SPECIAL ISSUE: FOSTERING SOCIETAL VALUES IN DIGITAL TIMES –
PEACE, CARE, AND TECH REGULATION

Editorial

Christoph Neuberger ⓘ \ Martin Emmer ⓘ \ Clara Iglesias
Keller ⓘ \ Hanna Krasnova ⓘ \ Martin Krzywdzinski ⓘ \
Axel Metzger \ Sonja Schimmler ⓘ \ Lena Ulbricht ⓘ \
Gergana Vladova ⓘ

Memories of Joseph Weizenbaum
   on the Centenary of his Birth

Joseph Weizenbaum would have been 100 years old on January 8, 2023. This anniversary was a welcome occasion to remember the life, work, and impact of the great computer scientist and public intellectual at the Weizenbaum Institute for the Networked Society. Fortuitously, his name was on everyone's lips in 2023: The hype surrounding ChatGPT and generative AI led many to remember the first chatbot ELIZA, which Weizenbaum programmed. For example, on January 20, *The New York Times* published an article that paid tribute to his pioneering work in posing the question, "How Smart Are the Robots Getting?" (Metz, 2023).

With ChatGPT, AI has arrived in people's everyday lives. Anyone can experiment with it and gain experience with its strengths (eloquence, efficiency, speed) and weaknesses (incorrect outputs, bias and lack of fairness, opacity, copyright violations, environmental costs) of generative AI (e.g., Feuerriegel et al., 2023; see also Floyd in this issue of the WJDS). We are also currently in the process of understanding and imagining how AI could harm and benefit society in the future (Dodgson, 2023, pp. 23–24). This situation reflects and revitalizes the questions that Weizenbaum already formulated some sixty years ago.

Parasocial Interaction and Empathy

ELIZA simulates a therapy session, setting up the interaction situation such that new information is fed into the dialogue from the human side (i.e., the patient), which the program only has to pick up on and process to ask a follow-up question. Despite the easily understandable conversational technique, Weizenbaum observed that some interaction partners felt understood by ELIZA and built up an emotional relationship with the chatbot (Bassett, 2019; Moses & Meldman, 2008; Natale, 2019; Pruijt, 2006; see also Berry in this issue of the WJDS).

Similar stories have also appeared in the media during the current AI boom, with reports of humans being in romantic relationships with bots and agents (e.g., Hillebrand, 2023; Voss, 2023) and widespread perceptions of AI having a consciousness of its own (Lemoine, 2022; Levy, 2022). Weizenbaum's observations of ELIZA prompted him to warn against the seductive power of AI and against the willingness of humans to overestimate the capacity of computers.

Notably, Weizenbaum's test subjects attributed empathy to the computer. Empathy is the ability to adopt another person's perspective and to recognize their emotions (Bloom, 2017, pp.16–17; Scudder, 2020, p. 53). In other fields of research, analyzing such illusory relationships has a long tradition (Tukachinsky Forster, 2023). "Parasocial interaction" is a term describing the relationship between protagonists on television – for example, a presenter or a politician – and the audience. By addressing viewers, they create an "illusion of intimacy" (Horton & Wohl, 1956, p. 217) that resembles the experience of a direct conversation. Such experiences can even develop into long-term parasocial relationships (Dibble et al., 2016).

Parasocial interactions also occur between humans and avatars in games (Bowman & Banks, 2021), in multi-user virtual environments (Procter, 2020), and when interacting with consumer chatbots (Tsai & Chuan, 2023). Research has shown that the quality of chatbots or conversational agents (such as Alexa) is essentially measured by whether they behave like humans, are perceived as empathic, or are attributed social intelligence and social presence (Görnemann & Spiekermann, 2022, p. 8; Tsai & Chuan, 2023, p. 262). In interactive marketing, "AI tools' 'empathetic intelligence' in recognizing and understanding people's emotional expressions, responding with proper emotions, and influencing others' emotions" is considered the "next key step for consumer acceptance" (Tsai & Chuan, 2023, p. 260).

However, one of the dark sides of empathy is that seeing through other people can also be used to manipulate them (Breithaupt, 2019), in marketing as well as in politics. Populist leaders seemingly have a special sensorium for people's fear of loss, which they pick up on and misuse for their political gain (Wigura & Kuisz, 2020, p. 47). As such, it is also conceivable that an empathic chatbot is particularly good at manipulating users (see Berry in this issue of the WJDS), for example, in the service of so-called "spin dictators" (Guriev & Treisman, 2022).

Uses of AI in Different Constellations

To investigate how users interact with computers, Weizenbaum chose a simple constellation, namely, a conversation between a therapist and a patient. The two participants have complementary roles in a professionally determined situation. Here, too, fields of application for AI are emerging, for example, in the context of training psychologists (Lemon, 2023; Wiederhold, 2023).

From a social science perspective, a future research agenda must move beyond this to systematically analyze the uses and effects of AI that can interact in different constellations in different roles and contexts. In the simplest case – exemplified by ELIZA – this concerns constellations of two participants. Such dyads often feature a relationship between service provider and service recipient. For example, ChatGPT can be instructed to write a text about a topic of the user's choosing. This sees a recipient delegate a task to a service provider. In such cases, the AI's output can be evaluated quite clearly and also improved without interference.

The situation changes significantly when AI is used in antagonistic constellations, for example, in cases of conflict and competition (Neuberger, 2022), and where large numbers of people can participate at the same time. That is, the collective action of providers in markets that compete or of political actors in the public sphere that engage in conflict is mostly undetermined and the results largely unpredictable. However, AI can now be used to predict and control these dynamics.

The situation becomes even more complex when agents programmed with antagonistic goals clash. Such "collective machine behavior" (Borch, 2022) can already be observed in the context of financial markets, where decisions on buying and selling are made on the basis of machine learning:

> [T]his strategy works fine if there is only one algorithm doing this at the same time. However, when several independent execution algorithms are

doing this simultaneously, they are likely to adversely affect one another and generate unanticipated collective interaction dynamics. (p. 513)

This also renders market crashes unpredictable and difficult to explain retrospectively. In the future, such dynamics of AI-based collective action must also be expected in the formation of public opinion during election campaigns in democracies or in the case of wars. Describing and explaining these dynamics is likely to become one of the major scientific challenges.

The Societal Value of AI

Equally important is the appropriate regulation of AI (see Zech and also Bryson in this issue of the WJDS) based on the values of liberal democracy, such as freedom, equality, justice, and truth (Coeckelbergh, 2022). However, this requires not only avoiding risks but also using the opportunities offered by AI to improve democracy. ChatGPT can promote equality because it helps those who have difficulty articulating their thoughts in words. Chatbots can increase the quality and success of public deliberation if they suggest better formulations for user statements (rephrasing) that are more convincing, polite, and empathetic (Argyle et al., 2023). AI could also be used to mediate conflicts in the context of peace journalism (see Malik et al. in this issue of the WJDS). Identifying these kinds of positive use cases for AI represents another important task.

Against the Overestimation of AI

Beyond the actual capability of AI, Weizenbaum saw as a danger AI's overestimation by humans who attribute superior intelligence to it. This means that he was primarily concerned with communication about AI and its consequent image. As a public intellectual, he wanted to correct this false image, leading to his considerable influence on the public discourse on AI with his narrative about the "ELIZA effect" (Natale, 2019; Switzky, 2020). However, Weizenbaum was by no means alone in his critical view, as demonstrated by his place in the history of ideas and discourse (Gill, 2019; Loeb, 2021; Pruijt, 2006, pp. 520 – 522).

Weizenbaum warned against propagating a false image of humanity, namely that of the completely predictable human being and its subsequent displacement by AI. He had the foresight to criticize the belief in the omnipotence of the computer and the willingness of people to voluntarily subordinate themselves and their thinking to the machine.

Weizenbaum also advised recognizing the limits of technology and using the computer in a value-based way (Weizenbaum, 1976, 1986, 2008; see also Pörksen in this issue of the WJDS). Furthermore, he saw a special responsibility for computer specialists in public, warning in an essay in Science (1972),

> Given these dismal possibilities, what is the responsibility of the computer scientist? First[,] I should say that most of the harm computers can potentially entrain is much more a function of properties people attribute to computers than of what a computer can or cannot actually be made to do. The nonprofessional has little choice but to make his attributions of properties to computers on the basis of the propaganda emanating from the computer community and amplified by the press. The computer professional[,] therefore[,] has an enormously important responsibility to be modest in his claims. This advice would not even have to be voiced if computer science had a tradition of scholarship and of self-criticism such as that which characterizes the established sciences. (p. 614)

In sum, Weizenbaum's work comprises many points of reference and inspiration for present and future research in the realm of digitalization, from addressing fundamental questions and interdisciplinarity (i.e., combining technological and societal perspectives) to reorienting values and using public appearances to impact the discourse on AI. This is what we are guided by at the Weizenbaum Institute.

The Special Issue

In this special issue, we bring together a series of articles directly or indirectly linked to the work of Joseph Weizenbaum. The first three contributions deal directly with Weizenbaum. Christiane Floyd reflects on the development of ChatGPT in the context of Weizenbaum's analyses. She examines the system in terms of accuracy, structure, context, perspective, and bias, identifying its key problems. David Berry discusses Weizenbaum's work on his chatbot, building on recent reconstructions of ELIZA's original code and using this to reflect on the current state of chatbots such as ChatGPT. That work is complemented by an interview conducted with Weizenbaum by Bernhard Pörksen, which is published here for the first time in English. In this interview, Pörksen talks to Weizenbaum about the development of AI and the philosophical foundations of Weizenbaum's work.

Importantly, Weizenbaum was not only a computer scientist but also a public intellectual in the progressive tradition. He was committed to the anti-war movement and the humane use of technology, with two contributions in this Special Issue entangled with this dimension of Weizenbaum's work. The article by Sehl, Malik, Kretzschmar, and Neuberger addresses the development of peace journalism in times of digitalization and identifies conditions for the successful work of peace journalists. The article by Schmid, Guntrum, Haesler, Schultheiß, and Reuter focuses on volunteerism and care work on social media and develops a feminist ethic of care for social media interactions.

The special issue concludes with two debate contributions in our "Voices for the Networked Society" section. Herbert Zech discusses key points for the regulation of AI in the context of the European AI Act debate. Joanna Bryson discusses the framework conditions for AI regulation, the EU's regulatory approach, and addresses perspectives on digital technology ethics.

We wish you an inspiring reading.

References

Argyle, L. P., Bail, C. A., Busby, E. C., Gubler, J. R., Howe, T., Rytting, C., Sorensen, T., & Wingate, D. (2023). Leveraging AI for democratic discourse: Chat interventions can improve online political conversations at scale. *Proceedings of the National Academy of Sciences*, *120*(41), e2311627120. https://doi.org/10.1073/pnas.231162712

Bassett, C. (2019). The computational therapeutic: Exploring Weizenbaum's ELIZA as a history of the present. *AI & Society*, *34*, 803–812. https://doi.org/10.1007/s00146-018-0825-9

Bloom, P. (2017). *Against empathy: The case for rational compassion*. Ecco.

Borch, C. (2022). Machine learning and social theory: Collective machine behaviour in algorithmic trading. *European Journal of Social Theory*, *25*(4), 503–520. https://doi.org/10.1177/13684310211056010

Bowman, N. D., & Banks, J. (2021). Player-avatar identification, relationships, and interaction: Entertainment through asocial, parasocial, and fully social processes. In P. Vorderer & C. Klimmt (eds.), *The Oxford handbook of entertainment theory*. Online edition. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190072216.013.36

Breithaupt, F. (2019). *The dark sides of empathy*. Cornell University Press.

Coeckelbergh, M. (2022). *The political philosophy of AI*. Polity.

Dibble, J. L., Hartmann, T., & Rosaen, S. F. (2016). Parasocial interaction and parasocial relationship: Conceptual clarification and a critical assessment of measures. *Human Communication Research*, *42*(1), 21–44. https://doi.org/10.1111/hcre.12063

Dodgson, N. (2023). Artificial intelligence: ChatGPT and human gullibility. *Policy Quarterly*, *19*(3), 19–24.

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2023). Generative AI. *Business & Information Systems Engineering*. Online first. http://dx.doi.org/10.2139/ssrn.4443189

Gill, K. S. (2019). From judgment to calculation: The phenomenology of embodied skill. Celebrating memories of Hubert Dreyfus and Joseph Weizenbaum. *AI & Society*, *34*, 165–175. https://doi.org/10.1007/s00146-019-00884-0

Görnemann, E., & Spiekermann, S. (2022). Emotional responses to human values in technology: The case of conversational agents. *Human–Computer Interaction*. Online first, 1–28. https://doi.org/10.1080/07370024.2022.2136094

Guriev, S., & Treisman, D. (2022). *Spin dictators: The changing face of tyranny in the 21st century*. Princeton University Press.

Hillebrand, F. (2023, September 21). Steffi sagt, es ist Liebe. Eine Frau hat sich in einen Chatbot verliebt. Er hat ihr einen Heiratsantrag gemacht. Ist das gefährlich – oder die Zukunft? *Die Zeit*, 60 – 61. https://www.zeit.de/2023/40/beziehung-chatbot-replika-kuenstliche-intelligenz-smartphone

Horton, D., & Wohl, R. (1956). Mass communication and para-social interaction: Observations on intimacy at a distance. *Psychiatry*, *19*(3), 215 – 229.

Lemoine, B. (2022, June 11). Is LaMDA sentient? – an interview. *Medium*. https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917

Lemon, C. A. (2023). Chatbots in social psychiatry education: A social phenomenon. *International Journal of Social Psychiatry*. Online first, 1 – 2. https://doi.org/10.1177/00207640231178484

Levy, S. (2022. June 17). Blake Lemoine says Google's LaMDA AI faces 'bigotry.' In an interview with WIRED, the engineer and priest elaborated on his belief that the program is a person – and not Google's property. *Wired*. https://www.wired.com/story/blake-lemoine-google-lamda-ai-bigotry/

Loeb, Z. (2021). The lamp and the lighthouse: Joseph Weizenbaum, contextualizing the critic. *Interdisciplinary Science Reviews*, *46*(1 – 2), 19 – 35. https://doi.org/10.1080/03080188.2020.1840218

Metz, C. (2023, January 20). How smart are the robots getting? The Turing test used to be the gold standard for proving machine intelligence. This generation of bots is racing past it. *The New York Times*. https://www.nytimes.com/2023/01/20/technology/chatbots-turing-test.html

Moses, J., & Meldman, J. (2008). In memoriam. Joseph Weizenbaum (1923 – 2008). *IEEE Intelligent Systems*, *23*, 8 – 9. https://doi.org/10.1109/MIS.2008.70

Natale, S. (2019). If software is narrative: Joseph Weizenbaum, artificial intelligence and the biographies of ELIZA. *New Media & Society*, *21*(3), 712 – 728. https://doi.org/10.1177/1461444818804980

Neuberger, C. (2022). How to capture the relations and dynamics within the networked public sphere? Modes of interaction as a new concept. In B. Krämer & P. Müller (eds.), *Questions of communicative change and continuity. In memory of Wolfram Peiser* (pp. 67 – 95). Nomos. https://doi.org/10.5771/9783748928232-67

Procter, L. (2021). I am/we are: Exploring the online self-avatar relationship. *Journal of Communication Inquiry*, *45*(1), 45 – 64. https://doi.org/10.1177/0196859920961041

Pruijt, H. (2006). Social interaction with computers: An interpretation of Weizenbaum's ELIZA and her heritage. *Social Science Computer Review*, *24*(4), 516–523. https://doi.org/10.1177/0894439306287247

Scudder, M. F. (2020). *Beyond empathy and inclusion: The challenge of listening in democratic deliberation*. Oxford University Press.

Switzky, L. (2020). ELIZA effects: Pygmalion and the early development of artificial intelligence. *Shaw*, *40*(1), 50–68. https://doi.org/10.5325/shaw.40.1.0050

Tsai, W.-H. S., & Chuan, C.-H. (2023). Humanizing chatbots for interactive marketing. In C. L. Wang (ed.), *The Palgrave handbook of interactive marketing* (pp. 255–273). Palgrave Macmillan. https://doi.org/10.1007/978-3-031-14961-0_12

Tukachinsky Forster, R. (ed.) (2023). *The Oxford handbook of parasocial experiences*. Online edition. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780197650677.001.0001

Voss, J. (2023, July 20). Mein Partner, der Chatbot: Kann man sich in künstliche Intelligenz verlieben? *National Geographic*. https://www.nationalgeographic.de/wissenschaft/2023/06/mein-partner-der-chatbot-kann-man-sich-in-kuenstliche-intelligenz-verlieben

Weizenbaum, J. (1972). On the impact of the computer on society: How does one insult a machine? *Science*, *176*(4035), 609–614. https://www.jstor.org/stable/1734465

Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation*. W. H. Freeman & Co.

Weizenbaum, J. (1986). Not without us. *ACM SIGCAS Computers and Society*, *16*(2–3), 2–7. https://doi.org/10.1145/15483.15484

Weizenbaum, J. (2008). Social and political impact of the long-term history of computing. *IEEE Annals of the History of Computing*, *30*(3), 40–42. https://doi.org/10.1109/MAHC.2008.58

Wiederhold, B. K. (2023). The virtual patient and you: How AI can enhance both sides of the therapeutic relationship. *Cyberpsychology, Behavior, and Social Networking*, *26*(10), 729–730. https://doi.org/10.1089/cyber.2023.29292.editorial

Wigura, K., & Kuisz, J. (2020). The pushback against populism: Reclaiming the politics of emotion. *Journal of Democracy*, *31*(2), 41–53.