Report on Working Group 2 "Ethics and Artificial Intelligence"

Prof. Dr. Petra Ahrweiler

"Digital Transformation in France and Germany:
Consequences for industry, society & higher education",
25th October 2019 at Télécom Paris Tech.

The Working Group (WG) consisting of about twenty international experts from the field was chaired by Isabelle Falque-Pierrotin (morning session) and Claude Kirchner (afternoon session). In her introduction, Isabelle located the political background for WG2 in the Treaty of Aachen and in the national policy strategies of both countries, France and Germany, where AI, and especially the Ethics in AI, enjoys high priority with a need for concrete propositions and solutions. The Franco-German cooperation targets a joint research network on Artificial Intelligence, a common AI ecosystem focusing on shared ethical framing.

Accordingly, the chair asked the WG at the beginning of the sessions to consider four important issues as tasks for reflection:

1. Develop concrete propositions due to high priorities
2. Look for arguments supporting a European Social Model "Ethics in AI"
3. Develop a template for Ethics in AI to make it measurable
4. Consider storytelling/narratives/"heroes" of Ethics in AI for advertising the framework

The morning session was opened by a lecture of Katharina Morik, Chair of Artificial Intelligence, Technical University Dortmund in Germany, about privacy, fairness and comprehensibility of AI systems. After a short introduction to Machine Learning (ML), Katharina pointed out in the first part of her talk that France had identified ML much earlier than Germany as an important topic and had started to fund this type of research. German scientists had hooked up on French funded research as cooperation partners, because funding for Machine Learning only started recently in Germany. Therefore, she argued, a long track record of cooperation exists between France and Germany in this area. Based on this opening, she introduced the two current network initiatives on ML in both countries.
In the second part of her talk, the lecturer developed her perspective on ethics: since complexities of ML disallow public understanding of explainable algorithms, the task for the AI researcher community is to give guarantees based on theoretical insights that algorithms are valid. Theoretical insights require much more basic research. This is the contribution to ethics the AI community can provide. She ended with a plea for further research.

The next lecture was delivered by Philipp Slusallek, scientific director of the Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI) and member of the High Level Expert Group on Artificial Intelligence. His topic was Trustworthy AI: bringing together AI Technology and Ethics. Philipp began with a short introduction of his initiative CLAIRE (Confederation of Laboratories for AI Research in Europe) and its contribution to Ethics in AI. From there, he developed his perspective on ethics: Starting from the deadly accident with an autonomous car in the US, where the person crossing the street was not recognised due to bad training data and killed, he identified the problem. The world is extremely complex: Where to get the models and the training data from to improve algorithms?

He pointed out that the only option for industry to get appropriate training data is by exploiting synthetic data and by using simulations. However, he stressed that industry uses game agents that are created by artists and that do not display realistic behaviour of real people. This is a clear deficiency, because behaviour is key, in his opinion.
Philipp put forward a big request to build real-world models simulating the environment in which algorithms are applied (example: mobility contexts in cities for applications such as autonomous cars). With this, it would be possible to simulate realistic scenarios and create data that are appropriate. These data can be used for validating AI algorithms. The challenge is to create "better models of the world" (such as about the behavior of pedestrians). This way, he argued, we can have more ethical AI. Behavioural models should be created by interdisciplinary approaches from sociology, psychology, philosophy, law, and other social science disciplines for understanding people and what they are doing. To bring this on the way, and as concrete next step, the European Social Simulation Association ESSA with its head journal Journal of Artificial Societies and Social Simulation was mentioned: the Social Simulation community has a long track record in this type of behavioural modelling. A close cooperation of European AI researchers with ESSA researchers was strongly recommended.

The third and last lecture of the morning session was provided by Julien Carme, Lead Data Scientist at Worldline by Atos, who talked about ethics of AI in the economy. Julien developed a company perspective within a practical, business-oriented approach. He started with introducing the Ethics-by-Design framework implemented at his company and illustrated its application using the example of some B2B projects, in which ethics in AI became relevant. He went further into one specific application about providing flexible car insurance based on driving behaviour (AI estimating the risk for making an accident from data; insurance based on "pay how you drive"). He explained the ethical issues involved:

- Privacy protection becomes a problem
- Transparency will become a problem (why has my insurance rate increased?)
- Road fairness (no solidarity, no shared risk anymore, behavioural change due to discrimination by design – rich people get good insurance and do not care how they drive; poor people get the bad insurance and have to be ethical)

The discussion following Julien's talk was vivid. It centered around the following questions:
What can the role of industry be in the European Social Model on Ethics in AI?
Is there a friction between the European Social Model and individual, i.e. industry interest?
Does the European Social Model harm European competitiveness?
Is there a good balance between optimisation for (certain) individuals and empowerment of the collective approach? What role does industry play in this? What role does AI play in this?

In the afternoon session, the four lead tasks outlined in the morning by Isabelle were taken up by moderator Claude for further elaboration and operationalisation. It became clear that the first two had been addressed more extensively than the latter. It was decided to concentrate further on the first two questions to arrive at some concrete propositions.

Philipp Slusallek started discussions by explaining more about CLAIRE to suggest this institution as a hub for moving the area forward. His claim was that the field needs something like a non-military/civil Boston Dynamics in Europe, a large infrastructure, where, for example, high-level robotics can be researched within a brand for European AI.

The question then arose how AI, and especially Ethics in AI, could more strongly feature in European education, i.e. higher education. Katharina Morik informed about Annex 1 of the Collaboration Agreement between France and Germany, where joint education programmes in AI are described. However, the assembly of experts agreed that there is much left to do on the ground-level of European universities for establishing this topic firmly in academic curricula. As concrete suggestion, the activities of the EU project HEIRRI (Higher Education Institutions and Responsible Research and Innovation) were mentioned, which could be used to develop a systematic approach with good coverage.

Another question intensely discussed was how public engagement and multi-stakeholder debates on ethics in AI could be organised - both, on the German and the French level respectively, and on the level of the two countries involved in unison. Perceptions, attitudes, discussions and acceptance of AI use vary between countries, as do the types and degrees of AI implementation, with reference to norms and values in-use, but also related to technology status, economic models, civil society sentiments, and legislative, executive and judicial characteristics. To understand and shape the role of AI for future societies, therefore, needs a participatory approach involving many relevant stakeholders, which includes research methods to compare empirical cases, to model future societal scenarios on detail level, and to create better, i.e. more responsible AI technology adapted to context-specific social value requirements. A meaningful interdisciplinary cooperation between the technical and the social sciences should therefore join forces with society within a transdisciplinary approach to build better, i.e. context-sensitive, ethically-acceptable, and socially-informed AI for future societies. As a concrete step, it was recommended to connect to participatory approaches in social sciences technology assessment within relevant funding schemes (e.g. research project AI FORA funded in the programme “Artificial Intelligence and the Society of the Future” of the German Volkswagen Foundation). Further funding schemes in this direction should be set on the policy agendas of both countries, maybe as a joint initiative and as a specific contribution to the European Social Model for Ethics in AI.

This type of recommendation also applied to the following discussion about certificates for ethically-responsive AI. The recommendation here was that Germany and France should go in the same direction here and coordinate activities. Concrete next steps were anticipated.

All in all, the assembly felt that a strategic research agenda for Europe is urgently needed, which would center around the German-French cooperation agenda. This agenda should focus on a democratic version of AI based on the European Social Model, where data ownership and data benefits are (back) in the hands of citizens - as opposed to central political control applications (cue: China), and as opposed to industrial surveillance capitalism applications (cue: Facebook). Huge interest was expressed in intensifying the French-German partnership here – not only in doing cooperative research, but also in research agenda-setting and in co-organising appropriate research funding schemes.