Responsible Technology Design: Conversations for Success

Susan J. Winter and Brian S. Butler

Abstract Digital humanism calls for new technologies that enhance human dignity and autonomy by educating, controlling, or otherwise holding developers responsible. However, this approach to responsible technology design paradoxically depends on the premise that technology is a path to overcoming human limitations while assuming that developers are themselves capable of super-human feats of prognostication. Recognizing developers as subject to human limitations themselves means that responsible technology design cannot be merely a matter of expecting developers to create technology that leads to certain desirable outcomes. Rather, responsible design involves expecting the technologies to be designed in ways that provide for active, meaningful, ongoing conversations between the developer and the technology, between the user and the technology, and between the user and the developer—and expecting that designers and users will commit to engaging in those conversations.

Digital humanism calls for new technologies that enhance human dignity and autonomy by infusing ethics into the design process and into norms and standards. These calls are even echoed by politicians in the international arena (Johnson 2019):

the mission ... must be to ensure that emerging technologies are designed from the outset for freedom, openness and pluralism, with the right safeguards in place to protect our peoples. ... we need to agree on a common set of global principles to shape the norms and standards that will guide the development of emerging technology (Boris Johnson, Address to the UN, 2019)

Although we should strive to achieve this vision of responsible technology design, we should also realize that we cannot meet it. Technologies are created by people. Both technologies and people are limited. Because of these limitations, the best we can hope for is that new technologies allow for ongoing dialogue that recognizes and responds to the need for human dignity and autonomy.

S. J. Winter (✉) · B. S. Butler
University of Maryland, College Park, MD, USA
e-mail: sjwinter@umd.edu; bsbutler@umd.edu

While there are many nuances and variations, one of the simplest ways to understand the foundational premise of digital humanism is to consider a digital humanist in contrast with a digital technologist. Both want to create a better world and improve the human condition. Both see technology as a way of overcoming human limitations and making things better.

Where they differ is that digital technologists see the creation of technologies that eliminate the need for human involvement through automation as the primary path to sustained, substantial improvement in the human condition. Self-driving cars seek to remove the need for a human driver. AI-based facial recognition seeks to remove the human from the recognition process. In contrast, digital humanists pursue change by encouraging development of technologies that place humans at the center, empowering them to enact their own well-being. Wikis enable humans to create collections of useful information for efficient learning. Social platforms allow people with similar goals to create mutually supportive communities of practice.

However, while they differ significantly with respect to the role of humans in the application of technology to improve the human condition, one thing that many digital technologists and humanists share is an assumption about the relationship between developers and the technologies they create. Whether it is implicit or explicit, it is often assumed that developers create technologies which in turn shape the actions and choices of the users (Gibson 1977). The actions of developers lead to the different technologies existing (or not), having particular features (or not), and creating affordances which enable (or prevent) users from taking particular actions or making particular choices (e.g., Anderson and Robey 2017).

It is this assumed power of the developer to shape the technology and subsequent user behavior that is the basis of efforts to bring about responsible technology design by educating, controlling, or otherwise holding developers responsible. Efforts to infuse ethics into Computer Science curricula reflect this assumption, for example, Embedding EthicsTM @ Harvard which:

embeds philosophers directly into computer science courses to teach students how to think through the ethical and social implications of their work. (<https://embeddedethics.seas.harvard.edu/>)

The premise is that if we nurture developers' appreciation of ethics, the self-driving cars that they create will be safe and efficient, expand mobility to underserved communities, and create a better, more equitable world. If we sensitize developers to the need for privacy, they will build privacy protections into AI for facial recognition that will enhance personal and societal security without intrusive over-surveillance. Educated, thoughtful, ethically aware developers will design wikis that reject unverified and inaccurate information. Developers who are aware of the ethical and social implications of their work will encourage regulation that will result in platforms that support pro-social communities of practice and block those that are pursuing goals of violence and hate.

Somewhat paradoxically, this approach to responsible technology design is based on the premise that technology is a path to overcoming human limitations while also assuming that developers are themselves capable of practically super-human feats of

prognostication and influence. Ethically trained, sensitized, well-regulated developers will still be surprised by how, when, and why their “responsibly designed” self-driving cars, facial recognition software, wikis, and social media platforms will be deployed. Contexts are infinitely variable, and users are infinitely “creative.” Assuming that a developer (or a user for that matter) will “get it right” drastically overestimates humans’ ability to imagine, anticipate, and influence the functioning of even the most basic socio-technical systems.

Responsible technology design cannot be merely a matter of expecting developers to create technology that leads to certain desirable outcomes. To posit this definition of responsible design necessarily requires a capability that is beyond the reach of any human designer and will lead to expectations about developer responsibilities and obligations that are at best unreasonable and at worst dangerously misguided.

Rather, responsible design involves expecting the technologies to be designed in ways that provide for active, meaningful, ongoing conversations between the developer and the technology, between the user and the technology, and between the user and the developer and expecting that designers and users will commit to engaging in those conversations. It is well within our ability to create systems and technologies that provide the affordances for iterative designer-tech, user-tech, and designer-user conversations. Indeed, in the face of the human limitations outlined above, the only forms of responsible technology design that are feasible are those based on repeated iterative, active, adaptive engagement with the technology by developers. Instead of defining success as developers creating a responsible design, we must expect that they engage in the never-ending process of responsible design.

One common approach to enabling these conversations involves developers incorporating (and committing to using) affordances that allow them to collect and attend to data about the technology. Self-driving cars record their state and action. AI facial recognition technologies track classification choices. Wikis support tracking of content changes and user actions. Social platforms track blocked content. This is seen by some as the minimum required for responsible technology design. Yet ultimately this approach still assumes that developers will have the somewhat super-human ability to use the data to review, monitor, track, and interrogate the performance and use of the technology while also having the control needed to not misuse the data.

As the quote from Boris Johnson above indicates, concerns about the outcomes of new technologies are not limited to the simple performance of the technological components of systems. Responsible design implies responsible engagement with the larger socio-technical system and the processes by which meaning, purpose, and values emerge. Developers will have to build in (and commit to using) affordances that allow them to collect and attend to data about not just the technological component, but the larger socio-technical system in which it is embedded. Records of actions by self-driving cars are important, but equally so are the choices of the driver. Responsible design of AI and facial recognition requires attention to issues of accuracy, but also to issues of appropriateness in their application. Responsibly designed wikis must filter misinformation and disinformation, but also must choose

how to balance the desires of readers with the inclusion of silenced voices and peoples. Social platforms must track the volume of posts and types of content, but must also continually consider trade-offs between the economic goals of the providers and the civic goals of the larger society. Engaging in dialog with a system requires that developers engage with these issues as well while balancing the needs for privacy and security.

These conversations can occur at multiple levels and in diverse forms. An agile or co-design approach creates a direct dialog between users and designers that is much richer than is possible with the waterfall method of development. Regulation also puts users, developers, and technologies in conversation with one another. Users can also use the marketplace to express their preferences. Of course, responsible technology design cannot mean that a developer is responsible for all of the outcomes of the technology. Rather, we argue that they are responsible for creating and engaging in systems that support ongoing dialog, engagement, and adaptation between developers, technological elements, and other stakeholders.

By their very nature, digital technologists necessarily set themselves up with a fundamentally harder problem with respect to enabling responsible design. By setting their sights on eliminating meaningful involvement of humans in systems through automation, they necessarily make the dialog between developers and those systems more difficult to support and achieve. Building in traceability, detailed logs, exception reports, and an extensive investigative operation to review and respond to this data in a timely fashion become essential. At best, building this capability for a more responsible system requires substantial additional cost and effort. At worst, it requires developers to incorporate features and functions which are counter to the goals of automation, setting responsible design up in opposition to what a digital technologist considers to be an effective design process.

In contrast, a digital humanist who seeks to improve the human condition by empowering people is already predisposed to enabling dialog between the human and technical element of a socio-technical system because that dialog is integral to their approach. Incorporating additional features and functions that enable developers to participate in this dialog is therefore a more straightforward proposition and less likely to be seen as counter to the goals of the design process. Digital humanism can help us recognize the limitations of humans and the role that technology can play in empowering humans to overcome those limitations. This is a significant contribution that digital humanism can make. Recognizing the limitations of developers and users and adopting models of responsible design and use that accommodate those limitations by putting these communities in continual conversation could be an even more powerful contribution of digital humanism.

Whether it is self-driving cars enabled by the internet of things, artificial intelligence for facial recognition, self-monitoring wikis, online community platforms, or some other application of emerging technology, it is not the consequences of designers who failed to anticipate the impact of their creations that we should fear. This we must expect even with extensive training in ethical decision-making. There is no other outcome that is possible. Instead, it is the designer who has the hubris to believe that they can fully anticipate the outcomes of their creations—and as a result

fails to allow for and participate in the conversations that are needed to adaptively engage the technology and its implications that are the irresponsible parties who should be the object of concern.

References

- Anderson, C. and Robey, D. (2017). Affordance potency: Explaining the actualization of technology affordances, *Information and Organization*, 27(2), 100-115, 2017.
- Embedding Ethics™ @ Harvard, <https://embeddedethics.seas.harvard.edu/>, retrieved April 15, 2021.
- Gibson, J. L. (1977). A Theory of Affordances. In R. Shaw and J. Bransford (Eds.) *Perceiving, Acting and Knowing: Toward an Ecological Psychology*, Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. pp. 67-82.
- Johnson, B. (2019). Prime Minister speech for the UN General Assembly, Sept, 14, <https://www.gov.uk/government/speeches/pm-speech-to-the-un-general-assembly-24-september-2019>.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.