Let's Talk AI with Andreas Sesing-Wagenpfeil

Andreas Sesing-Wagenpfeil[1] and Barbara Steffen[2]

[1] Saarland University, Institute of Legal Informatics, Germany,
andreas.sesing@uni-saarland.de
[2] METAFrame Technologies GmbH,
barbara.steffen@metaframe.de

"The greatest challenge of AI is, from my point of view, that humans might feel like they are losing control (or they even lose control) because of uncertainty raised by autonomy."

The Interviewee - Andreas Sesing-Wagenpfeil

My Personal AI Mission:
As a lawyer, I'm willing to apply the law with regard to technical insights into and limitations of AI systems. We need to find the sweet spot between over-regulating the technology (coming along with chilling effects) and under-regulation (bearing the risk of sacrificing our common societal values) which is difficult, but possible.

My Takes on AI

Artificial Intelligence: In my opinion a buzzword which is used for complex artefacts with specific abilities. Those artefacts – representing AI – are capable of learning, adapting to perceived situations, solving (more or less complex) tasks.

Trust: Relying on something that we do not have perfect information about.

Explainability: A property of something that happens (which is, in case of actions, close to justification or reasoning) or of an entity (which is close to transparency).

Essential Elements of Human Capabilities: Argumentation by means of spoken and written, but not formalised language is a very important one.

The Interview

Barbara *Today, I have the pleasure to interview Dr. Andreas Sesing-Wagenpfeil. Could you please briefly introduce yourself and your relationship to artificial intelligence?*

Andreas Of course, thank you for the invitation. My name is Andreas Sesing-Wagenpfeil. As you mentioned, I'm a lawyer. I'm associated with both Saarland University in Saarbrücken and Helmut Schmidt University in Hamburg, where I currently hold a deputy assistant professorship for a year. I'm attending the AISoLA conference because I'm part of the interdisciplinary project "Explainable Intelligent Systems" (EIS). It's a project funded by the Volkswagen Foundation, which we run at Saarland University together with other universities.

Barbara *Can you name one or two specific challenges you're addressing with your AI research?*

Andreas Sure. I think the main challenge from a legal perspective is that we have a lot of technical approaches for building explainability models, and creating explanations on a system level, such as how a system works, as well as how it arrived at a certain decision or recommendation [1]. This is what explainability approaches usually do. From a legal perspective, we are required to apply the law. Applying the law means that there is a specific legal provision with certain requirements, and there is some set of given facts. These given facts are becoming more complex when dealing with AI systems which we don't fully understand, and which even computer scientists don't completely comprehend. The core challenge in our project focuses on how the questions raised by lawyers can be answered or at least supported by existing explainability approaches. We need, at least, a 'match-making' between relevant legal terms or concepts on the one side and technical approaches to 'open' the black box on the other.

Barbara *Based on your expertise and experience, what role does trust play in the adoption of artificial intelligence applications?*

Andreas Trust is a complex concept, as I've learned at this conference and in previous discussions with the colleagues from our project. The law addresses the concept of trust in different ways. Two years ago, when I was not yet working on the project, I would have agreed that maximising trust is often valuable. In fact, this is of course true if you take into account 'justified' trust: Then you know that people using a system, won't be harmed and the system will not discriminate against people. What I learned, which was very insightful from a psychological perspective, is that there's such a thing as overtrust and undertrust, and psychologists often discuss the calibration of trust. This is a new perspective. I think we should adopt this concept of calibrating trust also in law because it's probably not the best idea to seek as much trust as possible, but to build trust where it makes sense and to avoid building trust, which hasn't been a significant field of law so far, in systems that don't work. And, of course, wherever law

requires a human in the loop, overtrust jeopardizes this concept. This is, I think, a significant challenge that we need to tackle together with other disciplines.

Barbara *So basically, making sure that people or users remain sufficiently skeptical where they should remain skeptical.*

Andreas Yes, of course. We had some talks on the upcoming European Union's Artificial Intelligence Act [5], which contains the so-called principle of human oversight. There's the principle of human-centered use of AI. A manufacturer of a system has to provide the user with information so that they can effectively oversee a system [4]. But effective human oversight is very complex [3]. I think effective oversight entails both building trust so that we can take advantage of the efficiency provided by an AI system and maintaining the idea that people who use AI systems have to be critical and alert. Users must be aware of potential risks and limitations of a system so that they can understand what a system does. Of course, this kind of trust must be limited. Otherwise, you have this kind of automation bias - which isn't a traditional legal term. But together with psychologists, I learned a lot about that and understood better how to connect it to the legal system, for instance in non-discrimination law. This is a good example of where we should not build as much trust as possible, just to enable users to have effective human oversight and to not just go along with the system's recommendation.

"From a legal perspective, we are required to apply the law. [It becomes] more complex when dealing with AI systems which we don't fully understand, and which even computer scientists don't completely comprehend."

Barbara *When it comes to the ethical adoption of AI, do you think there are any essential measures we should take?*

Andreas Indeed. But I think the challenging question is, how can we derive and incorporate ethical standards into law? Lawyers aren't competent for in-depth ethical research, but we can use the opinions of ethical boards [2] and give advice to regulators. The challenging question is, how do we incorporate ethical standards, moral standards, into law? How do we derive legal norms that meet the requirements of AI systems and how do we put this into place? It's always difficult. It would be easy to just copy and paste moral principles for ethical use of AI into law, but ethical positions are often vague, they are not shaped for fitting legal systems. So I think it's a very specific task, which is not very easy, to derive legal provisions that can be enforced from moral standards and ethical guidelines to basic legal principles that we have to develop in the next few years.

Barbara *When you think about AI and its potential technical capabilities in the future, on a scale of 1 to 10, where 1 represents artificial intelligence systems like ChatGPT and 10 represents an artificial general intelligence that surpasses human capabilities. What do you think will be possible in the future?*

Andreas I can't predict that. But as it stands today, I think we're not at the minimum level. We're not around 1, because we already have some systems capable of performing very unspecific tasks, which we call general-purpose AI systems that we can deploy in numerous scenarios. Thus, 1 isn't the right answer. But I would also say 10 isn't the right answer either. If I had to guess, I would go for (maybe) 7. We will undoubtedly have AI systems that are very powerful. But I think if we're talking about surpassing humans, it's not about being better than a specific individual, but better than the entire group of humans populating our world. And I guess that if you take the crowd of people, AI will not surpass the capabilities of humans as a species or as a group of people – at least as long as it is depending on human input and based on human developments.

Barbara *In addition to the capabilities of AI, many different future scenarios get discussed, ranging from dystopia to utopia. What is your position?*

Andreas As a lawyer, you might expect me to lean towards a dystopian view. But I personally prefer the utopian view, primarily because of history. We've had several industrial revolutions. AI is not an industrial revolution; it's not limited to industry. But it is, from my perspective, a revolution in how we use technology for a better way. I think there are many tasks that society has to deal with. And I think we also managed some revolutions before. The most significant revolution was the development of computers and deploying computers to the entire population and all businesses. Of course, there are always risks that you could use to paint dystopian pictures. But we have to make people aware of the risks. Let's take the example of online banking. Online banking makes our life much easier (and it's not rocket science anymore): we are used to it, no matter that it's (of course) much easier for attackers to get hands on our savings due to technology [6]. Of course, there are risks that come with any technology because you always have people who misuse it. I think we have to find a balance. That's the main challenge. We have to find a balance to maintain our utopian way while dealing with the risks that come along with AI systems.

"We need, at least, a 'match-making' between relevant legal terms or concepts on the one side and technical approaches to 'open' the black box on the other."

Barbara *Reflecting on the last couple of days where we had a variety of presentations from different disciplines. Was there a presentation or insight that you found particularly interesting?*

Andreas There was a talk that dealt with the question of whether explanations given by autonomous systems influence a user who gets the advice of the system, is willing to overrule the system, or stick to the recommendation of the system. The approach was to give different explanations, one category was not gender-fair, and the other tried to give more technical explanations [7]. The quality or the perceived quality of the explanation had a significant influence on the behavior of the user, in terms of whether to override the system or trust the system's

recommendation. However, the behavior was influenced, but the actual outcome, if you just compare it to the baseline and the truth, did not really change. The explanation changed the behavior in monitoring the system, but it did not really make the decisions better or more accurate. There were also false positives or false negatives, where people did not trust a correct recommendation because of a weak explanation. This highlights the importance of good explanations: If the user is not comfortable with a given explanation, he is likely willing to overrule decisions and recommendations which are correct, but poorly explained.

Barbara *Is there a specific research question that you would like to see addressed from a multidisciplinary perspective?*

Andreas I think the question that I mentioned before is, how can we make use of technology which enables people to analyze systems, to understand systems, to gain explanations, or even to certify that a system is working properly? This is a very interesting thing, where law especially plays a role, of course, but where law is completely lost without the input from other disciplines. As lawyers, we ask for a guarantee for the functioning of a system so that we can say: The use of the system must be permitted or should be allowed. These questions must be answered by lawyers, but they cannot do so without any advice from psychology or computer science.

Barbara *From your personal perspective, what should the AI vision be?*

Andreas My ideal AI vision is that we use AI in the future for good. By good, I mean that we can tackle problems that we have in society, like the aging of our population, and decreasing birth rates, and just make use of AI so that the societal, moral, legal, and economic standards that we have achieved until today can be preserved. So that we can say, AI enables us to preserve our life or even make things better. Not only for us personally, but also for our children. They will need this planet. And I think AI will enable us to make the world better if we, on the same side, just limit the risks and are aware of the risks. This is my wish and my vision: that we use AI for just preserving our lives as we know them and improve things that we don't like today (and won't like in future).

"It would be easy to just copy and paste moral principles for ethical use of AI into law, but ethical positions are often vague, they are not shaped for fitting legal systems."

Barbara *I noticed over the last few days that people, like the other researchers, were particularly interested in the legal perspectives on AI. Is this a trend you often see in your research or at other conferences? Or did it surprise you?*

Andreas No, it did not surprise me. We know that from other conferences. There are buzzwords that everyone hears and reads in the media and gets from various sources and people talking about AI. There are a lot of things where we must clarify some distinctions that we are used to and are very familiar with. I gave the example of the distinction of liability, responsibility, and accountability.

Everyone uses these buzzwords, but the meaning differs across different disciplines. 'Privacy' is another famous example which stands for a whole bunch of concepts. Another aspect is that one: Lawyers are always asked which rules must be followed to build a system which is safe, and therefore compliant with legal obligations. It comes to a clash when we answer: "Sorry, there is no specific rule addressing your problem – but if you take reasonable care, you will – at least – not be liable." What I want to say is: The law often provides for general rules written in unspecific terminology because law cannot elaborate on every single use case in detail [8]. But lawyers are always open for a joint development of standards and technical norms together with technical experts which is a win-win situation. The engineer has a guidance that he must follow, and lawyers have a checklist that helps to identify a faulty product.

Barbara *Do you think it will be possible in the future to make the legal framework or the law more accessible to other disciplines? I have the impression that for non-experts, the law seems quite complex, which explains why they prefer to keep their distance and talk to experts who can explain it to them.*

Andreas Of course, we should not do that because otherwise we would lose our jobs as lawyers if everyone could just apply their law by themselves (kidding). The real answer is that making laws is very complex. Making laws that are applicable to anything that comes along is a very complex field where you have to put much effort in. For instance, if you're just talking about the wording of a norm, it makes a difference if you say a norm is applicable if x, or if you say the norm is applicable "except for not x". From a logical and maybe from a computer scientist's perspective, there is no difference – both rules seem to be the same. But from a legal perspective, it makes a (huge!) difference, because the wording can lead to a reversal of burden of proof for this exemption [9]. And this is something that we cannot abolish, you need lawyers who have this knowledge. I think that it's important to stick to those basic legal principles which have developed over centuries. Therefore, it's always hard to find the right way of not being too abstract that no one can understand what the law says. And on the other hand, it is necessary to keep the law flexible for technical developments.

"As lawyers, we ask for a guarantee for the functioning of a system so that we can say: The use of the system must be permitted or should be allowed. These questions must be answered by lawyers, but they cannot do so without any advice from psychology or computer science."

Barbara *Is there anything else you would like to add?*

Andreas I thought about this question, and I did not really have an idea.

Barbara *Then thank you very much, Andreas, for your time and your legal perspective on AI. Have a great day!*

Andreas It was a pleasure. Thank you very much for the invitation.

References

1. On the scope and the extent of transparency in automated decision-making under the General Data Protection Regulation (GDPR) Sesing, MMR2021, pp. 288 et seq.
2. High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI, 2019, available at https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1.
3. Sterz/Baum/Biewer et al., On the Quest for Effectiveness in Human Oversight: Interdisciplinary Perspectives, preprint available at https://arxiv.org/abs/2404.04059.
4. Article14 of the draft proposal of the AI Act.
5. See esp. the initial Proposal for a Regulation of the European Parliament and of the Council laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative Acts, 21.4.2021, COM(2021) 206 final, available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex\%3A52021PC0206.
6. The German Federal Court of Justice (Bundesgerichtshof; BGH) has dealt multiple times with the civil liability for losses caused by phishing attacks, see BGH, 24.2.2012, case no. XI ZR 96/11; BGH, 26.1.2016, case no. XI ZR 91/14.
7. See on the discussion on a ‚right to explanation' under the General Data Protection Regulation (GDPR) Temme, EDPL 2017, 473; Wachter/Mittelstadt/Floridi, IDPL 2017, 76 et seq.
8. See, for instance, on the relevance of technical norms and standards under the German Product Liability Regime Borges, in: Borges/Hilber, Beck'scher Online-Kommentar IT-Recht, 13.Edition (1.4.2024), Sec.3 ProdHaftG, mn.44 et seq.
9. For the principles of the attribution of burden of proof under German civil laws, see Sesing-Wagenpfeil, in: Hoeren/Sieber/Holznagel (eds.), Handbuch MultiMedia Recht, 60th ed., sec. 18.5 mn. 5 et seq.