

Editorial: On IRIE Vol. 29

In May of 2018 Sundar Pichai introduced Google Duplex¹, a voice interaction technology that can book an appointment at a hair salon or meal at a restaurant. In the public demo they played a recording of what was presumably an authentic conversation between Duplex and someone at a hair salon. Pichai introduced the recording with “So what you’re going to hear is the Google Assistant actually calling a real salon to schedule the appointment for you.” Who could object to machines making our calls?

But no, at no point did Duplex identify itself as an AI to its human interlocutors. Google seemed to take pride in how good Duplex was at deceiving people. The audience of developers clapped, but the next day others pointed out the ethical issues. Zeynep Tufekci pointed out on Twitter, that Silicon Valley seemed to be ethically lost.

Google’s response was predictable ... principles to the rescue. Pichai introduced some vague principles a month later which alas, didn’t include anything explicitly about transparency or disclosure.² The closest they came was principle 4, “Be accountable to people” which they explain thus,

We will design AI systems that provide appropriate opportunities for feedback, relevant explanations, and appeal. Our AI technologies will be subject to appropriate human direction and control.

Even if providing relevant explanations includes explaining to people when they are working with an AI, we have to ask who the “we” is and how will Google ensure that they live up to their principles. And that is the deeper problem we all face, namely the gap between principles and a healthy culture of ethics. Principles, guidelines, codes of ethics, and other lists are all accessible ways to start thinking about ethics, but they are not a guarantee (Mittelstadt 2019). They communicate a commitment, but the hard work is getting beyond the announcement to the sustained engagement that follows and that is work we at the IRIE can contribute to.

This leads me to a research agenda for information ethics at this moment when there is public concern about the ethics of AI; and that is to ask what we can know about the development of a culture ethics in new technical fields like AI. There are all sorts of well-meaning initiatives to promote good AI. Many of these have developed principles and many of them offer tools or services, but few are looking to the rich philosophical tradition of information ethics. We can contribute by thinking through the ethics of ethical principles, to paraphrase Hagendorff (2020). We can think about how to go beyond principles in order to care about ethics in the information sector, and not just how to care today, but to continue to care. We can draw from across cultures of ethics to imagine an *ethos of ethics*.

Sincerely Yours,

Geoffrey Rockwell

¹ On Duplex see <https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html>. To see Pinchai’s presentation see <https://www.youtube.com/watch?v=D5VN56jQMWM>.

² See “AI at Google: our principles”, <https://blog.google/technology/ai/ai-principles/>.

References

Hagendorff, T. (2020) "The Ethics of AI Ethics: An Evaluation of Guidelines." *Minds & Machines* 30, 99–120.
<https://doi.org/10.1007/s11023-020-09517-8>

Mittelstadt, B. (2019). "Principles alone cannot guarantee ethical AI." *Nature Machine Intelligence* 1, 501-507.
<https://doi.org/10.1038/s42256-019-0114-4>