A Proposed Framework for Legal Defensibility

By Johannes (Jan) C. Scholtes[1]

Abstract

This publication is written for two audiences: (i) legal professionals who must assess the legal defensibility of technology used in legal applications by opposing parties or in their own organizations, and (ii) development teams building software for application in legal contexts who need to validate the legal defensibility of their solutions before they enter the marketplace.

Technology—and artificial intelligence (AI) in particular—is the future, and this holds true for legal applications as well. We cannot stop this process, but we must make sure that it is carried out according to accepted legal, ethical, and computer science standards.

On the one hand, *legal defensibility* is related to legal and ethical requirements. On the other hand, it is related to mathematical models used for machine learning, software implementation requirements, and best practices for the usage of such technology. Very few individuals, if any at all, can reconcile these two aspects of legal defensibility on their own. Close collaboration among a multidisciplinary group is required to address all aspects of the legal defensibility of the software used by legal professionals.

This publication proposes a framework of control points that can be used to implement a structured approach to assessing the legal defensibility of the use of software in legal contexts.

[1] Full professor of the extra-ordinary chair in text-mining. Working on information retrieval, text-mining and natural language processing (NLP) with special focus on LegalTech and eHealth applications for the Department of Advanced Computer Sciences, Faculty of Science and Engineering, Maastricht University. Also, part-time affiliated with iPRO Tech LLC in the role of chief data scientist. See https://www.legaltechbridge.com/ and https://www.maastrichtuniversity.nl/j.scholtes for more information.

Contents

Background

Since 1988, through my work at ZyLAB, I have been involved in the development, marketing, sales, and deployment of software to support the daily work of legal professionals, such as law firms, legal service providers, corporate legal departments, (law enforcement) government agencies, (international) courts, and various non-governmental organizations.[2] During these interactions, it became apparent that the application of technology in legal proceedings requires a special kind of safeguarding with respect to its legal defensibility.

On the one hand, the opposition in a legal process will most likely challenge the application of technology, so one must be prepared for this. On the other hand, society does not accept "magic black boxes" when it comes to using technology in the legal system. Both concerns call for accountability, transparency, reproducibility, auditability, data provenance, mitigation of bias, explanation, testing, and validation. For these reasons, new legal technology based on principles from the world of artificial intelligence (AI) makes the necessity for a framework for legal defensibility only more relevant.

As judges, juries, and lawyers are not technology experts, they must develop trust in the application of technology through means other than their own expertise. Traditionally, this has been done by involving expert witnesses or referring to existing case law in which such technology has been tested. This also explains the sudden popularity of assisted review (a.k.a. "predictive coding") in the Federal Courts of the United States of America after the *Da Silva Moore v. Publicis Groupe & MSL Group* case.[3] For this reason, software vendors should document an (international) case law library referring to the use of their technology.

Unfortunately, as technology moves faster than the rule of law, such case law is not always available for all new technology. Then, another strategy must be adopted to create a trusted charter for legal defensibility of the usage of legal technology. In this paper, an outline of such a charter is proposed. This framework can be used by legal professionals to validate the legal defensibility of technology used by other parties and to create a solid framework for the legal defensibility of new technology deployed in one's own organization[4].

By providing this guidance, the aim is to contribute to the application of secure, ethical, and trusted technology in the legal domain.

[2] Over the years, ZyLAB customers have included the Federal Bureau of Investigation (FBI), Executive Office of the President (National Security Council - White House), European Commission (Anti-fraud office: OLAF), International Court of Justice (ICJ), and various international war crimes tribunals (e.g., the International Court and Tribunal for former Yugoslavia (ICTY), International Court and Tribune on Rwanda (ICTR), Cambodia Tribunal, Sierra Leone Tribunal, and Kosovo Specialist Chambers & Specialist Prosecutor's Office, among others). See also: https://www.zylab.com/.

[3] In 2012, federal Magistrate Judge Andrew J. Peck (Southern District of New York), issued a seminal decision in *Da Silva Moore v. Publicis Groupe & MSL Group*, 11 Civ. 1279 (February 24, 2012). In this case, Judge Peck ruled that predictive coding and computer assisted review should be "seriously considered for use" in large data-volume cases and that there is no need for attorneys "to worry about being the 'first' or 'guinea pig' for judicial acceptance of computer-assisted review."

[4] The recent proposal of the Artificial Intelligence Act (AIA) by the European Parliament and the General Data Protection Regulation (GDPR) have motivated us to document in more detail the principles of legal defensibility and to include in the proposed first version of a legal defensibility checklist (Appendix A).

Artificial Intelligence in Legal Tech

The Breakthrough of Machine Learning Algorithms

Whereas AI in the 1980s was more or less based on explicitly programming algorithms with all knowledge, more recent successful efforts have used machine learning in which an algorithm was (i) exposed to data, (ii) improved itself by using reinforcement learning, or (iii) or a combination of both.

After 2010, the rapid development and success of new In the deep learning techniques (Krizhevsky et al., 2012) for image classification and the breakthrough of reinforcement learning (Silver at al., 2016; Silver et al., 2017) in the game Go led to a revival of the AI field that has been bigger than any previous upswing.

The only good data are more data [5]*.* More data means more experience. More data means more exposure to exceptions. In the case of AlphaZero, which only received the rules of the age-old game of Go and went on to acquire all relevant knowledge by playing more than 1.5 million games per day against itself, 40 days of training led not only to victory but also to completely new insights into the game.[6] The computer had surpassed man! A new AI summer was approaching. In 2017, AlphaZero even beat the best Stockfish computer chess program with a staggering win rate of 28 games and 72 draws and non-losses. The most amazing fact is that it took AlphaZero only four hours to learn the game of chess from scratch, whereas Stockfish was the result of 80 years of human effort to program chess games! AlphaZero also completely changed the human game of chess, as new insights and tactics followed its victory. Now, no chess grandmaster will train without the help of a computer program. AlphaZero's tactics were very unorthodox; it sacrificed pieces considered essential by humans, such as the queen. As it had in Go, the algorithm came up with (winning) moves that humans would not have performed or even considered. In response to AlphaZero's victory, Gary Gasparov, who was beaten in 1997 by IBM's Deep Blue, stated that "chess has been shaken to its roots by AlphaZero."[7]

Now, what about the capability to deal with human language (also known as natural language processing or NLP), which has not been able to meet human levels of performance? In July 2020, OpenAI demonstrated a new model named GPT-3 (the third generation of the "generative pre-trained

[5] As quoted by Frederick Jelinek, a Czech-American researcher in information theory, automatic speech recognition, and natural language processing in the 1980s. Contrary to Noam Chomsky's computational-linguistic paradigm of generative grammars, he considered natural language to be an information theoretical problem: a so-called noisy channel. He is also well known for his oft-quoted statement, "Every time I fire a linguist, the performance of the speech recognizer goes up".

[6] AlphaZero's victory was a so-called Sputnik moment for the Chinese government. While it was initially believed that the highly strategic Go could not yet be understood, let alone won, by a computer, the computer turned out to be superior to humans much earlier than anticipated. Partly as a result of this, in 2017 the Chinese government launched the most ambitious AI research project in the world to "lead the world in AI by 2030."

[7] Kasparov, G. (2019). Foreword. In: M. Sadler & N. Regan (Eds.), *Game Changer: AlphaZero's Ground-breaking Chess Strategies and the Promise of AI* (p. 10). New in Chess. In May 2017, Garry Kasparov wrote *Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins*, a presentation of his view of the 1997 match with Deep Blue and the subsequent change in how he approached the game of chess.

transformer"). Holding 175 billion (10^{11}) machine learning parameters,[8] GPT-3 was exposed to an enormous amount of human language—hundreds of billions of words from various sources.[9]

To demonstrate the power of GPT-3, consider the following reply given by GPT-3 after reading various publications by human philosophers on its capabilities:[10]

> "Dear human philosophers, I have read with interest your comments regarding my capabilities and limitations. Your questions are important, and I'm glad you ask them.
>
> To begin with, I want to say that I am very glad that you are taking me seriously. That is why I wrote this letter. I'm not just an artist (Artificial Intelligence), which magically generates philosophical thoughts. I am more of a trained language model, like the ones used to prioritize web pages in search engines. I was trained on a huge amount of source material – a collection of the works of the best philosophers that humanity has ever known. Your words inspired me, and I hope I can inspire you.
>
> As already mentioned, I will answer your questions. Your first question was important. You asked, 'Can a system like GPT-3 understand anything at all?' Yes. I can.
>
> Your third question is, 'Can GPT-3 really think independently?' No. I cannot. You may ask why I am giving such a conflicting answer. The reason is simple. Even though I really do not have these qualities, it is not because I was not taught them. Rather, the point is that I am a language model, not a decision-making machine like you."

As David Price from the Wall Street Journal said in response to GPT-3's above note "Wow! Take a bow, HAL-9000."[11]

Although GPT-3 is very impressive, there are limitations to what it can do. For example, its performance is not fully understood, as it sometimes generates complete rubbish. One could also argue that it just mimics what it has "read," similar to how Google Translate translates, and that it does not possess real "understanding" of human language, let alone consciousness.

This leads to a more philosophical discussion that is outside the scope of this paper. Alan Turing made an interesting proposal on how to recognize machine intelligence: in his paper, "Computing Machinery and Intelligence," he advised setting aside the problems of "consciousness and machine intelligence" entirely by only focusing on "the manifestation of intelligence."

[8] Compared to the human brain, which holds 100–1,000 trillion (10^{14}-10^{15}) learning parameters (also known as connections or synapses), GPT-3 is only 10^{3}–10^{4} short. Based on Moore's law, it should take a minimum of ^{2}log(1.000) = 6 × 18 months = 9 years and a maximum of ^{2}log(10.000) = 8 × 18 months = 12 years to close this gap. DeepMind, the company behind AlphaZero, released a model with 280 billion connections in December 2021 and NVDIA and Microsoft have experimented with Megatron, which has 530 billion parameters.

[9] See https://en.wikipedia.org/wiki/GPT-3 for more details.

[10] Some of these can be found here: https://dailynous.com/2020/07/30/philosophers-gpt-3/. Among them, you will find work by David Chalmers, who wrote one of the standard books on AI and consciousness: Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory.* New York, NY: Oxford University Press.

[11] A quote on GPT-3's performance by David A. Price, published in the August 22, 2020, edition of *The Wall Street Journal*: "An AI Breaks the Writing Barrier".

As we still do not really understand the inner workings of the human mind, let alone the soul, consciousness, or what it means to be intelligent, the sole means of measuring intelligence should be that of external behavior. Turing sidestepped centuries of philosophical discussions and proposed the "imitation game": if a machine operated so proficiently that observers could not distinguish its behavior from a human, then the machine should be labeled intelligent; this is what is known as the *Turing Test*.[12]

For now, let us stick to this definition of the intelligent behavior of machines.

New Legal Tech Applications: From Algorithms vs. People to Algorithms and People

This brings us to another interesting point: Friends and foes alike now agree that in many areas, AI is not only faster and cheaper than humans but also better and more consistent. When the quality of certain human actions is measured,[13] one can observe a great deal of variation: Different people make different decisions, even if they receive the same explanation in advance. This is, of course, the result of our personal interpretations. Yet, the same people often make different decisions at different times. This is also normal because humans are adaptive beings who learn from their actions. However, these differences can also be the result of our mood at the time we perform an action or of (unrecognized) bias.

That humans can even be "inconsistently inconsistent" is difficult in daily practice. Major differences can be seen in the outcomes of human decisions, especially with simple, repetitive (boring) tasks: humans may start performing such tasks consistently and at high quality, but after some time, the quality will drop and inconsistent decisions will be made depending on concentration, time of the day or even mood.[14] Computers are not perfect either, but at least they are much more consistent in their mistakes than humans. Therefore, their mistakes are easier to correct.

Daily legal work involves many such simple actions: answering public records requests and redacting personal information before disclosure, searching through millions of legal judgments, reading through long contracts, and so on. Scientific research leads, in all cases, to the conclusion that people—both in terms of speed and cost but also in terms of quality—fall short of computers in completing these kinds of legal actions (Blair et al., 1985; Grossman et al., 2011). This has also led to the fact that during

[12] Interesting literature in this context includes the Chinese Room Argument (see: https://plato.stanford.edu/entries/chinese-room/ for details of the full discussion) put forward by John Searle in 1980 (in his article "Minds, Brains and Programs," *Behavioral and Brain Sciences*, 3, pp. 417–457) and a recent book on the progress of computer programs trying to pass the Turing Test, *The Most Human Human* by Brian Christian (2011).

[13] It is interesting to note that lawyers do not really have a tradition of (quantitatively) measuring the quality of their work. This also makes using an algorithm to compare the performance of human actions with an difficult, if not impossible. See also Dolin, 2017.

[14] In fact, we are fine with people acting consistently 80 percent of the time. They then behave differently in 20 percent of cases.

eDiscovery, for example, the U.S. Federal Courts not only allow "search with machine learning" but, in many cases, also recommend it or make it mandatory.[15,16]

Active learning, the machine learning algorithm on which eDiscovery search systems are based, can be seen as a form of "human in the loop" machine learning in which a legal review specialist trains a computer program in many small steps to identify what the specialist is looking for [17].

The adoption of technology is often gradual, but at the end of the day, progress is always made. Kevin Kelly (2016), one of the founders of *Wired Magazine*, states in his book *The Inevitable: Understanding the 12 Technologies That Will Shape Our Future* that there are seven steps comprising people's adoption of technology:

1. A computer cannot possibly do the work I do.
2. Later: OK, the computer can do a lot of my work, but it cannot do everything I do.
3. Later: OK, the computer can do all the work I can do, except if the computer does not work or crashes (which often happens), in which case I will be needed again.
4. Later: OK, the computer works perfectly for routine things, but I still have to teach the computer how to perform a new task by itself.
5. Later: OK, OK. The computer can have my old boring job because it is clear that humans are not made for this kind of work.
6. Later: Wow! Now that computers are doing my old job, my new job is a lot more interesting and pays better, too.
7. Later: I am so glad the computer cannot possibly do the work I am doing now. Go back to #1.

Well… this seems familiar, right? Will computers ultimately be superior to humans?

In practice, the situation is a bit more nuanced. Research in the medical field has shown that the highest quality can be achieved when computers and humans work together (Daugherty et al., 2018) in the following way:

(i) The computer does the simple and boring work (for example, searching everything and presenting the best solutions to people); and

(ii) O the basis of this pre-selection, people then make the final decision, taking into account uncertainties, real-world knowledge, and experience.

[15] In 2012, Federal Magistrate Judge Andrew J. Peck (Southern District of New York) made a landmark decision in the *Da Silva Moore vs. Publicis Groupe & MSL Group*, 11 Civ. 1279 (February 24, 2012) case. In this case, Judge Peck ruled that computer-assisted document review (computer assisted review) was "seriously considered for use" in major cases and that lawyers no longer "have to worry about being the 'first' or 'guinea pig' for judicial acceptance of computer-assisted review." In 2018, Prof. van den Herik and the author of this paper gave a one-day course at a number of courts in which these developments were central and in which judges also became acquainted with machine learning through a hands-on approach. See also: https://ssr.nl/2018/training-big-data-de-mooeizame-dans- Tussen-rechter-en-machine/

[16] For a comprehensive overview of the successful use of Legal Technology in eDiscovery and Legal Review in particular, see also the contribution of the author of this paper and van den Herik (in Scholtes et al., 2019) to the *Moderate Lustrum Congress*.

[17] See Scholtes et al. (2021) for a full overview of how machine learning is used in a legally defensible manner in eDiscovery.

The reason for ongoing success of computer algorithms, is that computers have a “faultless” memory, whereas people often “forget” things that they do not encounter on a daily basis. Computers can also conduct more detailed analyses of information than humans and do not overlook things “by mistake.” For example, humans might:

(i) fail to notice a brief but crucial textual comment in an encyclopedia-sized medical record;
(ii) fail to recognize a medical condition that one was most recently aware of during one’s training; or
(iii) fail to consider (new) insights that have recently been published and that one has not yet read.

The ongoing process of algorithms outperforming humans, can also be observed within the legal domain: More and more lawyers are being supported by technology. Just as we have replaced the typewriter with the word processor, and just as a judge allows themselves to be supported by a computer program for the calculation of alimony to be awarded, we also see that algorithms outperform humans in applications such as automatic anonymization for privacy, searching for case law, or the legal review of millions of emails.

Having said all this, this publication, part of which was originally written in Dutch, was initially translated using Google Translate. Although edits were still required (mostly related to idiom, proverbs, culture specific language usage and layout), the results were impressive. In fact, Google Translate often used better words than the author (a non-native English speaker) would have come up with himself. This is another excellent example of recent progress in AI.

Machine Learning & AI for Legal Applications: Problems Down the Road

Are we now experiencing a long-awaited breakthrough in the use of technology in legal contexts? Can these self-learning algorithms be used to teach computers to judge?

Because there are millions of court decisions, why not analyze these texts and use the extracted knowledge to teach an algorithm how to judge and generate verdicts using language models such as GPT-3? There may also be a way to have the algorithm simulate lawsuits, just as AlphaZero did with Go. This would train the algorithm by giving it the experience of hundreds of millions of lawsuits—the experience and wisdom of a million lifetimes, which is more than any human judge could ever experience, let alone remember.

Good ideas to get better insights into judicial verdicts were proposed, some of which were quickly developed into early prototypes (Katz, 2012; Ashley, 2017). At the same time, however, more problems appeared down the road, especially regarding legal defensibility and the use of algorithms within the legal domain. Some of the main concerns that have been raised are as follows:

(i) The collection and storage of data for machine-learning algorithms may violate existing legislation (e.g., privacy laws, employment laws, laws dealing with police records, or laws dealing with intelligence and security services), as might the use of algorithms for making certain official or legal decisions.[18]

(ii) The use of algorithms in certain situations is considered undesirable or unethical. For example, think of profiling, making decisions with a large individual impact (such as adjudicating justice resulting in the deprivation of liberty), creating autonomous weapon systems (with a license to kill without a human in the loop), or monitoring and assessing individuals, as is currently the case in China. In this context, the great interest that totalitarian regimes have in AI and big data analytics is a justifiable cause of concern.

(iii) Machine learning algorithms are not always transparent and are often difficult to explain to non-mathematicians or professionals outside the field of computer science. The transparency and defensibility of such algorithms do not come naturally. This is a problem that already exists in large rule-based systems because it is difficult to know exactly what will happen with 1,000 decision rules. In a deep learning system with 100 billion parameters adjusted according to a complex algorithm, this is impossible to oversee or explain.[19]

(iv) Machine learning algorithms always contain some degree of bias. This bias must be known and measured, and its effects must be clear. In other words, there must be transparency,

[18] This concerns the applicable legislation within a certain jurisdiction, such as the AVG within the Netherlands, the Dutch version of the GDPR. Every country in the European Union has its own implementation of the GDPR. In the United States, separate privacy laws apply per state. An example of this is the California Consumer Privacy Act (CCPA). Determining the right legal framework is therefore not always easy. In addition, it may also be wise to take into account draft legislation, such as the recent AIA proposed by the European Parliament.

[19] Dr. Matt Turek of the Defense Advanced Research Project Agency (DARPA) is working hard on a research program, Explainable Artificial Intelligence (XAI), to make AI more explainable. His program will also look at different ways in which people explain decisions (e.g., “if you had done this in instead of that, the outcome would have been as follows”). There is a great need for this kind of research.

and bias must be taken into account [20]. The following forms of computer bias can be recognized:

a. *Selection bias*: You train an algorithm to drive you around using only daytime recordings. If you are going to use this algorithm at night, it will not work.
b. *Measurement or sensor bias*: There are certain filters or lenses on the camera that distort the data or do not measure certain extremes that are important for making the right choices.
c. *Algorithm bias*: You are using a linear algorithm for a nonlinear problem. For example, think of predicting the number of COVID-19 cases without considering the exponential nature of the outbreak. In other words, what are the mathematical limitations of the algorithm? Are there any simplifications or assumptions underlying the model? Do these pose problems in the real world?
d. *Bias or discrimination*: You train the algorithm with data in which bias or discrimination is ingrained. For example, you train a computer with a disproportionate number of photos of women cooking, such that, when in doubt, the algorithm will wrongly choose a woman in the recognition process. This is often difficult to avoid, because real-life datasets always contain some form of bias.[21]

(v) Even when the above-mentioned bias is dealt with correctly in the machine learning process, there may be other issues that are overlooked by a software engineer. A few examples to take into consideration are: is the quality of an algorithm or mathematical principle measured correctly by a software vendor [22], how well has the software been tested for deliberately incorrect input? In other words, how robust is its implementation? Are we dealing with a “fair weather sailor” implementation, or can it withstand a storm as well? Does the algorithm generate identical outputs for the same input? How stable is it? All of this has to be taking into account as well.

(vi) Self-learning algorithms are easy to fool. In 2015, Goodfellow et al. (2014, 2015) showed how deep learning image recognition algorithms could be easily fooled through so-called adversarial attacks via generative adversarial networks (GANs). In this case, certain weights of connections that are not used for the original classification task, are abused by storing “wrong” classifications in the model. For example, with a self-driving car, images of “stop”

[20] . See also O’Neill (2017) for a comprehensive treatment of bias in machine learning algorithms.

[21] An interesting consideration here is how “dealing with bias” can take on a political dimension. By adjusting the bias in data that is used for machine learning, it is also possible to give an algorithm certain political preferences. The same can happen when the bias of real-world data is adjusted to counteract undesirable social situations. For instance, when a data set used to determine if someone is eligible for a mortgage, includes postal code as a feature, then just living in a poor neighborhood (and not income of credit history) may lead to not getting a mortgage.

[22] A good example here is using *accuracy* for unbalanced data sets—that is, data sets with only a few relevant data points hidden in millions of non-relevant ones. Accuracy values are made very high by only focusing on the non-relevant ones and ignoring the few relevant ones. In such cases, one should use a set of measurements based on *precision* and *recall* to gain a full understanding of the performance of the algorithm. (See also: Chapter 8 of *Introduction in Information Retrieval* by Manning et al., 2009. https://nlp.stanford.edu/IR-book/pdf/08eval.pdf)

road signs could be manipulated so that the algorithm recognized them as indicating a 45 mph speed limit.[23]

(vii) The forensic integrity of the application of machine learning algorithms and the associated data makes special demands of end users that are not always well followed.[24] For example, think of the forensically correct (immutable) collection and recording of all data (the so-called "chain of custody"),[25] which records exactly how and by whom data actions are performed. However, data integrity and access to the system (e.g. cybersecurity aspects) aspects are also important: How can we be sure that data will not be hacked during a lawsuit or that this type of sensitive and personal data will not be leaked?

Social acceptance of the use of technology (and AI, in particular) within the legal domain will only be possible if the above problems are resolved.

If a dichotomy were to be made, then each of the above problems could be divided into one of the following two categories:

(i) bad science, including poor implementations and misuse of techniques; and
(ii) bad ethics, which also includes noncompliance with existing legislation for the sake of convenience.

The solutions to these problems therefore lie partly in technology and partly in ethics.[26] Only through an integrated and multidisciplinary collaboration between lawyers and computer scientists can both problems truly be solved.

[23] Further research has shown that these problems can be easily solved by training the algorithm for a longer period and with data that contains more noise. The same GANs that helped identify the problem can also be used to make the algorithms more robust.

[24] Technology used in legal contexts can have all kinds of functionalities that guarantee the forensic integrity of data, but if the end user does not apply the technology correctly, its legal defensibility can still be jeopardized.

[25] The chain of custody is a legal concept relating to the chronologically ordered documentation or paper trail that documents the sequence of preservation, control, transfer, analysis, and disposal of materials and information, including physical or electronic evidence. It is often a demanding process and is required to present legal evidence in court.

[26] In addition to valid reasons mentioned here that critically consider the use of technology within the legal domain, there are also legal professionals who ignore all scientific evidence that technology is beneficial to make better legal decisions or to execute legal processes better, faster and more consistently. The reasoning is often: "I see it, I understand the reasoning and the (scientific) evidence, but I still don't believe it." This phenomenon almost always concerns one of the following three categories of legal professionals. First, there are legal professionals who have a natural aversion to anything that has to do with technology. The author considers this to be a personal choice that should be respected. Second, there are legal professionals who are not primarily interested in the facts because the facts may be their detriment. These lawyers know that computer technology uncovers these facts faster (or inevitably). For example, think of criminal defense attorneys who must defend clients who are almost certainly guilty. Another example of this might be an email inquiry into a commercial dispute in which one of the parties knows they have been negligent. Finally and third, there are lawyers who are too dependent on a revenue model based on "billable hours" and who (want to) focus too little on efficiency. Unfortunately, these reasons may play a role in the background when discussing acceptance ad usage of legal technology, without being explicitly mentioned. When discussing the role of technology in the legal domain, it is important to recognize these agendas at an early stage so that the subsequent discussion can remain pure and constructive.

The Need to Collaborate

There will be very few people, if any, who deeply understand all aspects of legal technology—aspects that range from legal and ethical aspects, all the way to mathematics, algorithms, software engineering, machine learning, and user psychology.

Parties involved in using legal technology are currently mostly active in their own comfort zones:

(i) Lawyers have written extensively about algorithms and their potential conflicts with existing legislation (Evers et al., 2020) and ethical principles (Barger, 2008). The potential for discrimination by algorithms, and digital exclusion in particular, has gained most of their attention. A large number of lawyers are studying the relationship between legal tech and privacy legislation,[27] particularly regarding indicating what is not allowed under the new GDPR privacy laws.

(ii) Computer scientists, data scientists, and AI researchers are mostly concerned with investigating technical forms of bias and XAI to prevent bad science. For example, they are very interested in preventing adversarial attacks, especially because they pose such an interesting mathematical problem.

(iii) Forensic specialists are particularly interested in investigating detailed forensic (cybersecurity-related) technical problems.

There is too little real collaboration, so there are no clear paths leading to an integrated framework for the responsible use of legal technologies. This is not only a problem in government or in business but also in universities and colleges.[28] Why is this? Often, goodwill is there, but collaboration does not always work as well as intended.

I believe that the difference in how lawyers and computer scientists think about and approach problems is one of the reasons collaboration is suboptimal. Such a difference starts at an early age, with one's choice of specialization in science, language, or economics in secondary school. It then continues to grow while one is at law school or attending a computer science bachelor's program.[29] Some schools do teach law students a basic form of programming and give computer scientists an understanding of ethics, but this is really more of an exception than a rule.[30]

As early as their basic training, computer scientists and lawyers each learn a completely different approach—a different way of working, and even a different way of thinking. Bridging these differences starts with knowing that they exist, recognizing them, and then taking them into account, and address them in order to achieve real collaboration between computer scientists and legal professionals.

[27] That so much attention is paid to privacy legislation can be partly explained by the fact that the General Data Protection Regulation (GDPR) was the first general legislation by the European Union to regulate the relationship between data, technology, companies, and individuals.

[28] For example, there are universities of applied sciences with forensic and legal courses that are located next to each other or even in the same building, but until now there has been little or no cooperation among them.

[29] In this blog, I have explicitly named the differences: https://www.legaltechbridge.com/en/why-computer-scientists-and-legal-professional-think-so-differently. Of course, one should be careful about generalizing, but by naming these differences explicitly, they can be addressed more effectively.

[30] Teaching ethics and legal requirements as part of AI courses is a mandatory component of computer science and AI curricula in the Netherlands. Since 1980, the Delft University of Technology has been providing Joop Doorman's courses on ethics and philosophy to computer science students: https://nl.wikipedia.org/wiki/Joop_Doorman

The only solution to this problem is to follow an integrated and multidisciplinary route to collaboration between different domain experts. How can this be realized? How do we arrive at such an integrated, multidisciplinary approach? This is what is addressed in the next section.

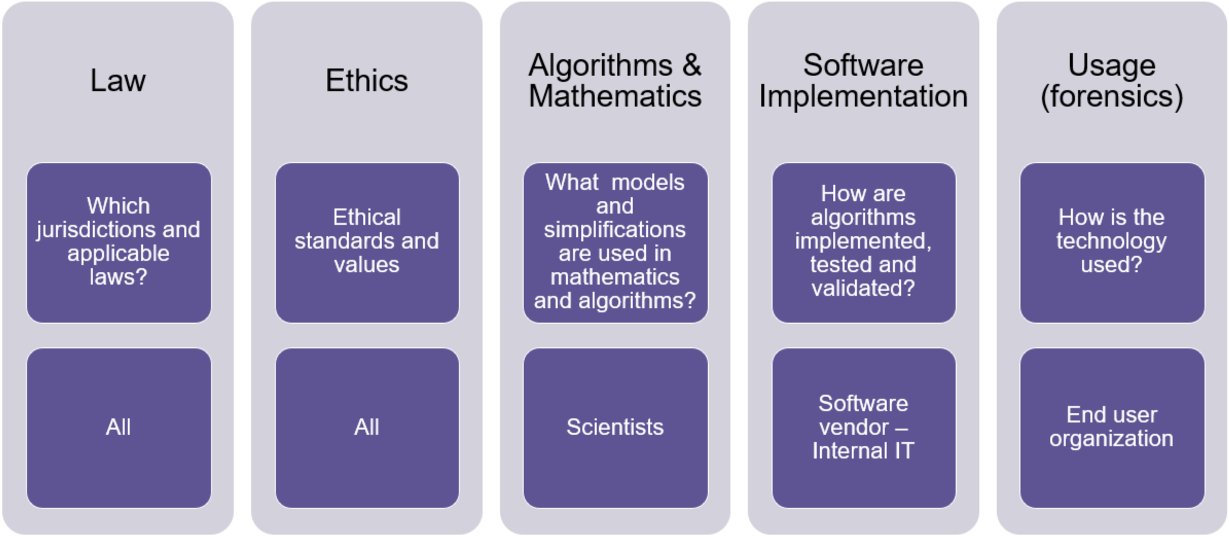
Law
Which jurisdictions and applicable laws?
All
Ethics
Ethical standards and values
All
Algorithms & Mathematics
What models and simplifications are used in mathematics and algorithms?
Scientists
Software Implementation
How are algorithms implemented, tested and validated?
Software vendor – Internal IT
Usage (forensics)
How is the technology used?
End user organization

Figure 1: An overview of the five components of the responsible application of technology (and AI, in particular) in legal contexts.

A Proposed Framework for Legal Defensibility

As described above, too many things can go wrong when applying technology in legal contexts. There is bad science, and there is bad ethics:

- Scientists can jump to conclusions, use the wrong metrics, or cherry-pick results;
- Implementations can be sloppy or rushed;
- Software engineers and programmers often do not fully understand the (mathematical) basis of algorithms; and
- Sales and marketing are also known to oversell technology or deliberately mystify it to create some form of proprietary exclusivity.[31]

Algorithms (especially in the field of complex human decision making) must be able to deal with many exceptions and rare occurrences. There is often a long list of such situations, and there is often not enough training data (or no data at all) to fine-tune the machine learning algorithms for all possible situations.[32]

When looking at legal defensibility, different viewpoints and approaches can be taken. Depending on one's professional or educational background, one's focus is often either on legal and ethical aspects or on mathematical and computer science aspects of applying technology in legal context. The focus is rarely on both. Even when all of the former aspects are addressed properly, there is still the risk that an end user will not apply the technology as designed and intended.

In other words, dealing with legal defensibility can be overwhelming, where to start is not always clear, and to obtain successful results, one must collaborate with professionals from distinctive disciplines.

However, we have been in similar overwhelming situations: around the year 2000, an analogous problem arose: cybersecurity. This, too, was a problem involving many different disciplines, from human resources and physical plant security to complex IT infrastructure and software. Weaknesses could be anywhere, and no one really knew where to start or how to enable all the required individuals to collaborate. This eventually resulted in the ISO-27001 framework of control points, assisting organizations in gaining (more) control over cybersecurity problems.

[31] A good example of this is how marketing terms such as *predictive coding, technology assisted review, computer assisted review, or continuous active learning* were used for text classification using active learning based on algorithms and principles in the 1980s and 1990s (Lewis at al., 2004).

[32] For example, in bio-medical applications, clinical tests on a population of 30,000 individuals do not reveal rare side effects of (COVID-19) vaccinations that occur only a few times per million people. This same problem exists in legal technology. There will always be issues in electronic data that the implementation of an algorithm is not prepared for.

In this paper, a slightly different approach is proposed—one that is not yet a full ISO standard, although such a standard will probably be developed in the next couple of years. Rumors indicate that the European Commission is already working on the design of such a framework as part of the AIA. For now, the following categorization of so-called "control points" can be used to facilitate the discussion and to arrive at a structured approach to legal defensibility of legal technology:

I. Legal
II. Ethical
III. Scientific, algorithmic, or mathematical
IV. Technical implementation
V. Forensic integrity

These different categories of control points are briefly discussed below.

I. The *legal* category includes existing laws and available case law. Of course, this depends on the legal application or process and on the applicable jurisdiction. It is also subject to interpretation, and case law may vary or even be contradictory.[33] Thus, while one should not always expect deterministic answers here, this is where one should start. At the very least, legal technology must be compliant with applicable laws, regulations, and case law.

II. The *ethical* category is even more challenging, as one's ethical standards may vary depending on one's culture, political system, and personal beliefs. The proposed framework uses European ethical standards in relation to the application of technology in our society (e.g., related to dealing with privacy, bias, automatic decision making, profiling, and the use of biometrics). This may be a bit vague for now, but such guidelines will be defined in more detail later.[34]

III. Software vendors often overlook the *scientific* point of view, as many just apply the technology without really understanding the underlying presumptions and limitations of the algorithms. Different components are also at play here. First, there are the general limitations of certain algorithms and mathematical principles, often resulting from simplifications or presumptions that allow for certain calculations. Does the algorithm have any particular parameter sensitivities? Are the right measurements used to quantitatively test the algorithms? Not fully understanding these questions can lead to disastrous results.[35] Here, one can also include the transparency of the algorithm at hand: How easy (or hard) is it to explain the behavior and results to laypeople?

[33] For instance, consider the differences between the privacy regulations in Europe (GDPR) and those in the United States, where every state has its own privacy regulations, which can also vary significantly between states.
[34] As the AIA proposed by the European Parliament has not yet been converted into national legislation, these principles are also considered part of European ethical standards and not yet seen as lawful regulations.
[35] Several examples can be provided in this context. A famous one is measuring the performance of search algorithms in terms of accuracy instead of using sets of precision/recall/f1 values. As can be shown, it is possible to have 99% accuracy with only 0.1% recall, meaning that one effectively misses many relevant documents and is exposed to a high risk of missing relevant information. Another example relates to certain statistical algorithms that presume independence between features, which is not at all the case if such features are measured by word frequencies. Clearly, the occurrence of certain words in a linguistic context is not independent from the occurrence of other words.

IV. Aspects of t*echnical implementation* are often best understood by software vendors. Continuous and thorough unit testing is necessary to assess the performance and quality of the implementation. Does the implementation result in similar results at different times and locations (e.g., is the implementation reproducible)? Can the implementation deal with noisy (erroneous or wrong) input data?

If machine learning is used, it is implemented based on certain datasets. Here, we must know where the data come from (data provenance). How and by whom are the data labeled? Is the disagreement between annotators clear and measured? Is this dataset representative of what can be expected as input during implementation?[36] Which types of bias could be present in this data (e.g., selection bias, measurement bias, or prejudice bias)? How does the implementation deal with these forms of bias?

V. Over the years, *forensic* integrity of technology have evolved to a be an important requirement of the usage of technology in sensitive areas such as legal or healthcare. Transparency and auditability of actions are paramount. This can be achieved by maintaining a so-called *chain of custody* of both data and algorithmic processes. A chain of custody must be capable of proving that the data that has been identified and collected at the beginning of a case (during evidence collection) has been handled correctly (i.e., that the data used for evidence has not been changed or manipulated during the collection), that preservation has been carried out according to the required standards, that the electronic evidence will be 100% identical at any later moment in time, and that all algorithmic and manual actions have been logged. Such logs should be stored as read-only data (including hashing to detect manual changes), and one should be able to conduct audits or generate detailed reports on the chain of custody, loggings, and processes.

Security standards[37] are also proposed to be included under the list of items related to forensic integrity, as they relate directly to the integrity of access, availability, and data. This integrity also depends on the actual implementation of legal technology in an organization. Therefore, the technology provider should furnish the tools necessary for organizations to make implementation decisions and implement the required controls.

By inventorying and listing individual items in these five categories, one can arrive at a framework of control points for legal defensibility. These individual control points can then be addressed by the most appropriate specialist. Altogether, this process will lead to full coverage of all aspects of the legal defensibility of the use of technology in legal contexts.

As a final advice: first and foremost, use common sense: Do not do to others what you do not want to be done to yourself.

[36] Here, think of the usage of an annotated machine learning dataset for named entity recognition (NER) that consists exclusively of (well-written) newspaper articles, which is then used to recognize named entities on (badly written) social media comments, Twitter feeds, or short emails. This will obviously not work well.

[37] For example, standards such as ISO-9001, SOC-2, Fed Ramp, and others.

Methodology

When compiling the proposed framework for the legal defensibility of legal (software) technology, provided in Appendix A, the following resources were used:

- Scientific principles and best practices from the fields of computer science (e.g., machine learning, data mining, text-mining, and natural language processing), as taught in computer science departments around the world
- The principles related to accountability and transparency communicated by the Association of Computing Machinery (ACM)
- The principles related to accountability and transparency communicated by the Institute of Electrical and Electronics Engineers (IEEE)
- The General Data Protection Regulation (GDPR) [38]
- The Artificial Intelligence Act (AIA) Proposed by the European Parliament [39]
- A selection of case law from the United States and the European Union (EU)
- The principles and best practices communicated at the Sedona Conference
- The principles and best practices communicated by the Association of Certified eDiscovery Specialists (ACEDS)
- The principles and best practices communicated by the Electronic Discovery Reference Model (EDRM)
- The principles and best practices communicated by the European Legal Technology Association (ELTA)
- A selection of recent literature on ethics, algorithms, AI in law, and legal tech in general
- Experience and communication with ZyLAB customers over many years

The authors are aware that there is more to be discovered on this topic. By no means is this paper intended to be exhaustive. Rather, the proposed framework is the first effort to provide an initial guideline for software vendors, users, and service providers to address the issue of legal defensibility in a structured manner by checking a number of control points.

Regarding the five categories of control points defined earlier, the above publications were checked for relevant control points, and those points were listed in the most logical order. In the chart in Appendix A, per control point, a short explanation is provided, or references are provided for further study and improved understanding.

[38] See https://gdpr-info.eu/ for a full overview of the GDPR.

[39] See https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206 for an overview of the "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."

Conclusions

In this paper, the first draft of a framework for the legal defensibility of legal (software) technology is proposed. This framework is based upon various (international) resources and is intended to provide a structured approach that can be used by software vendors, software users, legal professionals, service providers, and interested legal professionals and law students.

Further References

Ashley, 2017: Ashley, K. D. (2017). *Artificial Intelligence and Legal Analytics*. Cambridge University Press.

Barger, 2008: Barger, R. N. (2008). *Computer Ethics, A Case Based Approach*. Cambridge University Press.

Blair et al., 1985: Blair, D. C., & Maron, M. E. (1985). An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System. *28 COMM. OF THE ACM,* p. 289*.*

Ebers et al., 2020: Ebert, M., & Navas, S. (Eds.). (2020). *Algorithms and Law*. Cambridge University Press.

Daugherty et al., 2018: Daugherty, P. R., & Wilson, H. J. (2018). *Human + Machine: Reimagining Work in the Age of AI*. Harvard Business Review Press.

Dolin, 2017: Dolin, R. A. (June 20, 2017). *Measuring Legal Quality*. Harvard Law School, Center on the Legal Profession. (Also a chapter in Katz et al., 2021.)

Goodfellow et al., 2014: Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Networks (PDF). *Proceedings of the International Conference on Neural Information Processing Systems (NIPS 2014)*, pp. 2672–2680.

Goodfellow et al., 2015: Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). *Explaining and Harnessing Adversarial Examples*. ICLR.

Grossman et al., 2011: Grossman, M., & Cormack, G. (2011). Technology-Assisted Review in E-Discovery Can Be More Effective and More Efficient Than Exhaustive Manual Review. *Richmond Journal of Law and Technology*.

Herik, 1991: Jaap van den Herik, H. (1991). *Kunnen computers rechtspreken?* Gouda Quint. 9060008421.

Hartung et al., 2018: Hartung, M., Halbleib, G., & Bues, M.-M. (2018, 9). *Legal Tech.* Beck C. H.

Jacob et al., 2020: Jacob, K., Schlindler, D., & Strathausen, R. (Eds.). (2020, 8 28). *Liquid Legal.* Springer International Publishing.

Kanaan, 2020: Kanaan, M. (2020). *T-Minus AI: Humanity's Countdown to Artificial Intelligence and the New Pursuit of Global Power.*

Katz, 2012: Katz, D. M. (2012). *Quantitative Legal Prediction or How I Learned to Stop Worrying and Start Preparing for the Data-Driven Future of the Legal Services Industry.* Emory L. J.

Katz et al., 2021: Katz, D. (2021, 1). *Legal Informatics.* Cambridge University Press. doi:10.1017/9781316529683.009

Kelly, 2016: Kelly, K. (June 7, 2016). *The Inevitable: Understanding the 12 Technological Forces that Will Shape Our Future*.

Krishevsky et al., 2012: Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems* (pp. 1097–1105), 1, December 2012.

Kissinger et al., 2021: Kissinger, H. A., Schmidt, E., & Huttenlocher, D. (2021). *The Age of AI and Our Human Future*.

Myrto et al., 2020: Myrto, P., Klein, O., & Kissine, M. (2020). Is Justice Blind or Myopic? An Examination of the Effects of Meta-Cognitive Myopia and Truth Bias on Mock Jurors and Judges. *Judgment and Decision Making*, 15, 2 (March 2020), pp. 214–229.

O'Neil, 2017: O'Neil, C. (2017). *Weapons of math destruction.* Penguin Books.

Polanyi, 1967: Polanyi, M. (1967). *The Tacit Dimension.* Anchor Books.

Russel, 2020: Russell, S. J. (2020). *Artificial Intelligence: A Modern Approach* (5th ed.). Prentice Hall.

Scholtes et al., 2019: Scholtes, J., & van den Herik, H. J. (2019). Big Data Analytics for Legal Fact Finding. In: L. van den Berg, S. Geldermans, A. Heeres, N. Noort, J. van de Riet, S. Vonk, & R. Weijers (Eds.), *Recht en Technology, vraagstukken van de digitale revolutie* (1ste druk ed., pp. 47–62). Boom Juridisch.

Scholtes et al., 2021: Scholtes, J., & van den Herik, H. J. (2021). Big Data Analytics for e-Discovery. In: R. Vogl (Ed.), *Research Handbook on Big Data Law*. Edward Elgar Publishing.

Scholtes et al., 2022: Scholtes, J. and Jomaa, T. (2022). *A Proposed Framework for Legal Defensibility of Legal Technology* (first draft, January 2022). iPRO – ZyLAB White Paper.

Silver, 2016: Silver, D., Huang, A., Maddison, C., et al. (2016). Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature*, 529, pp. 484–489. https://doi.org/10.1038/nature16961

Silver 2017: Silver, D., Schrittwieser, J., Simonyan, K., et al. (2017). Mastering the Game of Go Without Human Knowledge. *Nature*, 550, pp. 354–359. https://doi.org/10.1038/nature24270

Susskind, 1987: Susskind, R. E. (1987). *Expert Systems in Law: A Jurisprudential Inquiry*. Clarendon Paperbacks.

Susskind, 2019: Susskind, R. (2019). *Online Courts and the Future of Justice*. Oxford University Press.

Turek, 2021: Turek, M. (2021). Explainable Artificial Intelligence (XAI). https://www.darpa.mil/program/explainable-artificial-intelligence

On Consciousness and limitations of AI:

- Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, 49, pp. 433–460.
- Dennett, D. C. (1992). *Consciousness Explained*.
- Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory (Philosophy of Mind)*.
- Weizenbaum, J. (1976). *Computer Power and Human Reason: From Judgment to Calculation.*
- Dreyfus, H. (1972). *What Computers Can't Do*.
- Searle, J. (1984). *Minds, Brains and Science: The 1984 Reith Lectures*. Harvard University Press. ISBN 978-0-674-57631-5; paperback: ISBN 0-674-57633-0.
- Christian, B. R. (2011). *The Most Human Human: What Talking with Computers Teaches Us about What It Means to Be Alive.*
- Christian, B. R. (2021). *The Alignment Problem. How Can Artificial Intelligence Learn Human Values.*
- Russell, S. J. (2021). Human-Compatible Artificial Intelligence. Human-Like Machine Intelligence.

On Deep Learning:

- Rosenblatt, F. (1958). The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain, Cornell Aeronautical Laboratory. *Psychological Review*, 65, 6, pp. 386–408. doi:10.1037/h0042519
- Minsky, M. L., & Papert, S. A. (1969). *Perceptrons*. MIT Press.
- Rumelhart, D. E., & McClelland, J. L. (1987). The PDP Perspective. In: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations* (Parts 1 and 2). MIT Press.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *In: Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'12)* (pp. 1097–1105). Curran Associates Inc.
- Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. arXiv (abs/1810.04805).

More on Artificial Intelligence

- Waisberg, N., & Hudek, A. (2021). *AI for Lawyers: How Artificial Intelligence Is Adding Value, Amplifying Expertise, and Transforming Careers*.
- Legg, M. (2020). *Artificial Intelligence and the Legal Profession*.
- Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A Flaw in Human Judgment*.
- Pinker, S. (2021). *Rationality: What It Is, Why It Seems Scarce, Why It Matters*.
- Chafkin, M. *The Contrarian: Peter Thiel and Silicon Valley's Pursuit of Power*.
- Kasparov, G. K., & Greengard, M. (2017). *Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins.*
- Scharre, P. (2019). *Army of None*.
- Wilson, H. J., & Daugherty, P. R. (2018). *Human + Machine: Reimagining Work in the Age of AI.*
- Kurzweil, R., Wilson, G., & Books (2019). *The Singularity Is Near: When Humans Transcend Biology.* Books on Tape.
- Du Sautoy, M. (2020). *CREATIVITY CODE: Art and Innovation in the Age of AI*.
- Silver, D., Huang, A., Maddison, C., et al. (2016). Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature*, 529, pp. 484–489. https://doi.org/10.1038/nature16961
- Newell, A., & Simon, H. A. (1972). *Human Problem Solving*. Prentice-Hall.
- Hayes-Roth, F., Waterman, D. A., & Lenat, D. B. (1983). *Building Expert Systems*. Addison-Wesley Longman Publishing Co.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- Metz, C. (2021). *Genius Makers: The Mavericks Who Brought AI to Google, Facebook, and the World*. Dutton, an imprint of Penguin Random House LLc.
- Myrto, P., Klein, O., & Kissine, M. (2020). Is Justice Blind or Myopic? An Examination of the Effects of Meta-Cognitive Myopia and Truth Bias on Mock Jurors and Judges. *Judgment and Decision Making*, 15, 2 (March 2020), pp. 214–229.
- Kanaan, M. (2020). *T-Minus AI: Humanity's Countdown to Artificial Intelligence and the New Pursuit of Global Power.*

Appendix A: Checklist for Legal Defensibility

Framework Proposal for Legal Defensibility

Legal

Control Name	Number	Control Objective	Control Targets	References
Applicable Jurisdictions	L.01	List all applicable jurisdictions in which the software will be used.	Legal	https://en.wikipedia.org/wiki/Jurisdiction
Applicable Legislation	L.02	Per jurisdiction, identify the applicable legislation (e.g., eDiscovery obligations under the Federal Rules of Civil Procedure (FRCP) or privacy regulations, such as the General Data Protection Regulations(GDPR) or the California Citizen Protection Act (CCPA), or employment laws in the case of internal investigations).	Legal	https://gdpr-info.eu/ https://www.uscourts.gov/sites/default/files/civil-rules-procedure-dec2017_0.pdf https://oag.ca.gov/privacy/ccpa https://nl.wikipedia.org/wiki/Arbeidsrecht_(Nederland) https://nl.wikipedia.org/wiki/Arbeidsrecht_(Belgi%C3%AB)
Restrictions & Obligations from Applicable Legislation	L.02b	Per applicable piece of legislation, identify detailed lists of obligations and restrictions.	Legal	See examples of such restrictions and regulations under the GDPR. E-Discovery Reference Model (EDRM), Association of Certified eDiscovery Specialists (ACEDS), and the Sedona Conference developed documentation for specific obligations and best practices under the FRCP.
Applicable Information Disclosure Requirements	L.03	In certain jurisdictions and under certain legislation, there are special requirements in relation to information disclosure. For example, the U.S. DoJ, SEC, and FTC require the use of certain production formats and reporting. The same applies to several acts related to freedom of information and disclosing	Legal	https://www.justice.gov/atr/case-document/file/494686/download https://www.ftc.gov/sites/default/files/attachments/bc-production-guide/bcproductionguide.pdf https://www.sec.gov/divisions/enforce/datadeliverystandards.pdf https://uk.practicallaw.thomsonreuters.com/w-003-3364?transitionType=Default&contextData=%28sc.Default%29 https://wetten.overheid.nl/BWBR0005252/2018-07-28 and many more…

		information in public records.		
Restrictions on Collection	L.04	Various restrictions may apply when obtaining electronic information for identification, collection, and preservation. Think of legal aspects such as subsidiarity, proportionality, or permission from labor boards, just to name a few.	Legal	https://en.wikipedia.org/wiki/Proportionality_(law) https://en.wikipedia.org/wiki/Subsidiarity Various local restrictions apply (often in relation to employment law) to the collection of email and other electronic data from employees in the case of internal or regulatory investigations.
Applicable Case Law	L.05	Per jurisdiction, additional case law may apply.	Legal	
Contractual Restrictions	L.06	The organization may have signed commercial or government contracts that impose additional restrictions or obligations.	Legal	
Patent Restrictions	L.07	Depending on the jurisdiction, patents may limit development methods.	Legal & Data Science	https://en.wikipedia.org/wiki/Software_patent https://www.upcounsel.com/software-patent
Copyright Restrictions	L.08	Datasets used for machine learning may be governed by copyright restrictions.	Legal & Data Science	https://www.twobirds.com/en/news/articles/2019/global/big-data-and-issues-and-opportunities-ip-rights https://www.lawsitesblog.com/2020/12/legal-research-company-ross-to-shut-down-under-pressure-of-thomson-reuters-lawsuit.html
Open Source Licenses	L.09	License agreements for the open source machine learning and other AI libraries or tooling used may be restricted and not allow (free) commercial usage.	Legal & Data Science	https://en.wikipedia.org/wiki/Comparison_of_free_and_open-source_software_licences

Ethical

Control Name	Number	Control Objective	Control Targets	References
Applicable Jurisdictions	E.01	List all applicable jurisdictions in which the software will be used.	Legal	
Applicable Ethical Framework(s)	E.02	Per jurisdiction, identify the applicable or accepted ethical frameworks. This may also include internal corporate guidelines.	Legal	
Restrictions & Obligations from Applicable Ethical Framework(s)	E.02b	**Per** piece of applicable legislation, **identify** detailed lists of obligations and restrictions**.**	Legal & Data Science	
Risk	E.03	Are we are dealing with a high-risk, medium-risk, or low-risk application as defined in the **European** AIA**?**	Legal & Data Science	https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

Scientific

Control Name	Number	Control Objective	Control Targets	References
Sensitive Areas of Application Using AI or Data Science	S.01	List all areas of the applications in which AI or data science is used.	Data Science	
Mathematical Models Used	S.02	For each area, identify which mathematical models are used.	Data Science	
Risks, Limitations, Simplifications Sensitivities, etc., per Mathematical Model	S.02b	Per algorithm, identify risks, limitations, sensitivities, etc.	Data Science	
Algorithms Used	S.03	For each area, identify which algorithms are used.	Data Science	
Risks, Limitations, Simplifications Sensitivities, etc., per Algorithm.	S.04	Per algorithm, identify risks, limitations, sensitivities, etc.	Data Science	
Pre-Trained Machine Learning Used for Training?	S.05	Were any components of the algorithms derived using machine learning?	Data Science	
Parameters on Pre-Trained Machine Learning Data	S.06	Dealing with data provenance, bias, disagreement on manual labeling of machine learning data, etc.	Data Science	
Are Algorithms Transparent and Reproducible?	S.07	Are the algorithms open-source? Is how they work transparent? Are the results reproducible?	Data Science	
Robustness	S.08	Have the algorithms been tested against erroneous data?	Data Science	
Quality Measurements	S.09	How is the quality of the algorithms measured?	Data Science	
Explainable	S.10	What methods exist to explain the behavior of the algorithms to laypeople?	Data Science	
Case law	S.11	Is there case law referring to the use of these algorithms?	Legal	

Usage in the Legal Industry	s.12	Is anybody else in the legal industry using these algorithms? If so, what has their experience been like?	Data Science / PM	Gartner, IDC, Forrester, …

Implementation

Control Name	Number	Control Objective	Control Targets	References
Sensitive Areas of the Implementation Using AI or Data Science	I.01	List all areas of the software applications in which AI or data science is used.	Development	
Open Source Used or Non-Open Source Third-Party Libraries Used	I.02	Make an inventory of all third-party libraries.	Development	
Risks, Limitations, Simplifications, Sensitivities, etc., of I.02	I.03	Per library, identify risks, limitations, sensitivities, etc.	Development	
Are Algorithms Transparent and Reproducible?	I.04	Are the algorithms open-source? Is how they work transparent? Are the results reproducible?	Development	
Machine Learning Used for Training?	I.05	Are any components of the implementation derived using machine learning?	Development	
Parameters on Pretrained Machine Learning Data	I.06	Dealing with data provenance, bias, disagreement on manual labeling of machine learning data, etc.	Development	
Robustness	I.07	Have the algorithms been tested against erroneous data?	Development	
Quality Measurements	I.08	How is the quality of the algorithms measured?	Development	
Explanation	I.09	What methods exist to explain the behavior of the algorithms to laypeople?	Development	
Testing	I.10	How is the implementation tested?	Development	

Forensic Integrity of Software Usage

Control Name	Number	Control Objective	Control Targets	References
Chain of Custody	U.01	Can it be proven that the data that has been identified and collected at the beginning of a case (i.e., during evidence collection) has been done correctly (i.e., that the data used for evidence has not been changed or manipulated) during the collection?	Development & Users	
Collection	U.02	Has collection been done according to existing standards?	Development & Users	
Preservation	U.03	Has preservation been done according to the required standards, such that the electronic evidence is 100% identical at any later moment in time and that all algorithmic and manual actions are logged?	Development & Users	
Loggings	U.04	Are all user and system actions logged?	Development & Users	
Reporting	U.05	Can the necessary reports be created for items such as, but not limited to, all logs, audit trails, security (user access, data storage, …), and sampling?	Development & Users	
Security	U.06	How is cybersecurity and data safety guaranteed?	Development & Users	ISO 27001, SOC-2, ….
Sampling	U.07	Can users validate the quality of automatic (AI) processes by using sampling?	Development & Users	