Garance Perrot
Elis Indebetou
Xiaochuan Ai

Seminar 2 Session 2 Group 6

Seminar 2: Responsibility for automated decisions

The three texts lead to some thoughts. The first thought is about Data Privacy and Responsibility. We believe that developers should take a big part of the responsibility to protect user data. Especially in today's digital age, data breaches are common. It's crucial for developers to follow laws like GDPR (General Data Protection Regulation). They need to be clear about how they collect and use data, informing users about their practices. Furthermore, they should use strong security measures to protect this data. If a breach happens, both the developers and their companies should be held responsible and provide help to those affected. Ultimately, ensuring data privacy is not just a legal duty; it's an important part of building trust with users.

The second thought is about accountability. Accountability is crucial for any system being developed and used. On the one hand, the company producing the system may be incentivized to offer a high-quality product in order to maintain its reputation. On the other hand, from a customer's point of view, it is necessary to have someone to blame to ensure just punishment or compensation for victims when harm occurs. With machine learning becoming more common, there are tricky challenges about who is responsible. When algorithm decisions negatively affect society such as physical harm, property damage, or data privacy issues, it's really important to know who is accountable. Right now, our legal system hasn't fully figured this out. For example, if a self-driving car crashes, it's hard to decide who is at fault. Should we blame the manufacturer, the software developer, or the user? The self-learning aspect of machine learning makes the algorithms extremely complex and opaque so it becomes difficult to explain the process leading to one specific decision, even for the developers who created it. So, we think we need clear rules about how to assign responsibility in machine learning decisions, especially when things go wrong. Plus, the increased use of such systems in critical fields such as medical care or transportation urges attention to the issue of accountability. In the case of self-driving cars, we should improve laws that explain the roles of passengers, manufacturers, and developers. We could also create a system to determine accident responsibility, with an independent agency to look into self-driving accidents and figure out who is to blame. Plus, we can require developers and manufacturers of machine learning systems to get liability insurance. This way, they can provide financial help and compensation when accidents occur.

Another thought is about generative AI. When we use these tools to replace or help with human decision-making, could it weaken our thinking skills? On one hand, as generative AI becomes more popular, people are relying more on technology for everyday decisions. For example, students might use AI tools to complete homework or research papers instead of looking up information and thinking for themselves. From the user's perspective, this convenience makes things easier and more efficient, but it can also hurt our ability to think independently. We might become so dependent on AI that we stop actively thinking or analyzing problems. Conversely, some people believe that the inherent human drive for innovation will ensure that we continue to pursue ambitious goals, and that technological advancements such as AI will serve to augment our efforts. Another common example is the social media algorithms we use most often. Their recommendation algorithms push relevant content that reinforces existing views based on the user's preferences and viewpoint bias, resulting in the user being exposed only to information that matches his or her own position, and this filtering of information not only limits the diversity of our viewpoints, but can also lead to a one-sided understanding of social issues, which can in turn weaken critical thinking skills.

We also mentioned a point during our discussion about the risk of the “Clever Hans” phenomenon where machine learning models relied on unintended features (e.g background artifacts) rather than solving the problem as expected. Such algorithms can be thought as accurate, but it is important to confront them to different data sets than the training ones. One example is the image classifier presented during the lecture, that exploited irrelevant text information rather than the main artifacts, leading to completely false behavior when challenged on images without this text. Thus, transparency and careful validation is necessary to build ML models.

We also discussed a bit about the third text and that being able to know “how/what” a ML model made its decision, since this could help prevent against unwanted biases, make sure that the model is working as intended, and also give insights to us humans on how we could be able to make better decisions.