

Summary Breakout-Session 01: Artificial Intelligence and Governance: Liberty, Trust, Security

Artificial Intelligence (AI) and big data have become increasingly important issues in public debates in recent years. Science, politics and business are hoping for new insights and more robust, presumably "evidence-based" recommendations for action. What are the political, social and ethical implications of "automated" knowledge production, circulation and use? What are the potentials, limits and risks? How do public policy and administration productively deal with the ethical challenges of AI and big data, and what might European answers to these challenges look like? What kind of (new) governance will be needed? A dedicated group of speakers discussed these questions from various points of view.

Patrice Chazerand, Director for Trade and Taxation at DIGITALEUROPE, starts the afternoon by talking about the complex challenges posed by AI and big data. He uses the example of the High-Level Expert Group (HLEG) on AI, in which DIGITALEUROPE participated. He argues that it is necessary to work together at the European level - while encouraging global partnerships - to be able to realize a governance model for AI and big data that not only protect EU ethics values, but also strengthen these values for the benefit of the European society and its economy. He introduces seven principles for AI governance developed by the HLEG: Transparency, robustness & safety, diversity and non-discrimination, human agency & oversight, societal & environmental wellbeing, accountability, and privacy & data governance.

After this introduction to the topic from an industry and policy perspective, **Sepp Hochreiter** talks about his current work and potential future projects on AI at the Johannes Kepler University Linz. He explains how deep learning has emerged as one of the most successful domains of artificial intelligence, with overwhelming success in industrial vision, speech and text processing. Zooming in on potential applications in governance, Hochreiter sketches out better services for citizens together with automated processing of service requests and talks about improved methods for anticipating road maintenance and replacement needs. This might lead to increased efficiency, better informed citizens, less crime, increased safety, improved healthcare, and better responses to natural disasters. While generally painting a positive picture of AI technologies, Sepp Hochreiter argues that the application of AI technology in governance must be carefully evaluated with respect to moral, ethical, social, and political issues.

Jack Stilgoe, Associate Professor at the University College London, introduces the notion that even the most mundane technologies have political qualities. Using the example of the introduction of the mechanical tomato harvester to American agriculture from the 1940ies he shows how such a seemingly mundane technological device led to a transformation of not only the U.S. agricultural system but also the tomatoes themselves. In order to "survive" the new machines they needed to grow hardier, sturdier and less tasty. Jack Stilgoe then asks for the ways in which AI is political and talks about some of his concerns and "worries": he is propelled by extraordinary hype and investment; he wonders about honesty in regards to purposes and processes, about the provenance of the data used by algorithms and about if and how questions of responsibility and inequality are being addressed; and he is concerned about the "myth of autonomy". Stilgoe argues that fundamental questions are either explicitly

rejected as inconvenient, or even ignored. He broadens the usual discussion on AI ethics by focusing on questions of politics, and calls for increased collaboration and “social learning” rather than relying solely on the dominant model of disruptive innovation.

Tim O’Brien of Microsoft takes us through various applications of facial recognition software, their possibilities and potential harmful uses. He uses facial recognition as a case study on how new approaches are required to realize the benefits of this powerful new technology while also putting in place safeguards to protect people and society from its outcomes. The basic functionalities of this technology can be grouped into face detection, face verification and face identification. O’Brien mentions flawed functionality, questionable use (e.g. without consent) and bias amplification (overuse on certain populations) as central issues in the use of facial recognition technologies. Microsoft therefore recommends regulating the use of facial recognition and advocates for a set of principles guiding both their development and deployment.

Data journalist and NYU Assistant Professor **Meredith Broussard** critically engages with our collective enthusiasm for new (computer) technology. She looks at the inner workings and outer limits of technology, and explains why we should never assume that computers always get things right. She challenges what she calls “techno-chauvinism”, i.e. the idea all our problems can be solved through technology, and argues that it is crucial to distinguish between the imaginaries and realities of AI technologies while paying close attention to how and by whom AI systems actually get designed. Our current ideas about technology come from a very small and homogenous group of people: well-educated white males. This is not a problem per se, but since unconscious biases get built into technologies we will re-produce the unconscious biases of this very particular group in the development of AI systems. From her perspective it is thus crucial to make this group more diverse and find more inclusive ways to create our collective futures.

In the final talk Paula Boddington of Cardiff University directs the attention of the audience to another central yet often overlooked issue in AI and related technologies: how can we manage to develop technologies that genuinely enhance our agency and our humanity rather than stifle it? While new technologies promise to enhance our agency and to deliver new means of expression and action, search engines for example tend to act in our place and make ‘judgements’ about what we seemingly need to be shown. Another way in AI technologies might stifle our agency is through subtly shifting understandings of human and machine intelligence: it’s quite easy to slip into judging humans by the standards of technology. The allure of new technology then is its promise to be ‘better’ than humans. Summing up her talk Boddington again stresses that liberty of thought and expression is a central human value, which opens up a number of key questions that need to be critically assessed in order not to let machines ‘think’ for us: “Open discussion and debate is not merely a means to *producing* ‘ethical’ AI. *It is part and parcel of what is needed alongside AI: forever.*”

After the talks **Ulrike Felt** from the University of Vienna gives a brief wrap-up, which also provides a transition into the panel discussion. The discussion after the talks takes up a number of talking points and circles around the following **main issues**:

- AI has evolved as an important sector of innovation, which makes it essential to do both assessing the achievements in an appreciative manner and continuing to ask critical questions.
- Whatever technologies we develop, they always contain and express the values (and unconscious biases) of those who develop them. Therefore, it is crucial to be attentive to biases but also as inclusive as possible along the technological development and implementation processes.
- Responsibility is also at stake when futures get imagined and promises are made. They frame innovation in many ways and therefore shape our lives.
- As AI will shape our lives in profound ways, we need to carefully consider who receives voice and who benefits from its development. Diversity, expressed as gender, race and age as well as in different socio-economic backgrounds, is key in a more inclusive development of AI technologies.
- Societal participation in AI development needs to happen through proactive engagement with the public. Allowing citizens to express concerns in online consultations will not be sufficient for receiving diverse and valuable inputs.
- Trust is critical in being more inclusive and responsive to manifold concerns, while being aware that it is always conditional. It would be a mistake to think trust can be restored and maintained by achieving certainty, by technological 'solutions' to ethical questions in AI, or by producing ethical codes and regulations.
- Regulation and ethical principles for AI are very important; yet, we need to think about issues that are often ignored or explicitly reject as inconvenient. These processes of questioning need to accompany the whole process of innovation, it needs to become a process of care and social learning.
- Related to questions of regulation and ethics, we need to critically reflect on accountability and how it is distributed.
- Finally, we should ask what acting in a world where AI plays an important role means and whether AI technologies might enhance or stifle our agency.