REPORT

Digitization – hopes and fears

Paul Grünke, *Institute for Philosophy, Karlsruhe Institute of Technology (KIT), Douglasstraße 24, 76133 Karlsruhe (paul.gruenke@kit.edu), https://orcid.org/0000-0002-3576-1921*
Aleksandra Kazakova, *Bauman Moscow State Technical University (kazakovaz@mail.ru), https://orcid.org/0000-0002-2952-8373*

The Budapest Workshop on Philosophy of Technology, hosted by the Budapest University of Technology and Economics on the 12th and 13th of December 2019, drew around 50 scholars from philosophy, history, sociology and Science and Technology Studies. Many of the talks approached the topic of Artificial Intelligence (AI) from socio-historical, epistemological or ethical perspective. Mark Coecklbergh gave the keynote speech and offered a comprehensive overview of ethical issues and policy directions concerning AI with a focus on Europe.

History and philosophy of technology

The philosophical legacy was revised with regard to current issues, for example by two reports discussing the "empirical turn" of the 1980s from different comparative perspectives.

Agostino Cera described the "empirical turn" as "transition from an over-distance to an overproximity". The new relativist and contextual approach to "technologies" (in the plural) confronted essentialism, determinism and dystopianism of the "classics", especially Heidegger. For Cera, this blurred the distinction between "problems" (requiring solutions) and "questions" (remaining open). "Engineerization of philosophy of technology" as problem-solving activity converged with scientific, technical or political expertise. Cera argued for leaving space for the genuinely philosophical "Grundfrage": the question of the historical or epochal phenomenon of technology as such.

For Darryl Cressman, the distinction between classical and empirical approaches is not relevant for critical philosophy. Marxist thought, he claimed, was always empirically oriented, recognizing the historical contingency of technology and disentangling reification of socio-economic interests in it. The "negative dialectics" reveals the inherent biases and reflects on alternatives, or "unrealized potential" of technology. This can be seen in the practices and imaginaries of users, which "contradict, or negate, the prescribed design and function". In our view, this focus on agency in production and use of technologies is in line with the objectives of participatory technology assessment (TA) and may become a conceptual resource for it.

Phil Mulins addressed the question of technological imaginaries of digitization, reconstructing Polanyi's participation in the mind-machine debate since the late 1940s. His concept of personal and tacit knowledge confronted both Cartesian dualism and reductionist theory of mind and underlay his scepticism about unlimited formalization, even though he acknowledged its indispensability for the complex modern societies. This makes Polanyi's post-critical philosophy relevant for the current debate on digitization, including questions of algorithmic bias and discrimination. Mulins formulated the task for TA nowadays: "Society needs both to set limits upon predictive analytics and to recognize better the potential of predictive analytics to promote the good society".

Epistemology

Two talks touched on epistemological questions concerning neural networks. Paul Grünke used the development of the neural network based chess engine AlphaZero to illustrate the different kinds of epistemic opacities, which are present in classical modeling and computer simulations compared to machine learning techniques. He claimed that the opacities in the traditional approaches are contingent, but machine learning introduced a fundamental kind of opacity. This model-opacity can make it impossible to trace the origin of specific results, because the underlying structure of the neural network cannot be understood.

Reto Gubelmann analyzed a claim made in the natural language processing community: the results of new translation systems using "neural machine translation" cannot be distinguished from those of professional human translators. They regard it as a clear step towards strong AI. Comparing this claim to Searle's classical Chinese Room thought experiment, Gubelmann showed that the techniques used (semantic representations via word embedding) are more advanced than classical statistical approaches and it could be argued that these systems do have some kind of linguistic understanding. Opacity of machine learning processes, however, limits our understanding of the machine's understanding.

Political science and media studies

A few talks focused on ideological implications of digitization. Jernej Kaluza explored fragmentation and polarization of the digital environment, manifesting itself both in collective and individual social action: from the rise of alt-right politics to mass shootings. He claimed that "digital hate" cannot be explained in psychological terms, but is a mutual reinforcement of individual choices and algorithms of personalization, creating the "endless You-loops", or a "static and narrowing version of oneself". Marginalization of hate speech, he argued, has the backfire effect of creating even more radicalized "rabbit holes".

https://doi.org/10.14512/tatup.29.1.60

Jacopo Bodini linked the subjective experience with the totality of digital culture. He claimed that hypothesis of the coming post-ideological era did not grasp recent metamorphosis of "ideology": from political or philosophical metanarrative to aestheticization of its own technological base. Its major feature is "transparency", which paradoxically conceals itself. Subjectively, it means ignoring the technological mediation of experience. The logic of "immediation" can be seen in the rise of populism and fake news, but also in the immersive way of digital life, shaping ("in-forming") individual perceptions and desires.

Ricardo Rohm investigated agency in opinion-making during the political crisis in Brazil, describing it as "architecture of misinformation". The interaction between "corporatocracy" and technocracy (IT-corporations and social platforms, advertisers and lobbies), he claimed, makes political marketing a real threat to representative democracy. The interplay of national and transnational actors reflects how through the globalization of politics civil society loses control at the national level. South American political systems are especially vulnerable, since their recent democratization evolves against the background of the digital revolution.

This discussion raises the further question for TA: what are the effective means of regulating the digital public sphere – both nationally and globally – for maintaining the argumentative and representative discourse?

Responsibility and decision-making

One of the salient topics during the whole conference was the question how to distribute accountability and responsibility of human actors, making decisions based on the suggestion of an automated expert system. A typical illustration: a doctor uses an expert system, which has a higher success rate of suggesting a treatment than the doctor him/herself. As a result, less time is allocated for decision making of the human, which discourages their disagreement with the expert systems, while in legal terms full responsibility and accountability are assigned to the doctor.

Chang-Yun Ku addressed this problem in a thought experiment of a conflict between a doctor's diagnosis and an AI recommendation: AI suggests a treatment and the doctor believes it is not necessary. Assuming that this fictional doctor would in this case be very likely to advocate the treatment, Ku pointed out a tension with Article 22 (1) of the General Data Protection Regulation, which requires that decisions, which will have significant impact on a person, should not be made solely by automated systems. Legally speaking, the decision is made by the doctor, but it can be argued whether he/she had enough autonomy in these circumstances.

Mark Coeckelbergh touched on similar problems, offering some insights into the ideas currently guiding European policy makers. Central to European policy is to ensure that humans make the decisions. Besides the philosophical problems of assigning responsibility to machines with their non-existent agency, he also referred to the "many hands problem", which makes it difficult to trace exactly who is responsible for the decision of a machine. In order for the human decision to be autonomous and at the same time, to include the recommendations by expert system, the human has to be able to understand its "reasoning". Efforts in the area of Explainable AI must be directed to enabling the decision maker, e. g. a doctor, to acquire enough information about the expert system to explain its result. Furthermore, the regulations on the use of AI technology must leave enough time for reflection on the results of the expert system. Otherwise, a decision will likely be replaced by a post-hoc rationalization.

How to distribute accountability and responsibility of human actors making decisions based on the suggestion of an automated expert system?

Many participants were critical of the idea that a human will always be legally responsible for decisions in these contexts. Going forward, there seem to be three different options: 1) accept the situation and act as described above; 2) find ways to distribute legal responsibility to the expert system; 3) reach a societal consensus that expert systems should not be used in some areas. The discussion did not result in a clear preference between these options.

The general leitmotif of this interdisciplinary workshop may be summarized as follows: what is human (knowledge, agency, responsibility) and what is not in the process of digitization. This theoretical question, however, is intertwined with a practical one (in a Kantian sense): what may we, a human society, hope to be? For achieving concrete results of TA, it seems that these questions need to be considered together.