

Explorative Experiments and Digital Humanism: Adding an Epistemic Dimension to the Ethical Debate



Viola Schiaffonati

Abstract The rise of Digital Humanism calls for shaping digital technologies in accordance with human values and needs. I argue that to achieve this goal, an epistemic and methodological dimension should be added to the ethical reflections developed in the last years. In particular, I propose the framework of explorative experimentation in computer science and engineering to set an agenda for the reflection on the ethical issues of digital technologies that seriously considers their peculiarities from an epistemic point of view. As the traditional epistemic categories of the natural sciences cannot be directly adopted by computer science and engineering, the traditional moral principles guiding experimentation in the natural sciences should be reconsidered in the case of digital technologies where uncertainty about their impacts and risks is very high.

1 Introduction

The rise of Digital Humanism calls for shaping digital technologies in accordance with human values and needs to possibly solve the critical issues of current technological development. Within this framework, ethics plays an increasing role at both a descriptive level and normative one, and, accordingly, several important results have been achieved in the last years. On the one hand, approaches such as the Value Sensitive Design have shifted the attention to the idea of *active responsibility*, that is, the design of technology to incorporate positive values (van den Hoven 2007). On the other hand, several regulatory frameworks have been proposed to address the ethical issues related to digital technologies, such as AI, and their adoption within our society.

Notwithstanding the importance of these initiatives, I argue that a further dimension should be added to this debate. This dimension concerns the analysis of the disciplinary and methodological status of computer science and engineering to better

V. Schiaffonati (✉)
Politecnico di Milano, Milan, Italy
e-mail: viola.schiaffonati@polimi.it

understand the radical paradigm shift promoted by digital technologies. Rather than considering this dimension as alternative to the other ones, I claim that it should be integrated with them to address the current challenges of digital technologies in a more comprehensive way. In this chapter, I focus in particular on the nature and role of experiments in AI and autonomous robotics. The main result of adding this further dimension to the current analysis is to set an agenda for the reflection on the ethical issues of digital technologies that seriously considers their peculiarities from a disciplinary and a methodological point of view. Constructing on some of my previous works, I argue that the traditional epistemic categories of the natural sciences cannot be directly adopted by computer science and engineering as an artificial discipline. Accordingly, the traditional moral principles guiding experimentation in the natural sciences should be reconsidered in the case of digital technologies, where uncertainty about their impacts and risks is very high.

This chapter is organized as follows. Section 2 discusses the nature and role of experiments in computer science and engineering and how experiments are perceived as ways to increase the scientific maturity of the field. Section 3 presents the novel notion of explorative experimentation emerging from the analysis of the practice of AI and autonomous robotics. Section 4 connects epistemic uncertainty, typical of explorative experiments, to the design of ethical frameworks based on an incremental approach. Finally, Sect. 5 concludes the chapter by stressing how explorative experiments can impact on the current shaping of Digital Humanism.

2 Experimental Method and Computing

In the last years, the debate on the nature and role of experiments in computer science and engineering has emerged as one of the ways to stress its scientific status: to adopt the same experimental standards of the natural sciences can make computer science and engineering more mature and credible.

AI and autonomous robotics make no exception. AI, for example, is facing a reproducibility crisis in which the importance of reproducibility is taken for granted: the specificity of reproducibility in AI is not investigated, and, in the end, only practical benefits are evidenced (Gundersen et al. 2018). Autonomous robotics presents two different tendencies (Amigoni et al. 2014). On the one hand, the traditional principles of experimental method (reproducibility, repeatability, generalization, etc.) are seen as golden standards to which the research practice should conform. For example, public distribution of code is promoted to achieve reproducibility. On the other hand, rigorous approaches to experimentation are not yet part of current practices. For example, the use of settings that can be applied to different environments is limited, jeopardizing the possibility of generalizing experimental results.

Only few exceptions have stressed the peculiarity of experimentation in computer science and engineering and emphasized that the term experiment can be used in different ways (Tedre 2015). Moreover, the question whether it does make sense to

apply the same standards of the natural sciences to the artificial ones has been seldom asked. The idea that computer science and engineering is an experimental science of a very special type has been advanced by Allen Newell and Herbert Simon already in the 1970s (Newell and Simon 1976). Even if the invitation to see each new machine as an experiment has remained largely unattended, some exceptions exist: they point out that experimentation is more multifaceted than usually depicted in computer science and engineering.

Two elements are particularly important. First, many experiments have the goal of testing technical artifacts rather than theories. *Technical artifacts* are physical objects with a technical function and use plan designed by humans in order to fulfill some practical functions (Veermaas et al. 2011). Second, in several cases, experimenters are designers, thus losing the independence of the experimenter prescribed in the classical experimental protocol. This is why I have proposed the notion of *explorative experimentation* to give reason of a part of the experimental practice in computing that cannot be subsumed under the traditional categories of the epistemic and controlled experimentation typical of the natural sciences (Schiaffonati 2020).

3 A Different Notion of Experimentation: Explorative Experiments

Explorative experiments are a technological form of experimentation devoted to test technical artifacts. They can be seen in continuity with the tradition of the so-called directly action-guiding experiments, that is, those experiments devoted to action and contraposed to traditional *epistemic experiments* devoted to knowledge (Hansson 2015). For example, a systematic test on an autonomous robot employed to assist an elderly person in her home is a technological form of experimentation, where the outcome looked for is the proper interaction of the robot with the person and the intervention is the careful tuning of the abilities of the robot to achieve the goal.

Moreover, explorative experiments have a normative component that epistemic ones do not possess. They are carried out to check whether the technical artifacts meet the desired specifications via their technological production. The normative element consists in determining how much the tested technical artifact conforms to its design. A robot system, for example, can be evaluated as better or worse with respect to a given function that works as a reference model. On the contrary, a natural phenomenon (which is usually what is investigated in an experiment in the natural sciences), such as an electron, cannot be good or bad: the electron in the experiment is evaluated without any reference to its supposed technical function, hence without any normative constraint with respect to its correct functioning.

To summarize, explorative experiments are devoted to test technical artifacts without the control boundaries typical of an epistemic controlled experiment. Their goal is to investigate the possibilities and limits of the technical artifact and its interaction with the surrounding environment. The design of this investigation is not

conducted on the basis of a well-formed theory or a systematic theoretical background. Rather, the initial hypotheses cannot always be formulated in a clear way, and the type of knowledge which is the goal of this experimentation is oriented to evaluate the performance of the technical artifact with respect to its technical function. The experimenter is often the same designer of the artifact, and, thus, her independence from the experimental context, as in the traditional epistemic experimentation, is not guaranteed. In conclusion, explorative experiments are not devoted to reject or accept a general theory, but to probe (iteratively) the possibilities and limits of the intervention. This makes them similar to some methodological reflections developed in the field of design science research, where the iterative and evolutionary nature of design improvements through exploration is emphasized (Gil and Hevner 2013).

4 From Epistemic Uncertainty to Ethical Incrementalism

The explorative experiment framework highlights how uncertainty plays an essential role at a theoretical level and how this has an impact on the experimental procedures that must renounce to a part of the experimental control traditionally associated to epistemic experiments. When considering AI and autonomous robotics, uncertainty concerns both the behavior of the complex systems themselves and their interactions with humans and complex environments. For this reason, they can be labelled *experimental technologies*: the operational experience relative to their effective behavior is limited, and, therefore, the attempts to precisely assess their societal risks and benefits are uncertain: this means that their impact on humans and societies is mostly unknown and difficult to predict (van de Poel 2016). To acknowledge uncertainty in the development of experimental technologies means to recognize that unexpected events can always occur and that a different approach is required to deal with their development and management. This approach is a form of *incrementalism*, where experimental technologies are gradually introduced into society to constantly monitor the societal effects that emerge and iteratively improve their design accordingly. In other words, the epistemic uncertainty emerging from an epistemological perspective in the case of explorative experimentation can be translated into a form of incrementalism from an ethical perspective. Explorative experiments, devoted to acquiring knowledge on the behavior of these experimental systems in the real world, are therefore crucial to address the ethical issues related to the impact on such systems on society.

Some ethical frameworks have already proposed to deal with these experimental technologies. For example, van de Poel (2016) incorporates the traditional principles of bioethics (beneficence, non-maleficence, respect for autonomy, respect for justice) and declines them in an incremental ethical framework. I have argued elsewhere that this a promising starting point in particular when integrated with explorative experiments (Amigoni and Schiaffonati 2018). This means that to concretely minimize the risks associated to experimental technologies, the first step is to understand

what it really accounts to experiment on technical artifacts. The framework of explorative experiments has thus an impact not only at the methodological level but also at the ethical one, where the traditional moral categories need to be revised to deal with experimental technologies.

5 Conclusion

In this chapter, I have suggested that to address some of the issues connected to digital technologies, the development of appropriate techniques is not enough. Rather, I have shown that some problems have to be addressed with methods having a philosophical nature.

To conclude, I emphasize how the framework of explorative experimentation is connected to the larger issue of the societal impact of digital technologies. The problem of how this approach can be better adopted in practice remains open. Yet I argue that a shift in the conceptualization has, at least, two important roles. The first one concerns the influence of novel epistemic categories, such as explorative experiments, on ethical ones, as I have discussed in Sect. 4. The second level regards the development of the disciplines of the artificial, to which computer science and engineering belong, by starting from methodological reflections. This is not only a disciplinary issue, but has an impact on how humans, digital technologies, and their interactions are conceptualized in the current discussion on Digital Humanism. If one of the goals of Digital Humanism is to “shape technologies in accordance with human values and needs, instead of allowing technologies to shape humans,” it is essential to recognize the centrality of technical artifacts and sociotechnical systems in the disciplines of the artificial. Sociotechnical systems are composed of physical objects, people, organizations, institutions, conditions, and rules. They, thus, have a hybrid character as they consist of components which belong in many different “worlds”: not only those requiring a physical description but also those requiring a social one (Veerma et al. 2011). So far, the components requiring a physical description have been addressed by scientific and engineering disciplines. Now it is time to consider all the components requiring a social description, like the ones promoted by the humanities and the social sciences, and to develop, accordingly, the new field of the artificial disciplines which should include and integrate both in a creative way.

References

- Amigoni, F., Schiaffonati, V., Verdicchio, M. (2014) ‘Good Experimental Methodologies for Autonomous Robotics: From Theory to Practice’, in F. Amigoni, V. Schiaffonati (eds.), *Methods and Experimental Techniques in Computer Engineering*, SpringerBriefs in Applied Sciences and Technology, Springer, pp. 37-53.

- Amigoni, F. and Schiaffonati, V. (2018) 'Ethics for Robots as Experimental Technologies: Pairing Anticipation with Exploration to Evaluate the Social Impact of Robotics', in *IEEE Robotics and Automation Magazine*, 25, n. 1, pp. 30-36.
- Gil, T. G. and Hevner, A. N. (2013) 'A Fitness-Utility Model for Design Science Research', in *ACM Transactions on Management Information Systems*, 4 (2): 5-24.
- Gundersen, O. E., Aha, D., Gil, Y. (2018) 'On Reproducible AI: Towards Reproducible Research, Open Science, and Digital Scholarship in AI Publications', in *AI Magazine*, 39, n. 3, pp. 56-68.
- Hansson, S.O. (2015) 'Experiments before Science? – What Science Learned from Technological Experiments', in Sven Ove Hansson (ed.) *The Role of Technology in Science*, Springer.
- Newell, A. and Simon, H. (1976) 'Computer Science as Empirical Inquiry: Symbols and Search', *Communications of the ACM* 19 (3): 113-126.
- Schiaffonati, V. (2020). *Computer, robot ed esperimenti*, Milano: Meltemi.
- Tedre, M. (2015) *The Science of Computing*. Boca Raton: CRC Press, Taylor & Francis Group.
- van de Poel, I. (2016) 'An Ethical Framework for Evaluating Experimental Technology', in *Science and Engineering Ethics*, 22, pp. 667-686.
- van den Hoven J. (2007) 'ICT and Value Sensitive Design', In Goujon P., Lavelle S., Duquenoy P., Kimppa K., Laurent V. (eds.) *The Information Society: Innovation, Legitimacy, Ethics and Democracy In honor of Professor Jacques Berleur s.j.*, IFIP International Federation for Information Processing, vol 233, Boston: Springer.
- Veermaas, P., Kroes, P., van de Poel, I., Franssen, M., Houkes, W. (2011) *A Philosophy of Technology: From Technical Artefacts to Sociotechnical Systems*, Morgan & Claypool Publishers.

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.

