Who is legally liable for AI?

Binary Dreams · Follow
11 min read · Mar 15

Artificial intelligence raises many legal questions that are likely to result in new regulations and laws in Switzerland and the EU in the future. Isabelle Wildhaber, law professor at the University of St. Gallen, clarifies the current situation and explains what this means for businesses.

OLIVER BOSSE SPOKE WITH ISABELLE WILDHABER

Mrs. Wildhaber, first of all, a very basic question: Why is it important, especially at this point in time, to deal with legal issues relating to artificial intelligence?

The US government agency OSTP is proposing a Bill of Rights for the AI era, and the EU Commission is making proposals for European AI regulations. It is therefore important to discuss which new regulations are necessary and sensible with regard

to AI and how these should be designed. On the one hand, we as a society want to exclude or at least reduce risks as far as possible; on the other hand, we do not want to nip innovation in companies in the bud. Since new AI products and AI services are springing up like mushrooms every day, now is undoubtedly the right time to take a critical look at AI regulations.

Can you give an example of a legal issue that arises as a result of AI?
One relevant and serious issue is algorithmic discrimination. The hype around the potential of AI leads many people and companies to believe that data always reflects the objective truth. However, this is not true at all. On the contrary, an unreflective belief in data is dangerous. Data is collected and interpreted by people — and this can result in discrimination. Even the definition of what is desired, e.g. what is meant by a good work performance, is determined by humans.

There are many causes of algorithmic discrimination — the data might be poorly selected, incorrect, or outdated. For example, if Google's face recognition algorithm incorrectly labeled black people as gorillas, it was due to insufficient training with black faces. Discrimination often enters the algorithmic process unintentionally, as the product of unconscious human bias or as simple error.

> *An insurance company's AI could, for example, screen out overweight people and thus discriminate against them. Legally, however, overweight people are not a protected group.*

Can't such discrimination be prevented with anti-discrimination laws?
From a legal point of view, it is unfortunately very difficult to capture algorithmic discrimination. There are several reasons for this. Discrimination is difficult to prove and there are hardly any deterrent sanctions against discriminatory behavior. Furthermore, most jurisdictions only selectively protect certain group memberships, e.g., individuals are protected on the basis of gender or disability. However, AI identifies and categorizes individuals based on entirely different criteria. For example, overweight individuals might be screened out by an insurance company. But overweight people are not part of a group that is protected by law.

And why is the issue of liability on everyone's mind?
In terms of liability, we need to ask whether AI creates certain risks that cannot be adequately addressed by our current liability rules. This question arises with many new technologies, as it did with biotechnology in the 1990s, for example, or with

nuclear energy in the 1970s. In each case, there is great concern about loss of control and unforeseeable, unattributable damage.

Regarding AI liability, the starting point is the discussion of the typical risks of AI. In particular, the three risks of black box, complexity, and autonomy of AI systems are cause for concern. This is because they can cause problems in the enforcement of liability claims.

Is the "black box" the main problem in terms of liability?
Yes, in terms of liability, the black box risk is a big problem. It can arise from the fact that AI systems improve their performance by learning from experience. Only input and output are known, while the model remains opaque. This can cause the AI to exhibit a lack of transparency.

The black box effect causes limited explainability and traceability of decision making by self-learning systems. Humans have difficulty understanding exactly how inputs lead to outputs in the context of AI systems. Why can't you get a loan even though you meet all the requirements? Aggrieved parties may not be able to identify what exactly caused the damage. However, transparency and traceability of causal processes are necessary for proving malfunctions in AI systems and for enforcing resulting liability claims.

Would establishing explainable artificial intelligence (XAI) be enough to address the "black box" problem?
The machine learning community has been working for some time to address the problem of algorithmic decisions lacking comprehensibility, e.g. in the context of the U.S. Defense Advanced Research Projects Agency's (DARPA) project on Explainable AI, launched in 2016. Entitled XAI, the project discusses how opaque machine learning models can be made explainable. The user should be given concrete explanations of individual decision steps, the general strengths and weaknesses of the model, and an understanding of how the system will behave in the future. But not every form of explanation is technically feasible or purposeful.

To reduce the problems of non-transparent AI for liability, science and policy makers have also been calling for a few years to make algorithms explainable. Transparency requirements are being introduced for this purpose, e.g., in the European General Data Protection Regulation. Transparent AI systems could reduce problems for plaintiffs in liability cases. However, it is not obvious that

transparency can solve the problems that plaintiffs will face in civil cases. In some cases, it is only attainable with great difficulty and is very expensive. Therefore, it cannot simply solve the "black box" problem.

> *Transparent AI systems could reduce problems for plaintiffs in liability cases. However, in some cases, transparency is only attainable with great difficulty and is very expensive.*

And where exactly lies the need for action? What are the relevant problems under liability law?
In my view, there are serious problems in proving fault and causality. These problem areas lead to the fact that the liability for AI is discussed very broadly and the demand for a harmonized liability regulation for AI has become prominent on the European level.

For example, let's look at the problems around proving fault. Fault consists of a breach of duty of care in deploying, training, maintaining, or monitoring the operation of the AI system. Providing proof for such fault is a huge hurdle for injured parties as they usually know far too little about the AI.

In addition, most systems are only decision support systems, they do not replace human decision making. If humans still have to monitor the system, the interface between humans and machines has to work well. But what can humans expect from an AI system in the first place? Under what circumstances may operators rely on the AI system, and at what point should they override its suggestions? We have found in our empirical studies of AI systems in the workplace that operators would rarely override the algorithmic system.

> *Should Tesla be liable as the manufacturer, or should the driver be liable as they have not maintained the car properly?*

Various countries and also the EU (Artificial Intelligence Act) are currently examining the extent to which new regulations might be necessary in connection with AI. What is the current state of affairs in the EU in this regard?
Unlike in Switzerland, liability for AI has been discussed up and down in the EU for several years. The EU is basically aiming for a special, so-called horizontal liability regime (see box) for defined high-risk AI applications.

On September 28, 2022, the European Commission published its proposed directives on AI liability. There are two pillars of liability. First, there is producer liability. Here

the European Commission is making a proposal for a new product liability directive to replace the current Directive 85/374/EEC. And secondly, there is operator liability, i.e. the liability of users, owners or holders of AI. Here, the European Commission envisages a proposal for a directive to adapt the rules on non-contractual civil AI operator liability. It is important to allocate manufacturer and operator liability appropriately and delineate the respective risk spheres. Should Tesla be liable as a manufacturer, or should drivers be liable if they have not properly maintained the car? If operators are also liable, they have to deal with the AI used. I think it's good for people to take a critical look at the use of AI and question it. Operator liability could thus improve digital literacy.

And what is the current state of affairs in Switzerland?
At the end of 2019, a report by the interdepartmental expert group on AI was published in Switzerland. The report did not identify any need for regulatory action and considered the existing liability regime to be sufficient at present. This assessment thus differs fundamentally from that at the European level. It calls — where necessary — for a sector-specific supplement to the existing regulation, without a horizontal approach.

Since 2020, however, based on the European discussions on AI regulation, more voices have been raised in Switzerland discussing whether there is perhaps a need for regulatory action with regard to AI.

In a report to the Federal Council in April 2022, the Federal Department of Foreign Affairs (FDFA) stated that Switzerland could set its own priorities for dealing with AI in its national legislation. However, too great a difference between international and national legislation would not be in the interest of Switzerland's ability to connect to international markets and supply chains on AI. It will therefore be interesting to see how Switzerland positions itself. It will exercise regulatory restraint, as it is keen to maintain its top position in countless innovation rankings.

What specific solutions are being discussed in the EU — and which do you think have the greatest potential and consensus to actually be introduced one day?
The EU is basically aiming for a horizontal liability regime for high-risk AI. Switzerland is more in favor of a sector-specific approach. However, there does not seem to be much left to change about the horizontal EU approach.

If a horizontal approach is chosen, various options are conceivable: One could tighten fault-based liability or introduce strict liability for operators of a selected group of AI systems.

In the European Commission's proposals on AI liability of September 28, 2022, in Article 5, the idea of strict liability is postponed to the future. Only after a few years should it be evaluated whether such liability is appropriate and whether insurance coverage is necessary. Since I am critical of strict liability, I think this is good.

On the other hand, the European Commission wants to introduce two rebuttable presumptions for proving fault and causality. So the legal procedure will be simplified for victims in terms of proof. They do not have to explain in detail how the damage was caused by a particular fault or omission. I think these two presumptions are welcome. They do not interfere too much with national legal systems. They can help overcome the black box challenge of AI in liability. I think consensus could be reached that presumed fault and causation presumption liability should be introduced for operators, each with the possibility to rebut the presumption. How exactly the presumptions should be formulated can still be discussed.

> *AI systems cannot generally be regarded as particularly risk-prone or hazard-increasing. On the contrary, AI systems are used to compensate for human inadequacies and thus reduce risks and damage.*

Are you in line with the EU's horizontal regulatory approach?
A cardinal problem of a horizontal regulatory approach is the definition of the scope, i.e. the definition of AI on the one hand and the definition of risk categories on the other. This is precisely what proves to be extremely difficult. What exactly is AI in the first place and which AI is actually high-risk? There is also the fact that a horizontal regime, even if it tries to differentiate with the definition of high risk, it cannot sufficiently do so regarding the context in which an application is used. For example, does the product come into contact with non-specialized users? How exactly is it used and how does it evolve? Horizontal regulation can also inhibit innovation. This was the case, for example, with the Genetic Engineering Law in Switzerland for agriculture.

Are there also examples of legally completely unobjectionable areas of application?

There are numerous AI systems that are not risky, e.g., AI for managing a computer's working memory, scheduling appointments, planning inventories and purchasing goods. AI systems cannot generally be considered particularly risky or hazard-increasing. On the contrary, AI systems are used to compensate for human inadequacies (delayed reaction times, overlooking essential data or information, and emotionality) and thus reduce risks and damages.

> *In our empirical studies on AI in the workplace, we have found that the introduction of AI tools in Swiss companies is sometimes not accompanied by legal advice at all.*

What would the current developments mean in terms of liability — who would realistically be held accountable?
It is and remains important that liability follows control over the liability-triggering properties of the thing in question. Accordingly, if the operator controls the "behavior" of the thing, he should be liable. If, on the other hand, the damage is based on the safety-related properties of the thing, the manufacturer should be liable. This is because they define and control these properties. The key difference between conventional products and AI products is that the latter tend to shift control over the "behavior" of the AI system to the manufacturer. It is the manufacturer that programs, trains, and distributes instructions to the AI systems. Therefore, in the future, the manufacturer will be the central figure of liability for AI systems.

However, the operator of the AI system will be the first visible point of contact for injured parties in numerous cases. An injured party will first turn to the operator — who will often also be their contractual partner — and not to the original manufacturer. An employee will therefore turn first to their employer. In the workplace, this should remain the case, as there are special procedural rules under labor law and special labor courts that are available to affected employees. Furthermore, I think it's right that operators need to address what kind of AI they are using, why and how.

What is your recommendation to companies that are already using AI tools today or would like to use such tools in the future — what legal questions should they consider?
In our empirical studies on AI in the workplace, we have found that the introduction of AI tools in Swiss companies is sometimes not accompanied by any legal advice at all. The tool is introduced in a so-called pilot and it is not mandatory that an in-

house or external lawyer has an eye on it. My recommendation would therefore be to have the legal issues — especially those of a data protection nature — clarified when introducing an AI tool.

> *I wouldn't wait to use AI tools, the wheels of progress turn too fast for that.*

As a company, waiting to deploy AI could put you at a competitive disadvantage from an economic perspective. Is there a legal argument for waiting?
No, I wouldn't wait, because the wheels of progress turn quickly. However, as I said, I would recommend that potential legal problem areas be clarified by a lawyer before introducing an AI tool.

From risks to opportunities, where do you see these in relation to artificial intelligence?
The rapid advances in AI offer countless opportunities for society. AI systems are being used in a wide variety of sectors, in industrial production (Industry 4.0), in the energy sector (Smart Grid), healthcare sector (MedTech), financial sector (FinTech), insurance sector (InsureTech) or legal sector (LegalTech). Also in public administration (e-government), in environmental and climate protection, in agriculture or in education. Many AI systems are here to stay and are full of opportunities! It is imperative that these be exploited, legal problems or no.

* * *

ABOUT ISABELLE WILDHABER
Prof. Dr. Isabelle Wildhaber, LL.M., has been a full professor of private and commercial law, with a special focus on labor law, at the University of St. Gallen since 2015, as well as managing director at the Labor Research Institute (FAA-HSG). She is the rector's delegate for equality, diversity and inclusion and president of the Equal Opportunity Commission of the University of St. Gallen. She researches, teaches and consults on topics of contract, liability and labor law, with reference to new technologies and developments.

Artificial Intelligence | Ai Regulation | Legal Issues | Ethical Ai | Liability