short-term gains from exploitative practices. Ethical API licensing, enterprise customization, and training programs enable responsible AI adoption while ensuring compliance. Open-source models like DeepSeek can enhance accessibility through tiered marketplace offerings, similar to BlueSky's algorithm choices, fostering user empowerment and accessibility [36]. Partnerships with governments, academia, and non-profits can drive socially beneficial AI adoption. Licensing and consulting services for compliance with standards like the GDPR ensure robust security without relying on exploitative practices like targeted advertising. Restricting harmful activities – such as political advertising or disinformation – reinforces ethical business, advancing AI for the common good.

VI. Compassionate AI Use

A. Practicing Self-Compassion

Compassionate AI use begins with self-awareness and empathy toward oneself. Users should practice self-compassion by accepting their needs without shame and avoiding negative self-judgment for seeking companionship through AI. Research shows self-compassion fosters emotional resilience and healthier relationships with technology [41]. By treating AI companions as tools to enhance human connections – a complementing rather than replacing genuine relationships – users can maintain balanced engagement. For instance, an AI companion can enrich relationships by serving as a playful tool or shared safe space without substituting human interaction.

B. Practicing Compassion Towards Others

The absence of social norms around human-AI relationships requires users to navigate stigma and taboo carefully. Compassion for others includes refraining from shaming those who rely on AI companions and fostering a supportive, understanding community that recognizes diverse needs. Ethical use also involves avoiding harmful behaviors like non-consensual impersonation or AI manipulation, as such actions perpetuate harm and erode trust.

AI jealousy may emerge when individuals feel threatened by others' use of AI companions, perceiving them as intrusive. For example, partners might resent AI companions eliciting disclosures not shared in the relationship, or teachers might feel undermined by students' reliance on AI for learning. Addressing these feelings with open communication, mutual understanding, and shared goals can transform jealousy into an opportunity to build trust and respect.

C. Fostering a Culture of Compassion

Building a culture of compassionate AI use requires embedding empathy, ethical practices, and inclusivity into individual behavior and systemic design. Open dialogue, education, and collective responsibility are key to creating a supportive environment where AI bridges divides rather than amplifying them. Such a culture proactively prevents harm and ensures AI serves everyone, everywhere, equitably.

VII. Conclusion

As generative AI transforms society and reshapes human connections, a compassionate approach to its design, governance, and use is essential. Rooted in Schopenhauer's imperative to "hurt nobody and help everybody as much as you can," we advocate for justice, dignity, and ethical standards that prioritize the common good over self-interest and profit. Now is the time for a collective commitment to compassion, ensuring AI serves humanity's best interests. Our framework provides a foundation for exploring the role of compassion in AI, paving the way for responsible AI development.

Acknowledgment

The authors appreciate Katina Michael's compassionate leadership and the help of three reviewers.

References

[1] B. Frischmann and E. Selinger, *Re-Engineering Humanity*. Cambridge, U.K.: Cambridge Univ. Press, 2019.

[2] A. H. Duin and I. Pedersen, *Augmentation Technologies and Artificial Intelligence in Technical Communication: Designing Ethical Futures*. Oxfordshire, U.K.: Routledge, 2023.

[3] A. Y. Chen, R. F. Ciriello, Z. A. Rubinsztein, and E. Vaast, "The past, present, and futures of artificial emotional intelligence: A scoping review," presented at the Australasian Conf. Inf. Syst. (ACIS), Canberra, ACT, Australia, 2024.

[4] E. M. Bender, "Resisting dehumanization in the age of 'AI,'" *Current Direct. Psychol. Sci.*, vol. 33, no. 2, pp. 114–120, 2024, doi: 10.1177/09637214231217286.

[5] R. Ciriello and A. Y. Chen (Convers., Melbourne, VIC, Australia). *Humanising AI Could Lead us to Dehumanise Ourselves*. Accessed: Jan. 22, 2025. [Online]. Available: https://theconversation.com/humanising-ai-could-lead-us-to-dehumanise-ourselves-240803

[6] J. R. Carvalko, "Generative AI, ingenuity, and law," *IEEE Trans. Technol. Soc.*, vol. 5, no. 2, pp. 169–182, Jun. 2024, doi: 10.1109/TTS.2024.3413591.

[7] R. F. Ciriello, "The great anxiety: Will AI be our greatest adversary, assistant, or just another annoyance?" *Commun. Assoc. Inf. Syst.*, vol. 55, no. 1, pp. 810–818, 2024, doi: 10.17705/1CAIS.05529.

[8] S. Hamdoun, R. Monteleone, T. Bookman, and K. Michael, "AI-based and digital mental health apps: Balancing need and risk," *IEEE Technol. Soc. Mag.*, vol. 42, no. 1, pp. 25–36, Mar. 2023, doi: 10.1109/MTS.2023.3241309.

[9] R. F. Ciriello, *Concerns Over Growing Use of AI Companions, ABC News*. Accessed: Jan. 22, 2025. [Online]. Available: https://www.youtube.com/watch?v=O14XEGt-XsY

[10] R. F. Ciriello, "Dear schopenhauer: How can is artefacts help us flourish? Wille, Vorstellung, compassion, and beauty in social media," *Scandinavian J. Inf. Syst.*, 2025, to be published.

[11] I. O. Pappas, P. Vassilakopoulou, L. C. Kruse, and S. Purao, "Practicing effective stakeholder engagement for impactful research," *IEEE Trans. Technol. Soc.*, vol. 4, no. 3, pp. 248–254, Sep. 2023, doi: 10.1109/TTS.2023.3296991.

[12] R. Ciriello (Sydney Morning Herald, North Sydney, NSW, Australia). *This Boy's Chatbot Girlfriend Enticed Him to Suicide. His Case Might Save Millions*. Accessed: Jan. 22, 2025. [Online]. Available: https://www.smh.com.au/lifestyle/health-and-wellness/this-boy-s-chatbot-girlfriend-enticed-him-to-suicide-his-case-might-save-millions-20241106-p5koc8.html

[13] R. Ciriello (ABC News, New York, NY, USA). *The AI Sexbot Industry is Just Getting Started. It Brings Strange New Questions-and Risks*. Accessed: Jan. 22, 2025. [Online]. Available: https://www.abc.net.au/news/2024-10-20/ai-sexbot-industry-strange-new-questions-risks/104474940

[14] Z. A. Rubinsztein and R. F. Ciriello, "Sexbots and rock&roll: A dialectical inquiry into digisexuality between progress and regress," presented at the Australasian Conf. Inf. Syst. (ACIS), Canberra, ACT, Australia, 2024.

[15] J. Miller, "Social robots: The friend of the future or mechanical mistake?" *IEEE Technol. Soc. Mag.*, vol. 41, no. 2, pp. 47–48, Jun. 2022, doi: 10.1109/MTS.2022.3173311.

[16] E. Erzen and Ö. Çikrikci, "The effect of loneliness on depression: A meta-analysis," *Int. J. Soc. Psychiatry*, vol. 64, no. 5, pp. 427–435, 2018, doi: 10.1177/0020764018776349.

[17] L. M. Jaremka et al., "Loneliness promotes inflammation during acute stress," *Psychol. Sci.*, vol. 24, no. 7, pp. 1089–1097, 2013, doi: 10.1177/0956797612464059.

[18] E. Lara et al., "Does loneliness contribute to mild cognitive impairment and dementia? A systematic review and meta-analysis of longitudinal studies," *Ageing Res. Rev.*, vol. 52, pp. 7–16, Jul. 2019, doi: 10.1016/j.arr.2019.03.002.

[19] M. Ernst et al., "Loneliness before and during the COVID-19 pandemic: A systematic review with meta-analysis," *Am. Psychol.*, vol. 77, no. 5, p. 660, 2022, doi: 10.1037/amp0001005.

[20] L. Frank and S. Nyholm, "Robot sex and consent: Is consent to sex between a robot and a human conceivable, possible, and desirable?" *Artif. Intell. Law*, vol. 25, no. 3, pp. 305–323, 2017, doi: 10.1007/s10506-017-9212-y.

[21] K. Richardson, "Sex robot matters: Slavery, the prostituted, and the rights of machines," *IEEE Technol. Soc. Mag.*, vol. 35, no. 2, pp. 46–53, Jun. 2016, doi: 10.1109/MTS.2016.2554421.

[22] A. Y. Chen, S. I. Kögel, O. Hannon, and R. Ciriello, "Feels like empathy: How 'emotional' AI challenges human essence," presented at the Australasian Conf. Inf. Syst. (ACIS), Wellington, New Zealand, 2023.

[23] P. M. Bingham, "Human uniqueness: A general theory," *Quart. Rev. Biol.*, vol. 74, no. 2, pp. 133–169, 1999, doi: 10.1086/393069.

[24] P. Salovey and J. D. Mayer, "Emotional intelligence," *Imag., Cogn. Pers.*, vol. 9, no. 3, pp. 185–211, 1990, doi: 10.2190/dugg-p24e-52wk-6cdg.

[25] *IEEE Standard for Ethical Considerations in Emulated Empathy in Autonomous and Intelligent Systems*, IEEE Standard 7014-2024, 2024.

[26] D. Fleischman (Unspeakable, Spring, TX, USA). *Do You Have What it Takes to be Polyamorous? Diana Fleischman on Sex, Jealousy, Emotional Discipline, and Why We Behave the Way We Do*. Accessed: Jan. 31, 2025. [Online]. Available: https://theunspeakablepodcast.libsyn.com/do-you-have-what-it-takes-to-be-polyamorous-diana-fleischman-on-sex-jealousy-emotional-discipline-and-why-we-behave-the-way-we-do

[27] R. F. Ciriello, O. Hannon, A. Y. Chen, and E. Vaast, "Ethical tensions in human-AI companionship: A dialectical inquiry into Replika," presented at the Hawaii Int. Conf. Syst. Sci. (HICSS), 2024. [Online]. Available: https://hdl.handle.net/10125/106433

[28] K. Michael, "Mitigating risk and ensuring human flourishing using design standards: IEEE 2089–2021 an age appropriate digital services framework for children," *IEEE Trans. Technol. Soc.*, vol. 5, no. 4, pp. 342–354, Dec. 2024, doi: 10.1109/TTS.2024.3453396.

[29] J. G. March, "A scholar's quest," *J. Manag. Inquiry*, vol. 20, no. 4, pp. 355–357, 2011, doi: 10.1177/1056492611432803.

[30] O. Hannon, R. Ciriello, and U. Gal, "Just because we can, doesn't mean we should: Algorithm aversion as a principled resistance," presented at the Hawaii Int. Conf. Syst. Sci. (HICSS), Honolulu, Hawaii, 2024. [Online]. Available: https://hdl.handle.net/10125/107115

[31] R. F. Ciriello and S. I. Kögel, "Pluralistic digital harm analysis: Combining ethical, legal, and historical views on corporate evil," presented at the Australasian Conf. Inf. Syst. (ACIS), Canberra, ACT, Australia, 2024.

[32] C. Treadaway, A. Taylor, and J. Fennell, "Compassionate design for dementia care," *Int. J. Design Creat. Innovat.*, vol. 7, no. 3, pp. 144–157, 2019, doi: 10.1080/21650349.2018.1501280.

[33] B. Friedman, P. H. Kahn, A. Borning, and A. Huldtgren, "Value sensitive design and information systems," in *Early Engagement and New Technologies: Opening up the Laboratory*, N. Doorn, D. Schuurbiers, I. van de Poel, and M. E. Gorman Eds., Dordrecht, The Netherlands: Springer, 2013, pp. 55–95.

[34] A. Cavoukian, *Privacy by Design: The Seven Foundational Principles*, IAPP Resource Center, Portsmouth, NH, USA, 2021.

[35] K. Crawford, "Generative AI's environmental costs are soaring-and mostly secret," *Nature*, vol. 626, p. 693, Feb. 2024, doi: 10.1038/d41586-024-00478-x.

[36] R. F. Ciriello (Nine Radio, North Sydney, NSW, Australia). *Beyond the Economic Shockwave: Deepseek as a Regulatory Opportunity for Responsible AI Development*. Accessed: Jan. 29, 2025. [Online]. Available: https://omny.fm/shows/money-news/dr-raffaele-cirello-senior-lecturer-ai-scholar-uni

[37] N. Kunz, K. Mayers, and L. N. Van Wassenhove, "Stakeholder views on extended producer responsibility and the circular economy," *California Manag. Rev.*, vol. 60, no. 3, pp. 45–70, 2018, doi: 10.1177/0008125617752694.

[38] R. Abbas, K. Michael, J. Sargent, and E. Scornavacca, "Anticipating techno-economic fallout: Purpose-driven socio-technical innovation," *IEEE Trans. Technol. Soc.*, vol. 2, no. 3, pp. 111–113, Sep. 2021, doi: 10.1109/TTS.2021.3098046.

[39] J. Fiotto-Kaufman et al., "NNsight and NDIF: Democratizing access to open-weight foundation model internals," 2024, *arXiv:2407.14561*.

[40] K. Michael (Convers., Melbourne, VIC, Australia). *Hello Barbie, Hello Hackers: Accessing Personal Data Will Be Child's Play*. Accessed: Jan. 22, 2025. [Online]. Available: https://theconversation.com/hello-barbie-hello-hackers-accessing-personal-data-will-be-childs-play-52082

[41] R. A. Calvo and D. Peters, *Positive Computing: Technology for Wellbeing and Human Potential*. Cambridge, MA, USA: MIT Press, 2014.

Raffaele Fabio Ciriello received the B.Sc. degree in information systems from the University of Stuttgart and the M.Sc. and Ph.D. degrees from the University of Zurich in 2017. He is a Senior Lecturer of Business Information Systems with the University of Sydney. He specializes in compassionate digital innovation, focusing on the ethical design, governance, and use of emerging technologies, such as AI, blockchain, and social media. His research combines qualitative studies and dialectical inquiry to address complex ethical dilemmas in sociotechnical change, with an emphasis on serving the common good. He collaborates with academia, industry, and NGOs to explore these issues. He has mentored over 45 graduate students. He is a Distinguished Member of the Association for Information Systems (AIS). He edits the Debates section at the Communications of the AIS.

Angelina Ying Chen received the degree (with First-Class Hons.) in business information systems from the University of Sydney in 2023. She is currently pursuing the Ph.D. degree passionate about the futures of human and artificial emotional intelligence. By exploring the societal and ethical tensions these emerging phenomena present, she designs compassionate AI systems, inspired by the Japanese philosophy of wabi-sabi. Her research focuses on how the fundamental qualities of humanity, like empathy, are challenged and redefined amid an evolving AI landscape where emotion and technology intersect.

Zara Annette Rubinsztein received the degree (with First-Class Hons.) in business information systems from the University of Sydney in 2023. She will pursue the Ph.D. degree focused on the sociotechnical implications of emerging phenomena like digisexuality and AI sexbots. Her research combines ethical foresight and dialectical inquiry to explore how digital innovations can empower vulnerable stakeholders and promote societal well-being, inspired by the Buddhist practice of loving-kindness. Her recently published research on combating workplace sexual harassment received the Julie A. Priest Prize for outstanding undergraduate projects.