AI, Law and Human Responsibility

By Gregor Noll

ABSTRACT

What do algorithmic technologies do to the law, how do they alter lawyers' work on legal issues, and how do they affect the allocation of legal responsibility? If it turns out that algorithmic technologies make it harder to identify a responsible subject, can we do something about it? These are the questions that I am trying to answer in this article. After mapping how AI affects the law and the legal profession, I inquire into the factors distinguishing legal normativity from such normativity as is expressed in algorithmic technologies. I conclude that law and the cybernetic basis of AI conflict with each other in a way beyond remedy. AI fundamentally undermines lawyers' ability to attribute responsibility, as humans and algorithmic technology amalgamate in practice. I propose that the lawmaker imposes strict responsibility on certain forms of AI to avoid a loss of accountability during the period where traditional law and cybernetic normativity overlap.

1. INTRODUCTION

What do algorithmic technologies do to the law, how do they alter lawyers' work on legal issues, and how do they affect the allocation of legal responsibility? If it turns out that algorithmic technologies make it harder to identify a responsible subject, can we do something about it? These are the questions I shall answer in this article. In the next section, I will characterize three relationships between algorithmic technologies and law: tech law, tech as law and legal tech. At the end of this section, I hope to have made clear, in the most general terms, what algorithmic technologies do to the law. In the next step, I will explain why law and algorithmic technologies cannot be reconciled (section 3). I will present my argument by drawing on concrete examples: the assessment of evidence and of intent, respectively (section 4). Both sections 3 and 4 give us an idea of how algorithmic technologies affect lawyers' work on legal issues and challenge basic tenets of legal responsibility. At this stage, the strategic question emerges what shape to give to the relationship between algorithmic technologies and the law in the future. By proposing a strict form of liability, I shall point to a possible way forward allocating responsibility for negative consequences of algorithmic technologies (section 5). Finally, I will present my conclusions in section 6.

As my questions are about the changes brought to law by algorithmic technologies, they impose a comparison between old and new on us. But what is that 'old' against which these novel technologies are compared to? It is probably some kind of established working method in legal practice, where people read, underline, take notes, think, write and discuss, all prompted by a legal issue. It is based on human language, and it is enacted in the passing world of paper envelopes and archive binders, meetings around worn office tables and conversations with colleagues at the coffee machine. While we might want to streamline and automate our way out of this world of routines, we also want to remain in it – at least in part. But could it be the case that we *idealise* the human and language-borne elements of this world, as if the time of reading, thinking, writing and discussing were unlimited, our colleagues were extremely wise and easy to deal with at all times, and our desire for truth limitless?

Let me use an example to guard against such idealisations. Let us say that at some point in the future we will analyse a concrete AI application that a Swedish court considers using in its decision-making. Let us also assume that we will soon find strong arguments that this application would remove something human from judicial decision-making and sentencing by that court. However, the ordinary work situation at the court is perhaps so demanding on staff that it is less than human. The number of judges employed to deal with cases before the Court may not have increased to the same extent as the influx of cases. Maybe the judges have too little time for their cases. Politicians may not be willing to inject additional funds to meet the increased workload at the courts in the midst of a recession. Our intended comparison now includes quite a few parameters: on the one hand, a new technology that may in part replace human legal work and, on the other hand, the influx of cases, the size and competence of staff, as well as the impact of politics and the economy on the concrete working situation of a court. Such a comprehensive and nuanced analysis comparing the introduction of a new technology with the real situation before its introduction is beyond the reach of this article, but we can all agree that it is necessary. Technology shifts, such as the transition to AI-based applications, never take place in the abstract, but in a concrete field, a concrete factory floor, or, for that matter, in a concrete office in a court in a medium-sized Swedish city. When I try to draw out big lines for the development in the following, I have to reserve myself for what I cannot do here: namely to analyze how these lines interact with the concrete life situation where the concrete changes through algorithmic technologies happen.

2 TECH LAW, TECH AS LAW AND LEGAL TECH

Today's legal rules do not necessarily resonate well with digitalisation and AI. We hear this from those who develop and implement technology: law is often perceived as an impediment, an innovation barrier.[1] On the other hand, AI and digitalisation are perceived by lawyers as an increasingly important normative force that competes with the law as we know it. An example of the latter is Facebook's so-called *Oversight Board*, a group of people who monitor how the platform applies its own rules on what can be posted. Doesn't that seem like a privately owned Supreme Court for the limits of free speech in a significant part of the internet? Also, we are seeing more and more applications that bring AI directly into the lawyer's everyday practice. They change legal work from within and confront us with the question of who should take responsibility for the results of this new way of working.

Let's take a closer look at the relationship between algorithmic technologies and the law, and put words on what characterizes them. It can be divided into three categories: tech law, tech as law and legal tech.

Tech law is the term for a growing field where law is applied to digitalisation and AI. As these technological solutions dominate an increasing part of the economy, society and everyday life, they engender new legal issues and conflicts. Think about how much of your private and professional transactions are made through digital platforms. Think about how you continuously consent to app developers' terms and conditions when downloading or updating apps. Or take the legal issues that need to be resolved every time a new app is placed on the market: stretching from liability limitation to intellectual property. More and more practicing lawyers are engaged in tech law on a daily basis. Contract law and copyright law are important regulations in this work, as is public law. The threat to ban TikTok made by the United States (US) in 2020 is a reminder that states may well use powerful legal tools to assert their regulatory power vis-à-vis large technology companies with millions of users. In a nutshell, tech law is an area where law seeks to dominate technology.

In *tech as law*, the opposite is the case: here AI and digitalisation are placed between the human and the law, gaining power and influence from that intermediary position. If a search engine constantly responds with the same top-rated hits to my search, it is no wonder if I start to think that these hits really are the most relevant ones. If a healthcare application repeatedly suggests a certain diagnosis for a certain bodily condition, it is no wonder that a doctor advised by it begins to build his treatment strategy based on this diagnosis. Or, if a decision supporting application processing claims for income support repeatedly proposes rejection for applicants from a particular population group, it is no wonder that administrators would start to assume that such applicants are generally not eligible. AI and digitalisation shape human behaviour in a way reminiscent of how law shapes human behaviour. As long as we are convinced that these applications do a better job than humans (whose time, attention span and cognitive capabilities appear to be much more limited), we accept the answers technology provides to our questions. Then these answers are normative. They come with a presumption of validity very similar to that of the law: we assume that legal rules shall be followed, unless that presumption has been rebutted. There are examples where law clashes with tech as law. Take the example of discrimination, which is proscribed by international law, EU law and Swedish law, but which some AI applications have actually been shown to promote. A well-known example is COMPAS, a decision support system for probation in the US, which was found to be biased against persons of colour.[2] Another example is the decision support system for school grades in England, which has been shown to disadvantage pupils from poor areas and benefit pupils from private schools.[3]

Legal tech is about the automation of tasks that are otherwise carried out by lawyers. Why let a colleague go through contracts or court cases page by page when a legal tech application is so much faster? In particular, if technology permits us to review *all* the agreements entered into by a company, or *all* legal cases that have been adjudicated by a particular court tier? Of course, new technology has implications for the professional role of the lawyer and for legal method *tout court*. A 2016 study by Deloitte, the consulting firm, claimed that jobs for 'traditional lawyers' will diminish as legal tech gains ground in the coming years.[4] Less skilled jobs are those expected to disappear. At the same time, technological solutions that can quickly solve simpler legal tasks will drive down the price of legal services while opening up new markets. Those who previously could not afford to hire The Big Law Firm, or were intimidated by its symbolic attributes, may be able to purchase automated legal services on a simple and consumer-friendly platform that combines a chatbot with a chatting *in vivo* lawyer. As in journalism, there is a real risk that certain legal jobs will be trivialised or simply considered superfluous when such platforms take over tasks traditionally performed by lawyers.

1   Simon Larsson and Kristoffer Bengtsson, *An explorative study of social dimensions of intelligent automation of the final assembly in the automotive industry: The views of Swedish expert stakeholders* (School of Business, Economics and Law, University of Gothenburg, 2021).

2   Adrienne Brackey found that African Americans were more likely than Caucasians to be given higher scores under COMPAS regardless of their recidivism rate. Adrienne Brackey, *Analysis of Racial Bias in Northpointe's COMPAS Algorithm* (2019).

3   A detailed explanation of the decision support equation can be found here: <https://www.theguardian.com/education/2020/aug/21/ofqual-exams-algorithm-why-did-it-fail-make-grade-a-levels> Accessed 4 December 2021.

4   Deloitte, *Developing Legal Talent. Stepping into the Future of the Law Industry* (2016) 16 <https://www2.deloitte.com/uk/en/pages/audit/articles/developing-legal-talent.html> Accessed 4 December 2021.
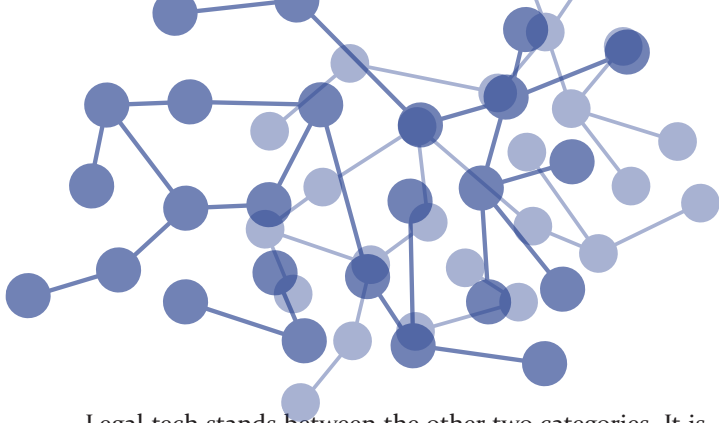
Legal tech stands between the other two categories. It is related to both tech law (consider, for example, the contract on whose basis a legal tech application or its services are purchased) and to tech as law (as decision support in legal matters can produce normative effects in its own right). Legal tech has the potential to reform law from within, as opposed to the legislator or the market, both reforming the law from its outside. The fact that it reforms the law from within means that it changes the very craft of the legal profession. By choosing automation and AI over our traditional craftsmanship, we as lawyers contribute to changing our own profession.

Now one might argue that legal tech brings about change, but that the law in its totality remains the same, much like a river remains the same even though new water is constantly flowing through it. Could it be that we are succumbing to the hype around new technologies, and that the real changes are no greater than those brought about by the iron plow, the steam engine and nuclear power or other technology at the time of its introduction? Could it be the case that the changes brought about by AI are felt, but not critical to the very identity of law?

3 IS LAW FACED WITH EPOCHAL CHANGE?

I think the law is facing a veritable epochal shift. Take a simple description of how the law works: we go back a few thousand years, taking three monotheistic religions – Judaism, Christianity and Islam – as examples. In these traditions, God hands over stone tablets with a law to humans, making divine norms available in writing. This process we usually call codification. Humans then study the law - often with the help of scholars - and try to understand what it means for their daily lives. They attempt to live by this understanding, sometimes they succeed, at other times, they fail. New questions arise and humans return to the law to understand it better and live a more law-abiding life. Fast forward less than two thousand years to our secular societies: not much changes. The Parliament hands over a law to humans, humans study the law - often with the help of lawyers - and try to understand what it means for their daily lives. They attempt to live by this understanding, sometimes they succeed, at

other times, they fail. The failures cause humans to go back to the law to study it and its meaning anew (again with the help of lawyers), hoping to do the right thing on the next occasion. What is central to this description is that the law is *written law*. By being written down, it is separated from us as humans: it has its own existence outside ours. The written law must be studied by us as humans. Such study is a prerequisite for compliance. 'Studying the law' might mean that we sit with our heads bowed over the text of the law, or that we let a trusted person interpret the law for us. There is a commonality between the laws of monotheistic religions and today's secular laws: the law is outside of us, outside our body, and it is supposed to enter our thinking, our acting and our daily life. The study of the law is *the* central factor that makes the law enter our lives. As you read this article, you are part of this tradition.

What I have described here is a simple sequence based on three monotheistic religions – Judaism, Christianity, Islam – but which work equally well in today's secular societies. Replace God with the legislator, who codifies political norms in the form of the law, and you will obtain the following model:

Legislator -> codification -> human study of the law -> compliance with the law

What is characteristic of this model is its strict distinction between the law and the human. Let us compare it with a corresponding process involving artificial intelligence, such as a machine learning application in legal tech. Even such an application rests on codification, but that is where resemblances end. Most prominently, human language is relegated to the sidelines. In machine learning, coding is done through code languages that enable suitable combinations of algorithms and learning data. These are chosen by app developers and data experts. Compared to the parliamentary lawmaker (or God in the three monotheistic religions named earlier), they do not possess formal authority. The algorithm is based on mathematics, not human language, and data is binary. Another difference is the meaning of codification is not studied by a human. Instead, it is the encoded application that studies data – data that in some form is assumed to depict reality, such as X-rays of human lungs in medical AI or PDF files with legal cases in legal tech. The application studies an excerpt of reality rather than any form of legislation. When it gives us the results of these studies, it is often not possible to fully reconstruct how it arrived at these results – perhaps because it has processed so much data that it would be impossible for us as humans to reconstruct its

5 Ronald Yu and Gabriele Spina Alì, 'What's inside the black box? AI challenges for lawyers and researchers' [2019] 19 Legal Information Management 2.

6 Daniel L. Chen et al., 'Early Predictability of Asylum Court Decisions' [2018] Proceedings *of the ACM Conference on AI and the Law*

ICAIL '17. <https://users.nber.org/~dlchen/papers/Early_Predictability_of_Asylum_Court_Decisions.pdf> Accessed 4 December 2021.

7 Theodore W. Ruger, Pauline T. Kim, Andrew D. Martin and Kevin M. Quinn, 'The Supreme Court Forecasting Project: Legal and Political

Science Approaches to Predicting Supreme Court Decisionmaking' [2004] 104 *Columbia Law Review* 1150. <https://doi.org/10.2307/4099370> Accessed 4 December 2021.

approach step by step within a reasonable time. Then there are machine learning algorithms whose steps cannot be reconstructed by a human, however much time one would spend on it (usually, these steps are referred to as taking place in a 'black box').[5] Algorithmic technologies are thus based on a completely different normative model, which can be summarized as follows:

The question of reality - > a human encodes and uses training data -> the machine learning application learns and processes a larger amount of data taken from reality -> the result is used in reality if humans so decide

At first, it may seem that humans play a heavy normative role, as they are the ones starting the whole process. It is a human being who asks and encodes the question of reality into an algorithm. But notice how cognitive work, namely the studies, has been removed from humans and delegated to the machine. The traditional legislative model relies heavily on human beings at the study stage: humans who act, witness events, interpret legal norms and pass judgments. What these humans have in common is that they unite experience and reflection in one body. In the algorithmic model, this is complicated by the algorithm 'experiencing' and 'thinking' as it works its way through the data. Although it is a human who decides whether or not to use the results rendered by the AI application, this human finds it difficult to understand the cognitive work of the machine as well as his or her own. I am not saying that the old model with a human in the middle is good, and that the algorithmic model with the machine in the middle is bad. What I am saying is that they are different, and that we do not yet have a strategy for how they should interact coherently.

On reflection, it is the question 'what is reality?´ that governs here, not some form of legislator. On the one hand, it is the reality as perceived by those humans who have worked to define the problem that the algorithm is supposed to solve, designed the algorithm, produced the data to train the algorithm with and selected the data that the algorithm will eventually process. On the other hand, it is reality as it emerges in the data that the algorithm is trained on or ultimately processes. The algorithm is assumed to reflect reality in such a way that the results produced by the algorithm can be used as a basis for decisions. If the traditional model with the legislator at the top raises the question 'what does the legislator want?', then the new model raises the question 'what does reality want?'. Anyone who can say 'how things really are' with the help of the algorithm – someone who defines the problem, designs and trains the algorithm – will win in the battle for power in society. However, we know that the algorithm is just one among several interpreters of reality and, as an interpreter, it can make mistakes both legally and ethically. As my previous examples of tech as law show, these errors can lead to violations of discrimination prohibitions and other human rights. What seemed to be a shortcut to more objective and rational governance of our societies at first is, on closer inspection, an amalgamation of subjective and irrational elements. In the monotheistic model, I had

to believe in a God. In its secular version, I believe in the ability of parliament to make well elaborated laws. The new algorithmic model also requires me to believe in something, namely the machine's superior ability to sort out reality for us.

4 TWO EXAMPLES: EVIDENCE ASSESSMENT AND INTENT

In section 3 above, I highlighted the risk of discrimination when algorithmic decision support is used. But each coin has two sides: with the help of algorithms, we can also find out how flawed human decision-makers are. My first example leads us straight into a central question when judges ponder a concrete legal case: how do they assess evidence? This issue is important in legal doctrine as well as in the everyday life of legal practice. Lawyers tend to believe that evidence assessment plays a major role for the outcome of legal proceedings. Trained judges and decision-makers ensure that justice is being done by carefully considering everything adduced in the case, sifting proof from it and assessing its importance as objectively as possible. This image is probably also shared by the public.

AI can help us test whether this image is correct. Concretely, we might ask whether it is possible to confidently predict the outcome of an asylum case simply by knowing the nationality of the applicant and the name of the judge, thereby excluding evidentiary assessment? A 2018 study by Chen, Dunn, Sagun and Sirin confirms that this is possible.[6] Using machine learning, the four authors plowed through 21 million documents involving 800,000 asylum cases in the United States. Having access to the name, judge and the nationality of the asylum seeker only, their algorithm was able to accurately predict about 80% of the outcomes. What is surprising – and troubling to us as lawyers – is that such predictability is possible without drawing on the concrete evidence in the case or its evaluation by the judge. A reasonable conclusion is that the assessment of evidence in the individual case does not actually matter at all as much as we think. That, in turn, leads us to question the whole story of how trained judges and decision-makers produce just judgments by carefully considering all evidence invoked in the case, deciding what it proves, and passing judgment on its basis.

Now, one might object that this example is extreme. Asylum law is an area of law with relatively vague rules that leave a certain margin of discretion in its interpretation of material norms. The discretionary margin of the judge increases in size, as the asylum seeker's oral testimony is often the only evidence on offer. But, then I would point to a 2002 study comparing an algorithmic prediction of outcomes in the US Supreme Court with human experts' predictions of the same. Although predictive technology was not as developed back in 2002, the algorithm won over the human experts with a 75% accuracy of the machine compared to a 59% accuracy of the expert.[7] This study covered judgments in all areas of law within the jurisdiction of the Supreme Court, not merely the right to asylum. This tells us that algorithms can be trained to predict outcomes in a wide range of cases better than legal experts.

The ability of AI to predict judgments and decisions has created its own market in legal tech. Let me add a few examples of commercial services and applications on the market at the time of writing: *Intraspexion* sends alerts when the system has identified a certain risk that the client will be sued. It bases its warnings on document analysis. *Premonition* predicts a particular lawyer's success prospects by analyzing the number of processes he or she has won, together with the length and the type of process. *Bryter* provided an experienced lawyer in a major international law firm with a decision support tool where he could integrate his own mathematical method of process risk analysis.[8] The last example shows that lawyers may leave their own mark on certain apps, integrating legal competence into an algorithmic platform.

As was the case for the 2018 study by Chen, Dunn, Sagun and Sirin, *Premonition* is also based on the human factor in law. It sells its ability to predict outcomes for the individual lawyer, and to put them into an economical context. Being so open with your own processual track record is something that *practising* lawyers are rather unfamiliar with. AI has the ability to make our successes and failures transparent in a new way. This impacts everyone, from the junior lawyer at a smaller law firm to the respected judge in a top-level court. As the 2018 asylum law study has shown, patterns of performance brought into the open by AI may call into question the legitimacy of the judge's privileged position. As one computer scientist commented in an oral presentation of the 2018 study: 'Why are we spending our tax dollars on something functioning as badly as lawyers?'

It is not only the legal sector that is exposed to a critical review enabled by AI and digitalization. In the financial industry, human investment advisers have been partly overtaken by *quants*: quantitative investment funds whose choices rest on AI-driven analysis. In quantitative history research, mathematicians have sought to identify the cyclical laws of history, equipped with large databases and extensive computer processing power. Traditional historians have felt this competition for interpretative supremacy. Why should lawyers and their business be spared the question why the taxpayer or a client should pay more and get less done (or be worse off) compared to if the algorithm were to do the job? Tech as law rests exactly on this logic: the algorithm starts to dominate traditional law by undermining its legitimacy and by replacing or modifying its rules as it goes along.

Take the app *Premonitions*, an AI app predicting the success rate of a particular lawyer: its appeal rests on *the normative power of the factual*, the respect for a description that we take to mirror reality. The same applies to quantitative investment funds or quantitative history writing. *They seem to be able to tell us what is going on in real life.* If lawyers truly are so predictable, or if the stock prices or history as a whole are, why should we not make direct use of this predictability? Since the beginning of the 17th century, we are living in an era marked by the Enlightenment, the scientific revolution and industrialisation. The logical order of things forms the basis of these three, and whoever manages to articulate the logic behind this order has access to what governs reality. This view gave rise to cybernetics back in the 1940s: a way of thinking the world in terms of controllability (*kybernetes* is the Greek term for helmsman). Being able to govern the world is exactly what the law is about, and thus the law and cybernetics enter into direct competition with each other. The law is based on the responsible person, both in the form of the sovereign legislator and that of the law-abiding subject. Cybernetics do not give humans a special status and make no distinction between them and other life forms. For law, it is a problem if we can no longer distinguish between a human and a machine in algorithmic technologies, because law assumes that the world is run by people who take responsibility for legal compliance. For cybernetics, it may not matter if it is a human, a machine, or a symbiosis of both that causes a phenomenon. It has not invested in a worldview with a responsible person at its heart. The algorithm in today's AI solutions delivers a response based on defined parameters. By contrast to a human being, it has no intent, and neither does it judge anyone as a human does. It merely produces a logical outcome, based on the data it is processing.

Therefore, the normative power of algorithms cannot be subordinated to the normative power of law, since the former neither assumes nor reproduces the central role of human beings responsible for themselves and for the world around them. The two forces simply cannot be integrated with each other. It is as if attempts were made to integrate market liberalism and command economy with each other, or Shintoism with Catholicism.

Now, we might suggest that cybernetics are but an extension of the secular mindset, freeing itself from old theological notions affording a special status to the human. As an extension of centuries of enlightenment, scientific revolution and industrialization, would it not be logical to afford it a greater normative space, together with the algorithmic technologies growing from it? Assumptions about history as a progress story and the actual spread and dominance of technology add further legitimacy to cybernetics and AI, and it is this factor that can be decisive in a situation where legal norms and algorithmic technology compete with each other.

What do we miss when this progress story grows and seizes influence at the expense of law? We are missing out on being a *responsible human being*. The human as such remains, with all her cognitive deficiencies and prejudice, but with the algorithm as a judge, she does not have to bear the consequences of his contribution to tech as law. Why is that so? Replacing part of the human cognitive process with algorithmic technology will reduce our ability to track human actions and responsibilities for the same, both in terms of the basis on which a judicial decision is made and in terms of criminal liability. With algorithmic technology, the criteria of intent or negligence as a prerequisite for individual responsibility are gradually phased out. Here, I would like to add my second example: in my previous research, I have investigated whether an advanced weapons system with elements of machine learning could be used in such a way that the laws of war and international criminal law remained fully applicable to the user.[9] Operating an algorithmic and largely autonomous weapon of this kind *and* subjecting it to the law is a con-

tradiction in terms. The use of the weapon system presupposes that human cognition and machine cognition are integrated to such an extent that we cannot know, for example, whether an attack on civilians – which is illegal under the laws of war – is covered by the intent of the system's human users (as the *actus reus* needs to correspond to a *mens rea* in order for it to be punishable). Without intent, no criminal liability, and without criminal liability, the international law prohibition of attacks on civilians remains toothless.

These problems are not specific to international law and criminal law. They will appear elsewhere, for example in contract law. How are we to ascertain the intentions of the contracting parties, if they have been shaped and formulated in interaction with an advanced algorithmic system? The insidious thing is that the law seems to be taking care of algorithmic weapon systems just like any other weapon system, or that the law is taking care of an automated contract just like any normal contract. In order to understand what the problem is, it is important to understand how technology undermines the conditions of the law – namely the possibility of the law to understand human cognition, prove human intentions and make humans responsible on that basis.

5 WHAT TO DO?

The problem at the core of this article is actually quite hard to perceive in the course of our daily lives. The law just keeps marching on, dealing with an increasing number of tech law questions, and, step by step, being reformed from within by legal tech. To the extent that legal tech is all about streamlining document search, case management and due diligence, everything seems to be in order. But gradually, legal tech is being used by practicing lawyers for a sort of analytical task where human and mechanical cognition, human intent and the inscrutable pathways of the algorithm amalgamate into each other. And we are increasingly confronted with legal tech claims regarding algorithmic decision support systems. This seems to be the price of progress.

What is the alternative to its incremental acceptance? Is it possible simply to ban algorithmic applications that risk blurring the dividing line between human and mechanical cognition and intentionality? I do not think so. Such applications already exist in a variety of sectors such as trade, medicine, economics, finance, the arts and the media. Because they exist, they generate tech law issues, which in turn risk undermining fundamental assumptions of the law on the central position of humans. Similarly, it would be difficult to put an end to the rapidly expanding market of advanced legal tech systems. The temptation to streamline and automate is huge, for the business sector as much as for the public sector.

What might it look like when a lawyer pays the price of progress in a concrete work situation? Let me briefly sketch up a fictitious situation where Simon, a lawyer working for a business named *LeanIn!*, meets Ayla, his boss. The scene takes place in a meeting room at the head office in Gothenburg.

'Words are superfluous.' Ayla turned her back on him, exhaled audibly, walked a few steps towards the door, paused, turned around. Simon just sat there and did not move his eyes from the point where she had been standing a few moments ago. 'As we have recommended it to the customer, we have to show that we trust it ourselves, in our own work', Ayla continued. 'Just give them a number indicating how many documents the app sifted through, an abstract of the three type-cast judgments that are trending in it, and then sign off on it. Lengthy justifications send the wrong signal! Let the material speak for itself!' She left the room. Evidently, the conversation had slipped out of Simon's hands.

His name and his signature were in the documents that the app had sifted through. In a previous life at the Administrative Court in Gothenburg, he had written judgments, and signed them, too. That was the reason why he was recruited to work here. LeanIn! helped their customers to significantly reduce payroll costs through client-adapted AI solutions. His new employer sold two things: the feeling of not buying the first best system coming along in a market full of lingo, but really the best. And, second, the feeling of being able to significantly reduce wage costs while maintaining full productivity. The promise of 'sustained productivity' was exactly what the conflict with Ayla was all about. Was it really the same if three human beings agreed on a sentence, or if the application compiled and analyzed lots of judgments and determined what kind of judgments had the best chance of not being overturned in a court of appeal?
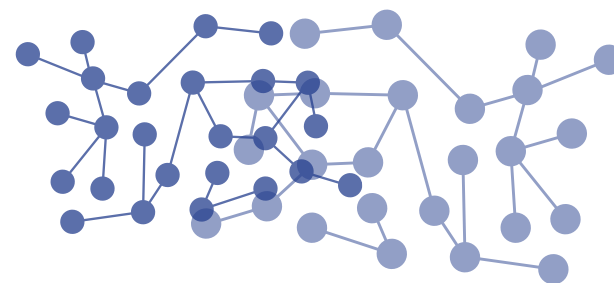
8   Bryter 2021. *Lego for Laywers* (undated press release), <https://bryter.com/press-releases/lego-for-laywers/> Accessed 4 December 2021.

9   Gregor Noll, 'Weaponising neurotechnology: International humanitarian law and the loss of language' [2014] *London Review of International Law*, 2(2): 201-231. <https://doi.org/10.1093/lril/lru009> Accessed 4 December 2021; and Gregor Noll, 'War and Algorithm: The End of Law?' in Max Liljefors, Gregor Noll and Daniel Steuer (eds.), *War and algorithm* (Rowman & Littlefield International 2019) <https://rowman.com/WebDocs/War_and_Algorithm_Open_Access_Liljefors_Noll_Steuer.pdf> Accessed 4 December 2021.

Again, Martin's name and signature were required, but just no more. LeanIn! had made a pilot for the Migration Court in Gothenburg, and, according to the contract, it was required that the first run of the application be certified by a lawyer as equivalent to legal work performed by a human being. Simon was this certifying lawyer, and he would attest that the client could well replace a number of persons – in fact, persons who had the same degree as Simon himself, and who were distant colleagues in some sense. Per Ayla's directive, he would not write a longer justification for the customer, detailing why the application worked as well as a number of lawyers would. Was it really OK to sign straight off? After all, it was Simon's signature that opened the way for the pilot to be tested at the Migration Court, and perhaps put into service for good. At each funeral, a few words are said about the deceased – why not here?

Is Simon acting as a real lawyer if he certifies with a mere sentence and his signature that the outcome of the application was sufficiently similar to the outcome that a human judge would generate? Does this statement not require motivation, so that others understand how he thinks? Should he assume responsibility for this, tell Ayla that he simply cannot put his signature under the certificate as the lawyer he is, and prepare to look for another job? Or is this a much larger question, beyond Simon and Ayla's horizon? Can it be answered at all at the individual level, or does it require a say for *everyone* affected by the technology – now and in the future?

What to do? Responsibility is a political question of great consequence, and so are the answers we give to it. During the 19th Century, the law was confronted with a not too dissimilar challenge: industrialisation brought an ever greater division of labour and increased efficiency. It also made work processes in manufacturing, distribution and consumption much more complex, to the point where it was hard to understand how a chemical plant or a steelworks functioned in all its capillaries. Risks to workers, consumers, the environment and the outside world grew in tandem to growing complexity and division of labour. It became more difficult for the courts to identify those responsible for the most dangerous activities as long as the concepts of human intent or negligence were central to the process. In Germany, the lawmaker intervened and

prescribed strict responsibility for certain dangerous industries. This placed responsibility with the person who runs the business in which an accident occurred. Human intent or negligence needed not be demonstrated in those cases. Rail, electricity networks, aviation or nuclear power are classic examples of dangerous activities where strict responsibility has been applied in one form or another.

Now, one might object that strict responsibility is a deterrent to innovation. If applied to AI, it would remove at least some of the efficiency gains that algorithmic technologies engender. However, I would be prepared to pay this price in order to retain the human as a subject of responsibility. Strict responsibility would be assumed for algorithmic systems that use strong forms of machine learning – a form of learning taking place in a black box.

It was precisely the human ability to understand and reconstruct what AI does that formed the core of Simon and Ayla's conflict. Simon is a lawyer; he is asked to certify that the AI system achieves the same results as human lawyers would. At the same time, Ayla argued that he must not justify his certification in lengthy explanations, as such a justification can open up for criticism and, paradoxically, make the AI platform appear less credible. The absence of justification means that the court who buys the services of the *LeanIn!* platform cannot know why Simon certifies that human and machine render equivalent legal results, while the machine does the work faster and more efficiently. Simon has not shown that he actually achieved the same results as the application. Then we understand that Ayla and *LeanIn!* at large simply ask us to believe in the ability of its platform to produce legally correct results. As we are asked to believe that, it makes sense that *LeanIn!* bears the legal and economic consequences when this belief proves to be unfounded (for example, if it turns out that using the AI system leads to discriminatory consequences in sentencing). Therefore, placing strict responsibility on *LeanIn!* and other sellers of black box-applications would be a reasonable governance move.[10]

Imagine for a moment that Sweden had legislated on strict liability along these lines. What would have changed in the conversation between Simon and Ayla? Simon's wish to write a lengthy justification for his certification would have been met with open arms. Ayla, his boss, would have felt the weight of responsibility on her shoulders; she would not dare to deploy the system at the court

[10] Drawing on a human rights framework, Karen Yeung argues that states that have committed to human rights must ensure that those who derive benefits from automated data processing and different forms of artificial intelligence are held responsible for their risks and consequences. Karen Yeung, *A Study of the Implications of Advanced Digital Technologies (Including AI Systems) for the Concept of Responsibility Within a Human Rights Framework* (SSRN 2018) <https://ssrn. com/abstract=3286027> Accessed 4 December 2021.

[11] Xiaoxuan Liu et al., 'Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension' [2020] 26 Nat Med 1364 <https:// doi.org/10.1038/s41591-020-1034-x> Accessed 4 December 2021.

[12] European Commission, Proposal for an Artificial Intelligence Act, COM 2021/206 final, 21 April 2021.

[13] The relevant statement is formulated in a more hedged way: 'The ethical responsibility and liability for the decisions and actions based in any way on an AI system should always ultimately be attributable to AI actors corresponding to their role in the life cycle of the AI system'. UNESCO, *Recommendation on the Ethics of Artificial Intelligence* (2021) Doc. No. 41 C/73 Annex, p 11, para 42.

without basing it on a thorough and detailed analysis of the effects of the system, including any risks of discriminatory outcomes. Simon's analysis would have had to be understandable for non-lawyers. The pilot had become more complicated and expensive, and the application might be less effective, but potential damage had been prevented.

Would strict liability legislation weaken Sweden in a technological competition with countries such as the US and China? What if it came with offset measures: substantial government funds would be invested into AI research so that better algorithms are developed. This might secure a sort of pole position for a future market for algorithms that perform very well (such as strong machine learning algorithms may do), but whose outcomes are sufficiently transparent to humans to exclude discriminatory effects.

What is more, machine learning research has run into a fundamental problem: an experiment made by a research team with a particular method often does not produce the same results when repeated by another research team using the very same method. This disparity in results has provoked a lively debate in computer science; reproducibility is, after all, a fundamental criterion for a research experiment or a practical application based on science. In recent years, researchers have begun developing guidelines to increase the degree of reproducibility, as in medical research using AI.[11] Computer scientists want to increase transparency, as they see the risk of losing their scientific credentials. In fact, imposing strict responsibility on AI applications would provide another reason to intensify work on these guidelines. Strict responsibility gives Ayla a reason to have the risks of the application for the Migration Court investigated, as much as it gives programmers a reason to make applications more transparent and rationally reconstructable to humans. We are far from a suitable change of the law, however. In 2021, the European Commission has proposed legislation for AI systems that shies away from the imposition of strict responsibility.[12] In the same year, UNESCO has proposed a set of non-binding guidelines, which emphasizes the observance of human rights, but also stops short of calling for the imposition of strict responsibility.[13]

6 CONCLUSIONS

What does AI do to law, to the legal profession and to human responsibility? Those were the questions this article sought answers to. At this point, it is evident that the introduction of tech law, legal tech and tech as law, all enabled by massive technological advances in AI and digitalization, challenge law on a fundamental level. Today, the operation of law rests to quite some extent on the lawyers' ability to isolate and analyze human cognition in a chain of events. AI fundamentally undermines this ability, because humans and algorithmic technology amalgamate in practice, and cannot be isolated from each other for the purposes of responsibility attribution. As during earlier epochal shifts driven by technology and science, law is slow in finding a response, while markets are moving fast, setting norms as they go along. Neither is

there a quick fix that would permit the law to offset the challenge, as it rests on a fundament of ideas that are incompatible with cybernetic thinking underlying algorithmic technologies.

To buy time and to create space for democracy as well as for the legal profession, I proposed the introduction of strict liability for certain forms of algorithmic technologies. Is that not a contradiction in terms, as I have just submitted that there is a fundamental conflict between law and AI? Not necessarily. Epochal change tends to come with long transitional phases, where societies gradually adapt to a new order. Traditional law, based on a monotheistic heritage, has overlapped with cybernetic normativity since the 1940s; this overlap will remain with us for quite some time. Strict responsibility is the lawyer's way of pointing out the existence of a serious conflict between law and the cybernetic basis of algorithms. This form of liability would be gravel in the gearbox that drives history. Whether this means that the gearbox only hacks for a short time to then resume work, if it needs major repair, or if it breaks completely, is a concrete question. As we come into our own as humans, the answers will be evident to us.
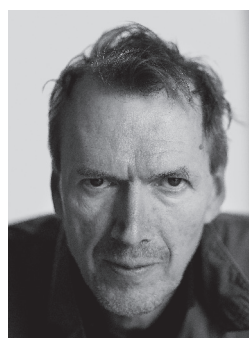
Gregor Noll

Gregor Noll is a Torsten Söderberg Research Professor at the School of Business, Economics and Law, Gothenburg, Professor of International Law, Department of Law, University of Gothenburg. Research for this article has been undertaken with funding by Riksbankens Jubileumsfond.