JOURNAL OF
CROSS-DISCIPLINARY RESEARCH
IN COMPUTATIONAL LAW [CRCL]

Scoping AI & Law Projects: Wanting It All is Counterproductive

Denis Merigoux*

Abstract

The intersection of law and computer science has been dominated for decades by a community that self-identifies with the pursuit of 'artificial intelligence'. This self-identification is not a coincidence; many AI & Law researchers have expressed their interest in the ideologically-charged idea-utopia of government by machines, and the field of artificial intelligence aligns with the pursuit of all-encompassing systems that could crack the very diverse nature of legal tasks. As a consequence, a lot of theoretical and practical work has been carried in the AI & Law community with the objective of creating logic-based, knowledge-based or machine-learning-based systems that could eventually 'solve' any legal task. This 'want-it-all' research attitude echoes some of the debates in my home field of formal methods around formalization of programming languages and proofs, and this position paper is the occasion for me to expand the line of reasoning developed in my PhD dissertation. Hence, I will argue here that the quest for an unscoped system that does it all is counterproductive for multiple reasons. First, because these systems perform generally poorly on everything rather than being good at one task, and most legal applications have high correctness standards. Second, because it yields artifacts that are very difficult to evaluate in order to build a sound methodology for advancing the field. Third, because it nudges into technological choices that require large infrastructure-building (sometimes on a global scale) before reaping benefits and encouraging adoption. Fourth, because it distracts efforts away from the basic applications of legal technologies that have been neglected by the research community.The critique presented in this paper is mostly technical. However, I also believe that a shift towards smaller-scale and domain-specific systems and tooling can foster genuine cross-disciplinary collaborations. These collaborations could form the basis for bottom-up approaches that respect the Rule of Law rather than twisting it for the needs of the system.

Keywords: AI & Law, legal formalism, cybernetics, semantic web, prolog

Journal of Cross-disciplinary Research in Computational Law

DOI: pending

www.journalcrcl.org

* Starting researcher at Inria Paris, denis.merigoux@inria.fr.

Introduction

In their recent presentation, Gebru and Torres [2023] relate how Gebru, originally an electrical engineer, evolved from criticizing technical aspects of machine learning models to discovering and explicating the underlying ideologies that drive the latest advances of machine learning. According to Gebru and Torres, the leaders of the field of Artificial Intelligence (AI), are influenced by the TESCREAL[1] bundle of ideologies. These ideologies view machine learning as a key enabler to artificial general intelligence (AGI), a god-like entity that could bring about either utopia or apocalypse for mankind. While the debate around AI and its future is raging and goes beyond the scope of this paper, the presentation of Gebru particularly struck me as I share the externality of her point of view. Indeed, I was trained in formal methods–a subfield of computer science close to its theoretical foundations and aimed at raising the level of assurance of critical software–and incidentally ended up the intersection of Computer Science (CS) and Law. This intersection is dominated by the field of AI & Law, as per the names of the eponym flagship journal and main conference (*International Conference on Artificial Intelligence and Law*, ICAIL).

Why is AI dominating this intersection? As new academic venues like ACM's CS & Law conference, the Programming Languages and the Law (ProLaLa) workshop and the CRCL conference and journal are emerging for scholars at the intersection of CS and Law, the consensus around the affiliation to the broader field of AI is being questioned. In this paper, I will investigate what the affiliation to the field of AI means, ideologically and technically, for legal technology research. First, I will make a link between the self-proclaimed goals of leading pieces of research in AI & Law and two powerful ideologies that aim at defining the ideal form of government, legal formalism and cybernetics. This link will help me emphasize the technological goals pursued by the field of AI & Law, and why they create a tension with the Rule of Law and democratic decision-making as explained by Hildebrandt [2020]. Second, I will analyse how the overreaching goals of AI & Law created such unrealistic expectations towards technological solutions, that researchers are basically compelled to come up with systems that 'want it all'. Finally, I will argue that pursuing systems that 'want it all' is counterproductive technically, but also socially as it delays the adoption of advanced legal technologies in domains where it is really needed.

This position paper aims at sparking a debate in the community and maybe shuffle the priorities of the field. More research will be needed to fully validate the claims that I make here, and I am sure valid opposing arguments can be raised. But my personal position expressed in this paper is that, rather than frantically looking for a way of integrating the latest glowing artifact of AI – like ChatGPT – into the existing AI & Law agenda, the field ought to return to the roots of the scientific method and start building more adequate evaluation frameworks for the output of its research. By adapting its technical solutions to the imperative of the Rule of Law and the real needs of the legal profession, a new era of applied research at the intersection of CS and Law could be ushered in.

The Tantalizing Promise of Government by Machines

In this section, I will take the leading scholars of AI & Law at their own words and try to make some links with broader ideologies that share an ideal vision of government by machines.

A first stop is the white paper by Genesereth [2015] that describes the orientation for research done at CodeX, the Stanford center for legal informatics. At first, Genesereth starts with a real-world problem: laws can be very complex and that affects compliance, efficiency and trust. Then, without considering how actual lawyers and subjects of law deal with legal complexity, he goes for textbook techno-solutionism: '*fortunately, these problems are not insurmountable. To the extent that they are information problems, they can be mitigated by information technology*'. The technical solution put forward here is computational law, which basically amounts to building a formal representation of legal knowledge and have legal decisions computed from it. He then lists examples where this has been put in

[1] Transhumanism, Extropianism, Singularitarianism, Cosmism, Rationalism, Effective Altruism, Longtermism.

practice, and acknowledges a fundamental limitation: '*the resolution of [the problem of the open texture of laws] is to limit the application of Computational Law to those cases where such issues can be externalized or marginalized*'. But this limitation does not seem to prevent computational law from being deployed ubiquitously: '*you are walking through the woods of Maine and see an attractive flower. You take a photo with your iPhone. Your plant app identifies it as a type of orchid and lets you know. At the same time, your legal app tells that, no, you may not pick it*'. Finally, going further than the ubiquitous use, Genesereth claims that '*in a way, Computational Law is the next step in the evolution of the legal system*'. The discourse here ends on a messianic note and computational law evolves from a technical solution to real-world problems into the inevitable future of the legal solution.

When looking closely, we can see that computational law, as touted by Genesereth, is heir to the old ideal of legal formalism. In the context of AI & Law, the important aspect of legal formalism is that laws and their enforcement should follow logical rules and must not depend on contextual elements that can change over time and place. Legal formalism is an old idea that peaked at the junction of the XIXth and XXth centuries and was already criticized back then, for instance by Dewey [1924]. But the field of AI, which was dominated at the end of the XXth century by the formal representation of knowledge, quickly recognized that its goals coincided with legal formalism: '*AI and Law is much more than an applications area. Its concerns touch upon issues at the very heart of AI: reasoning, representation, and learning. For the AI researcher interested in symbolic methods—or methods of whatever stripe—that are focused on providing explanations and justifications, AI and Law is an excellent arena*' [Rissland et al. 2003]. Thus, the general promise of efficiency, accessibility and uniformity of legal formalism are the primary motivations of legal knowledge representation projects. The formal representation of legal knowledge and concepts is thus viewed as the gateway to finally access the benefits promised by legal formalism, though this intervention of AI into Law and society, in general, gathered criticism early for its tendency to conflate judgment with computation, as per the thesis of Weizenbaum [1976].

The specific uses for formal representations of legal knowledge are diverse: most AI & Law scholars state that the intended use is to explain, justify, teach or better understand legal concepts. However, once a formal representation is made, it can be used for more than that. In that sense, the applications of AI & Law technologies can be qualified as opportunistic; once created for benevolent purposes, an artefact can be repurposed for more involved applications. I will present here three examples of this potential switch in applications for a given technology. First, the question whether judges should be replaced by the computer execution of a formal legal knowledge representation has been asked in the past by D'Amato [1977]. Some AI & Law scholars have rejected this idea categorically: '*A juridical machine can thus only be an aid to the jurist and not a substitute for him. We shall have no "electronic judges" in the world to come, any more than we shall have a machine to rule us*' [Mehl 1959]. But recently, replacing judges with machines has taken a very concrete outlook with the advent of machine-learning based prediction, as criticized by Medvedeva et al. [2023]. Second, a classic use for formal representations of legal knowledge is assistance to legal drafting: '*an executable, logic-based representation of rules and regulations can be used not only to apply the rules, but to aid the process of drafting and redrafting the rules in the first place–a point that was made by Allen [1956]*' [Sergot, Sadri, et al. 1986]. But today the 'assistance' to legal drafting is switching to a mandatory requirement that laws should be drafted in a 'digital-ready' fashion, with explicit guidelines about how the legal text should follow a precise logical structure. Indeed, Denmark passed its Digital-Ready Legislation Act [Plesner and Justesen 2022], and the OECD [2019] and the European Commission[2] are supporting similar initiatives. Third, the rise (and fall) of blockchain technology and its derivatives made the idea of automating contracts far more realistic, as noted by Crafa et al. [2023]: '*Since parties are free to express their agreement in the language and medium they choose (freedom of form, a principle shared by modern legal systems), drafting a contract by using a programming language (rather than, as usual, natural language) seems a valuable option. Advantages are in terms of speed-up, lack of ambiguity, and automatic and transparent enforcement of the contractual clauses.* '

[2] See the 'Digital-Ready Drafting' track at the SEMIC 2023 conference.

The tantalizing opportunity of using AI & Law applications beyond academic circles to directly intervene in public policies is where AI & Law meets a second powerful ideology: cybernetics. The science of systems founded in the second half of the XXth century was quickly endorsed by scholars who wanted to apply it to society and the State. But before that, in France, Mehl [1957] theorized how the State Apparatus could benefit from cybernetic principles to increase the efficiency of its administration: '*the administration can thus be seen as a cybernetic system, but with its own specific aspects. Administration operates solely on information. Its counterpart in the world of machines is the "computer", not the machine tool. Administrative information is rarely imprecise and sometimes erroneous. Administrative action is altered by random phenomena. As a result, it takes on the appearance of a strategy*'[3]. Later, Mehl and Breton [1970] and Catala et al. [1974] created one of the first French legal databases in order to materialize the cybernetic ambitions of automating the analysis of legal cases by the administration [Mehl 1959]. Knapp and Vrecion [1970] corroborates this link between cybernetics and the nascent AI & Law techniques, and provides more examples of similar projects in the USSR, Czechoslovakia and the United States. In Italy, Contissa et al. [2021] relate the theoretical work of Frosini [1968, 1973] inspired by the same ideas. More recently, cybernetics has been making a comeback in AI & Law, for example through the work of Bourcier [2017], Mehl's former student, Potvin [2023] and Potvin et al. [2021] and their idea of a computer infrastructure for distributing rules to actors, or Sileno [2016], who cites a couple cybernetician references and explicitly takes a cybernetics approach: '*[...] in order to fulfill their mandate, the responsible authorities must put in place adequate activities to target known and hypothetical non-compliance patterns, along with anticipatory, discovery mechanisms to unveil new ones. But non-compliance is only half of the story. [...] In short, public administrations have to adapt their allocation of resources and scheduling of activities in accordance with the social environment in which they operate, and to the requirements set by the legal system*'. The crux of the issue to integrate AI & Law with cybernetics is to make a precise enough formal model of law and society as a system to be able to analyze it. This modelling activity involves formalization and knowledge representation, which is why the technological solutions developed in the context of AI & Law are highly relevant for cybernetician endeavours.

So far, we have highlighted the links of the AI & Law community with two powerful ideologies: legal formalism and cybernetics. These links all point to a shared goal, which is building formal representations of law and the objects it regulates, in order to first explain and justify existing phenomena and maybe later to automate aspects of the legal system and the administration of society. While AI & Law projects are careful when stating their ambitions and most scholars would reject the literal idea of government by machines when asked about it, it becomes evident that the field is enticed by a utopia. In this utopia, law is formally codified and instantly accessible through digital mediation, automatically enforced, applied without bias and enables a very efficient society where uncertainty, risk, frictions and delays are reduced to a minimum. I will not discuss here whether that utopia is desirable or not; however I can remind the reader some of the contradictions between this utopia and the Rule of Law as it exists currently in democratic societies. I would refer the reader to the formidable work of the COHUBICOL typology [Diver et al. 2022] for further analysis and examples.

Hence, the technologies developed in the field of AI & Law share the common goal of modelling law and society to get closer to the benefits of the tantalizing promise of government by machines. What have been the (practical) results since the 1980's? So far, not many; as pointed out already by Oskamp and Lauritsen [2002] who justify the lack of success by offering two reasons (summarized). First, law is too complex and hard to model. Second, the users (lawyers) are too conservative and ignorant about technology. '*For most of the past fifteen years practicing lawyers and AI researchers appear to have been locked into parallel worlds of theoretically uninspiring implementations and tiny brittle research applications. Robust traffic across that disciplinary divide has yet to develop.*' That was true in 2002 and, I claim, to some extent still true in 2023, so perhaps there is a deeper explanation behind the two reasons provided here.

[3] My translation from French.

Technical Systems that Want It All

I claim here, that because of the utopian nature of the goals that the field of AI & Law has imposed upon itself, it has been compelled to only research technologies that are all-encompassing, attempting to solve every problem at once while twisting the practice of the Rule of Law. In a nutshell, AI & Law technical systems 'want it all' and this is the cause of their lack of success in the real world. To substantiate this claim, I will discuss technical aspects based on the recent AI & Law retrospective by Governatori et al. [2022], Sartor et al. [2022] and Villata et al. [2022].

In the first decade (1992-2002), the focus of AI & Law was, discovering the logical features necessary to formally represent laws, norms, regulations, court cases and argumentations. This focus yielded several key theoretical results (deontic logic, defeasible logic, argumentation schemes, isomorphism) that constitute an efficient state of the art for legal formalization. However, and since the beginning, the confrontation between theory and practice has yielded poor results: '*Sergot, Kamble, et al. [1991] argued that it is appropriate to follow an isomorphic approach if the legislation is itself well structured, but otherwise this approach might become cumbersome. However, legislation is very often not well structured. In such a case, isomorphism would lead to a poorly structured knowledge base, one which fails to correspond to the "real world" problem.*' [Governatori et al. 2022]. The crux of the issue is that law does not always conform to the expectations of legal formalism: some enacted and enforced norms may be contradictory or ill-conceived, but they are nonetheless in force. This problems is solved through social processes, where parties choose an interpretation that suits their interest. A vague or ambiguous regulation can even be made on-purpose to serve the interest of the regulator [Torny 2005]. Hence, AI & Law projects tackling real-world legal situations with knowledge representation tools are forced into a dilemma, that I will illustrate with quotes from the seminal AI & Law paper by Sergot, Sadri, et al. [1986]:

1. Either developers simplify the model to avoid the complexity of the real-world and stay within the realm of what is formalisable but then the model loses a lot of its utility: '*the simplest way to handle vagueness is to assume that the vague concepts always apply and to use this assumption to generate qualified answers*'.
2. Or developers complexify the model by incorporating more and more bits of real practice into the model. However, this may require escalading weird logical features to account for the irrationality of reality, at the risk of rendering the model unusable (impossible to execute or maintain it): '*a more sophisticated approach might combine this with the use of rules of thumb that reduce vague concepts to concrete ones, but are not guaranteed to cover all cases. The rules of thumb arise from the analysis of previous cases. We deliberately avoided such complications and chose the simpler alternative in our implementation of the act*'.

Because of the influence of the ideologies discussed in the previous section, AI & Law scholars need their model to be as precise and rich as possible to preserve their potential for infinite reuse into all areas of applications. Hence, they usually want to choose option (2) of the dilemma, unless they hit the hard technological limitations of their tools, in which case they stick with option (1). Consequently, most AI & Law scholars tend to choose technological platforms that allow for very general and open-ended computer modelling. During the first decade of AI & Law, at the turn of the 1990's, the most versatile formal modelling tool around was Prolog [Colmerauer and Roussel 1996], which is not surprising as the goal of Prolog was to be the ultimate meta-language in which to declare formal systems. But modelling versatility in computer science tools comes with trade-offs. For example, running Prolog programs requires the use of a Horn clauses solver that requires a heavy runtime and may be a source of inefficiency. While it is possible to write efficient programs in, for example, SWI-Prolog [Wielemaker et al. 2012], the efficiency is conditioned by the use of a very strict subset of the features of Prolog that is correctly optimized by the interpreter, a subset that may not match what is required to elegantly and concisely model the law. Moreover, the diversity of opinions and research projects around Prolog transformed it into a family of languages [Körner et al. 2022, Tables 1 and 2] that share a common core, but where each has its strengths, weaknesses and quirks. Furthermore, there isn't yet a consensus in the AI & Law community about the exact variant of Prolog to be used. For instance, new AI & Law projects by Arias et al. [2021] or Lim et al. [2022] have switched to using Answer Set Programming (ASP) or its variants, ASP being itself a variant of Prolog with a completely different semantics and set of implementations.

The diversification and non-interoperability of models built by AI & Law projects attracted the attention of another branch of the AI research community, focused on semantic representations and ontologies. This shift corresponds to the second decade of the AI & Law retrospective, and at first the goal of introducing ontologies is well circumscribed by Breuker et al. [2004]: '*an ontology makes explicit the concepts and their properties one is committed to in modeling a domain. Note that we do not consider an ontology itself to be a model of a domain: it is used to have unambiguous and shared terms in the model*'. So then, ontologies could be viewed as a soft tool for aligning diverses models into interoperability. As such, they were object of great developments in the 2000's and 2010's. For instance, Barabucci et al. [2010] and Monica Palmirani and Vitali [2011] designed Akoma Ntoso, an ontology for structuring the presentation of legal documents around the world, which then became a UN-endorsed [Peroni et al. 2017] standard adopted by many legal publication offices around the world. The success of this ontology relies on its very low semantic content: its killer features are the ability to declare precisely what is a paragraph, a list, a document, etc. and cross-reference these items. Akoma Ntoso as an ontology does not try to express what the law means, but merely how it is structured. Thus, it falls short of actually modelling the law in the sense of the ideologies discusses in the previous section.

But more generally, as a very versatile tool, ontologies can also turn themselves into the formal models they were meant to align: '*as ontologies contain generic knowledge, cost-effective knowledge engineering may benefit from its reuse potential. Indeed, one can argue that the use of ontologies in AI comes from research in the late 80s and 90s that aimed at improving knowledge engineering by creating "well-structured" knowledge bases that would not only solve the problem at hand but be more maintainable, easier to extend, etc. In this sense, ontologies are then very much an engineering tool. This role of ontologies implies the use of an inference engine that is used to conclude specific goals*' [Breuker et al. 2004]. Here, the tantalizing opportunity to expand the use of ontologies in the legal domain into more computationally involved uses thrived: '*there is urgent need to find a robust and expressive XML annotation, compliant with the Semantic Web technologies, able to meet all the unique particular aspects rising from the legal domain and in the same time close the gap between legal text descriptions, using XML techniques, and norms modeling, in order to realize an integrated and self-contained representation of legal resources available on the Web*' [Monica Palmirani, Governatori, et al. 2011][4]. One of the most salient works in this line of research led to the creation of the LegalRuleML ontology [Athan et al. 2015] on top of Akoma Ntoso, which embedded all the logical features needed for legal formalisation discovered in the previous decade. In a *grand finale* proof of concept, Monica Palmirani and Governatori [2018] present the integration of all the ontology-based tools to showcase an example of a legal model of GDPR being used for automatically enforcing it (on a specific platform). However, this proof of concept is criticized by Novotná and Libal [2022]: '*[...] they do not deal with multiple interpretations and they do not specify the cooperation with legal experts. Secondly, they don't provide use examples and any evaluation of correctness or usability of the system.*'

As [Monica Palmirani and Governatori 2018, Figure 1] depicts, the actual technological artefact behind this *tour de force* is comprised of a dozen different software tools interacting with each other, making up a complex architecture that required years of infrastructure building, typical of the high-modernist (in the sense of Scott [1998]) approach to computer science that I criticise in my PhD dissertation [Merigoux 2021]. '*In our framework, presented below, we find LIME and RAWE [M. Palmirani et al. 2013], which are two web editors (JavaScript) capable of semi-automatically marking up the text in Akoma Ntoso and the manually formalized norms in LegalRuleML. PrOnto [Monica Palmirani, Martoni, et al. 2018] is a legal ontology for*

[4] A similar quote can be found in [Monica Palmirani, Contissa, et al. 2009].

modelling GDPR concepts and axioms. It feeds concepts and predicates to the legal rule-modelling layer in order to make the formalization consistent and harmonized. Regorous [Governatori 2015] is a tool (written in Java) that makes it possible to design BPMN 2.0 and to connect each step of the process with the legal rules. Regorous provides an API to SPINdle [Lam and Governatori 2009], a defeasible legal reasoning engine. Regorous presents at the end the results of compliance checking in a user interface for the end user' [Monica Palmirani and Governatori 2018]. The problem with this approach is that in the process of creating this integrated platform for modelling GDPR and its evolutions in time to enforce it automatically, the authors have ended up at reimplementing a whole dedicated software engineering toolchain relying on *ad-hoc* tools, that do not interoperate with standard software engineering tools. The ontology inference engine used to actually execute the LegalRuleML annotations may entail a significant reconfiguration of an existing IT system in order to be used in real-world applications. The LegalRuleML rules are themselves a *de facto* form of source code that can only be edited and viewed through custom editors and visualizers, that may not support all the features of modern standard code editors. For instance, there is no version control system for this source code (though the rules themselves track code of the versions of the law). More generally, the *ad-hoc* tools are not at the state of the art of the things they want to replace. Ironically, the ontology architecture made for aligning legal standards goes against the Unix philosophy that underlies a lot of modern software engineering best practices: '*Write programs that do one thing and do it well. Write programs to work together. Write programs to handle text streams, because that is a universal interface*' [Raymond 2003]. Simply put, integrating the toolchain of Monica Palmirani and Governatori [2018] in an existing IT system is highly disruptive technically and its ability to scale up is not yet demonstrated.

But this paradox is not surprising, as all big ontology-making endeavours are doomed to become more and more integrated and monolith-like as they approach actual real-world usage. This was already mentioned by Breuker et al. [2004]: '*An interesting problem that arises is the introduction of an inference bias. Valente et al. [1999] show that ontological choices are strongly influenced by the purpose of the ontology. That is, the same knowledge will be structured or formalized differently depending of how it will be used by the reasoner in reaching the desired conclusions in a specific context. This indicates that reusability is a good idea, but it can never be accomplished completely*'. This critic is formulated even clearer by Brewster and O'Hara [2004]. I add here two points to the argument. First, that the inference bias also translates to a technological bias for ontologies as software artifacts that tend to build separate ecosystems from the rest of software engineering, preventing them from harvesting the scale-up benefits coming from the huge infrastructure investments made in standardized software engineering tooling. Second, that the technological bias leads to a bias in the users of the technological tools: rather than being usable directly by either lawyers or programmers, the tools require both lawyers and programmers to learn new concepts (logic programming, ontologies, defeasability, deontic logic, etc.) before they can put the tools to use. The risk is that these artifacts meant as reusable and accessible models, create a third class of model-makers, distinct from lawyers and programmers, that act as an intermediary and barrier that controls how the worlds of CS and Law interact with each other.

These developments strike a heavy blow to the dream of machine-consumable and infinitely reusable legal modelling that would conform to the shared utopia of legal formalism and cybernetics. Consequently, the third decade of the AI & Law retrospective is dominated by machine learning approaches. These aim to directly address a specific task without going through the intermediate step of making a general, reusable model of the relevant law first. However, the hopes have been recently revived by the latest developments in AI & Law's parent field, namely generative Large Language Models (LLMs) such as ChatGPT. Spearheading this research effort is Stanford CodeX's Doulcet [2023] who has tried to use ChatGPT to directly generate a formal and reusable model of the law from the legal text. However, he quickly realized that the reusable model he created that way was less efficient than using ChatGPT to translate legal texts directly to specific pieces of executable code: '*[...] once you can automatically translate any piece of legislation/contract into code, the most interoperable format is the words themselves. And as the cost of building a code representation of a legal text goes down, you care less about reusability of your legislation model for* *multiple use cases. [...] Instead we need to build systems that automatically transform legal text into some code in a popular programming language that solves the task at hand. Using popular programming language, it is easier to let the LLM generate the code (it likely already knows Python or Typescript), and we get better tooling!*'. According to Doulcet's observations, technology may have gone full circle and in the end, the correct way of modelling the legal text is... the legal text itself. More pragmatically, tenants of logic programming and ontologies are currently trying to integrate LLMs into their modelling processes, in order to compensate for the limits of their technology of choice and try and finally achieve the dream of machine-consumable and infinitely reusable legal modelling. But is this dream really desirable? Is it worth spending another decade of collective work trying to achieve it at the expense of other goals for the research at the intersection of CS and Law? Here, it is useful to recall the analysis by Leith [2016] about the fall of the movement of legal expert systems powered by logic programming in the 1990's: '*why was there optimism, was there ever any success, and–if as I suggest–there was none, then why was [there] such a huge extravaganza of funding for expert systems research in a field (legal technology) which has been practically starved of funding in all the other decades outwith the 1980's? I tried to answer these questions in my* Formalism in AI and Computer Science *[Leith 1990], suggesting that the focus on the machine rather than the user had led technicians into fields which they little understood, and I still believe that was the underlying reason for the decade. [...] the AI community now rewrite their AI projects to suit funders who are less keen on the AI approach - XML technology being one such funding source*'.

Wanting it All is a Counterproductive Approach

In what follows, I will try to generalize some arguments from the critique of the technical systems made in the previous section, about why the dream of creating machine-consumable and infinitely reusable legal models is counterproductive. These arguments are not philosophical, nor are they related to how the Rule of Law should be respected or not. They are instead very technology-oriented and stem from the experience of engineering and applied science best practices. By providing these new ideas to the debate, I hope to reframe the current antagonism of bold disruptive technologist *vs.* conservative and scientifically illiterate lawyers with the more nuanced proposition that good engineering and legal technological work will benefit from detaching oneself from the ideologically-imposed goals and methods discussed in the previous sections. First, a general-purpose legal model can never be precise enough to cover the needs of critical applications. Second, infinitely reusable legal modelling is an unscoped system that is impossible to evaluate correctly, thus evading any rigorous scientific process. Third, the quest for the all-encompassing model requires a lot of infrastructure work upfront before any benefits can be reaped, which also means wasting a lot of time if the approach ends up not working.

My first argument builds on the notion of inferencing bias of Valente et al. [1999]: '*knowledge is usually modeled with certain types of inferences in mind. For example, if we expect to use the Loom classifier to infer whether or not two intervals meet (that is,* `(meets int1 int2)`*), we need to add enough information in the definition of the relation* `meets` *to enable the classifier to use it. If, however, we only want to assert that the intervals meet and use this information for other inferences, it is enough to state the range and domain of the meets relation*'. Similarly, real-world applications of AI & Law need to cope with number representations, rounding errors, data structure layout, error handling, null data, etc. All of these actually change the result returned to the user, so the question becomes: what is the level of precision we expect from AI & Law tools? I would argue that the killer applications of AI & Law would be expected to be the ones where the stakes are high and we would need computer precision to compensate for human deficiencies. But an all-encompassing, reusable model is more likely to spread out its precision and accuracy over its diverse uses. Minimizing the precision problem saying that the tool is only here to 'assist' is not satisfactory in my opinion: a tool that requires a lot of infrastructure and time-commitment to get answers that are often wrong is not convivial in the sense of Illich and Lang [1973].

The second argument points out the impossibility to correctly evaluate the usefulness or performance of a general, infinitely-reusable model (of law or of something else). This is a broader problem for the field of AI. The evaluation problem in the field of AI & Law is correctly pointed out by Novotná and Libal [2022]: '*In [Cohen and Howe 1988], the authors state that the evaluation of the experiments and the methods "expedites the understanding of available methods and so their integration into further research". The authors in [Conrad and Zeleznikow 2015] argue, that "a performance-based ethic signifies a level of maturity and scientific rigor within a community". However, the meta-analysis of research studies in the field of artificial intelligence and law in [Conrad and Zeleznikow 2015, 2013; Hall and Zeleznikow 2001] shows that great part of studies does not contain any kind of evaluation whatsoever*'. Machine learning researchers have addressed the evaluation problem by building shared benchmarks and evaluation metrics like Imagenet [Deng et al. 2009]. These benchmarks have biases and may lead to overfitting models, but they are a necessary step for advancing the field. Works by Holzenberger et al. [2020] or Guha et al. [2023] are starting to fill that gap but the practice should extend to non-machine-learning-based AI & Law too. Otherwise, the field's quest will be similar to the unscoped quest for AGI described by Gebru and Torres [2023].

The third argument concerns the very prospective nature of making general models that we hope can be reused afterwards. AI & Law projects that start without precisely (or properly) identifying who the developers and the users of their tools are, will very likely miss any target they later set for themselves. The technical choices should follow the needs of the users and not vice-versa. By choosing architectures and objectives with only the dreams of legal formalism and cybernetics in mind, the field will continue hitting hard barriers for adoption by institutions and companies that operate under different ideologies and conceptions of legal practice. While the theoretical foundations of AI & Law have been deeply studied, the applied research branch of the field is missing killer applications and adoption. In applied research, it is not sufficient to 'propose an innovative architecture' or 'offer some ideas on whether certain techniques can help users'; the goal is to have technical solutions battle-tested and some of them ultimately adopted to become industry standards.

Conclusion: Bring Back the Scientific Method

I come from a research field – formal methods – whose relationship with applied research and practical applications is complicated [MacKenzie 2004]. Formal methods, coincidentally, is also a scientific discipline born out of the early field of AI. But today, formal methods have dropped their affiliation to AI and have resolutely chosen concrete areas of applications: critical software, embedded systems, model checking, etc. Some tools from formal methods have become industry standards in railroads, avionics, the nuclear sector, etc. At the same time, the theoretical activity in formal methods is still significant, and we're starting too see thirty-year-old theoretical foundations (such as linear logic by Girard [1987]) being used as the basis for popular programming languages (such as Rust). What I retain from my experience in formal methods is the healthy divide between the evaluation criteria for theoretical *vs.* applied research, even when faced with contributions that are basically formalisms or models. If your contribution is theoretical, then you must show how it is more expressive, concise, elegant, etc. than related work. If your contribution is applied, then you must show how it is more performant, practical, adopted, etc. than related work. Negative results can be contributions but only if the paper explains why the results are negative and what we can learn from them (other than that something does not work). In short, follow the scientific method! Deviating from it threatens the ability of the field to weed out unproductive approaches and allocate its research ressources optimally. And if funding bodies continue to be fascinated by legal formalism, cybernetics or the tantalizing promise of government by machines, then it is up to us, as peers in our scholarly venues, and not as cogs of a system caught up in a feedback loop, to judge the value of contributions, with technical and legal sharpness alike.

Acknolwedgments

I want to thank the many persons, especially in the Catala research group, that helped me refine my argumentation and proof-read this paper. Also, I am in debt with the COHUBICOL team for welcoming me in their internal seminar, where I presented what would be the premises of this paper.

References

Layman E Allen. 1956. 'Symbolic logic: A razor-edged tool for drafting and interpreting legal documents'. *Yale LJ*, 66, 833.

Joaquín Arias, Mar Moreno-Rebato, Jose A Rodriguez-García, and Sascha Ossowski. 2021. 'Modeling administrative discretion using goal-directed answer set programming'. In: *Conference of the Spanish Association for Artificial Intelligence*. Springer, 258–267.

Tara Athan, Guido Governatori, Monica Palmirani, Adrian Paschke, and Adam Wyner. 2015. 'LegalRuleML: Design principles and foundations'. In: *Reasoning Web International Summer School*. Springer, 151–188.

Gioele Barabucci, Luca Cervone, Angelo Di Iorio, Monica Palmirani, Silvio Peroni, and Fabio Vitali. 2010. 'Managing semantics in XML vocabularies: an experience in the legal and legislative domain'. In: *Proceedings of Balisage: The markup conference*. Vol. 5.

Danièle Bourcier. 2017. *Artificial intelligence and law: what perspective?* Presentation at the Congress of the World Organisation of Systems and Cybernetics. (2017).

Joost Breuker, André Valente, and Radboud Winkels. 2004. 'Legal ontologies in knowledge engineering and information management'. *Artificial intelligence and law*, 12, 241–277.

Christopher Brewster and Kieron O'Hara. 2004. 'Knowledge representation with ontologies: the present and future'. *IEEE Intelligent Systems*, 19, 1, 72–81.

Pierre Catala, Lucien Mehl, and Edmond Bertrand. 1974. *Constitution et exploitation informatique d'un ensemble documentaire en droit (droit de l'urbanisme et de de la construction)*. Ed. du CNRS.

Paul R Cohen and Adele E Howe. 1988. 'How evaluation guides AI research: The message still counts more than the medium'. *AI magazine*, 9, 4, 35–35.

Alain Colmerauer and Philippe Roussel. 1996. 'The birth of Prolog'. In: *History of programming languages—II*, 331–367.

Jack G. Conrad and John Zeleznikow. 2015. 'The Role of Evaluation in AI and Law: An Examination of Its Different Forms in the AI and Law Journal'. In: *Proceedings of the 15th International Conference on Artificial Intelligence and Law* (ICAIL '15). Association for Computing Machinery, San Diego, California, 181–186.

Jack G. Conrad and John Zeleznikow. 2013. 'The Significance of Evaluation in AI and Law: A Case Study Re-Examining ICAIL Proceedings'. In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law* (ICAIL '13). Association for Computing Machinery, Rome, Italy, 186–191.

Giuseppe Contissa, Francesco Godano, and Giovanni Sartor. 2021. 'Computation, cybernetics and the law at the origins of legal informatics'. *Italian Philosophy of Technology: Socio-Cultural, Legal, Scientific and Aesthetic Perspectives on Technology*, 91–110.

Silvia Crafa, Cosimo Laneve, Giovanni Sartor, and Adele Veschetti. 2023. 'Pacta sunt servanda: legal contracts in Stipula'. *Science of Computer Programming*, 225, 102911.

Anthony D'Amato. 1977. 'Can/should computers replace judges?' *Georgia Law Review*, 11, 11–36.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. 'Imagenet: A large-scale hierarchical image database'. In: *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 248–255.

John Dewey. 1924. 'Logical method and law'. *Cornell LQ*, 10, 17.

L. Diver et al.. 2022. *Typology of Legal Technologies*. COHUBICOL. (2022).

Pierre-Loïc Doulcet. 2023. *Computational Law, AI status update*. https://twitter.com/hexapode/status/1691386964170207232. (2023).

Vittorio Frosini. 1968. *Cibernetica, Diritto e società*. Ed. by Edizione di Comunità. Milano.

Vittorio Frosini. 1973. 'L'automazione elettronica nella giurisprudenza e nell'amministrazione pubblica'. *Bollettino bibliografico d'informatica generale e applicata al diritto*, 2, 3-4, 101–104.

Timnit Gebru and Émile Torres. 2023. *Eugenics and the Promise of Utopia through Artificial General Intelligence*. Keynote presentation at the SaTML conference. (2023).

Michael Genesereth. 2015. *Computational Law: The Cop in the Backseat*. Stanford CodeX Center. (2015).

Jean-Yves Girard. 1987. 'Linear logic'. *Theoretical computer science*, 50, 1, 1–101.

Guido Governatori. 2015. 'The Regorous approach to process compliance'. In: *2015 IEEE 19th International Enterprise Distributed Object Computing Workshop*. IEEE, 33–40.

Guido Governatori, Trevor JM Bench-Capon, Bart Verheij, Michał Araszkiewicz, Enrico Francesconi, and Matthias Grabmair. Dec. 2022. 'Thirty years of Artificial Intelligence and Law: the first decade'. *Artificial Intelligence and Law*, 30, 4, (Dec. 2022), 481–519.

Neel Guha et al.. 2023. *LegalBench: A Collaboratively Built Benchmark for Measuring Legal Reasoning in Large Language Models*. (2023). arXiv: `2308.11462 [cs.CL]`.

Maria Jean J Hall and John Zeleznikow. 2001. 'Acknowledging insufficiency in the evaluation of legal knowledge-based systems: strategies towards a broadbased evaluation model'. In: *Proceedings of the 8th international conference on Artificial intelligence and law*, 147–156.

Mireille Hildebrandt. Nov. 2020. 'Code-driven Law: Freezing the Future and Scaling the Past'. en. In: *Is Law Computable? : Critical Perspectives on Law and Artificial Intelligence*. Ed. by Simon Deakin and Christopher Markou. Hart Publishing, Oxford, (Nov. 2020), 67–84.

Nils Holzenberger, Andrew Blair-Stanek, and Benjamin Van Durme. 2020. 'A Dataset for Statutory Reasoning in Tax Law Entailment and Question Answering'. *arXiv preprint arXiv:2005.05257*.

Ivan Illich and Anne Lang. 1973. 'Tools for conviviality'.

Viktor Knapp and Vladimir Vrecion. 1970. 'The Possibility of applying cybernetic methods to law and administration'. *International social science journal*, XXII, 3.

Philipp Körner et al.. 2022. 'Fifty years of Prolog and beyond'. *Theory and Practice of Logic Programming*, 22, 6, 776–858.

Ho-Pun Lam and Guido Governatori. 2009. 'The making of SPINdle'. In: *Rule Interchange and Applications: International Symposium, RuleML 2009, Las Vegas, Nevada, USA, November 5-7, 2009. Proceedings 3*. Springer, 315–322.

Philip Leith. 1990. *Formalism in AI and computer science*. Ellis Horwood.

Philip Leith. 2016. 'The rise and fall of the legal expert system'. *International Review of Law, Computers & Technology*, 30, 3, 94–106.

How Khang Lim, Avishkar Mahajar, Martin Strecker, and Meng Weng Wong. 2022. 'Automating defeasible reasoning in law with answer set programming'.

Donald MacKenzie. 2004. *Mechanizing proof: computing, risk, and trust.* MIT Press.

Masha Medvedeva, Martijn Wieling, and Michel Vols. 2023. 'Rethinking the field of automatic prediction of court decisions'. *Artificial Intelligence and Law*, 31, 1, 195–212.

Lucien Mehl. 1959. 'Automation in the legal world: from the machine processing of legal information to the "Law Machine"'. In: *Mechanisation of Thought Processes: Proceedings of a Symposium Held at the National Physical Laboratory on 24th, 25th, 26th and 27th November 1958*. Vol. II, Her Majesty's Stationery Office, London.

Lucien Mehl. 1957. 'La cybernetique et l'administration'. *La Revue administrative*, Juillet-Août, 58.

Lucien Mehl and Jean-Marie Breton. 1970. 'L'automatisation de la recherche de l'information juridique par le procédé Docilis : Documents, interrogations libres'.

Denis Merigoux. Dec. 2021. 'Proof-oriented domain-specific language design for high-assurance software'. Theses. Université Paris sciences et lettres, (Dec. 2021).

Tereza Novotná and Tomer Libal. 2022. 'An evaluation of methodologies for legal formalization'. In: *International Workshop on Explainable, Transparent Autonomous Agents and Multi-Agent Systems*. Springer, 189–203.

OECD. 2019. *Embracing Innovation in Government: Global Trends 2019*. Tech. rep. Observatory for Public Sector Innovation.

Anja Oskamp and Marc Lauritsen. Dec. 2002. 'AI in Law Practice? So far, not much'. *Artificial Intelligence and Law*, 10, 4, (Dec. 2002), 227–236.

M. Palmirani, L. Cervone, O. Bujor, and M. Chiappetta. 2013. 'RAWE: an editor for rule markup of legal texts'. In: *CEUR Workshop Proceedings, Seattle, USA, 11–13 July 2013*. Ed. by P. Fodor, D. Roman, D. Anicic, A. Wyner, M. Palmirani, D. Sottara, and F. Lévy.

Monica Palmirani, Giuseppe Contissa, and Rossella Rubino. 2009. 'Fill the gap in the legal knowledge modelling'. In: *Rule Interchange and Applications: International Symposium, RuleML 2009, Las Vegas, Nevada, USA, November 5-7, 2009. Proceedings 3*. Springer, 305–314.

Monica Palmirani and Guido Governatori. 2018. 'Modelling Legal Knowledge for GDPR Compliance Checking'. In: *JURIX*. Vol. 313, 101–110.

Monica Palmirani, Guido Governatori, Antonino Rotolo, Said Tabet, Harold Boley, and Adrian Paschke. 2011. 'LegalRuleML: XML-based rules and norms'. In: *International Workshop on Rules and Rule Markup Languages for the Semantic Web*. Springer, 298–312.

Monica Palmirani, Michele Martoni, Arianna Rossi, Cesare Bartolini, and Livio Robaldo. 2018. 'Pronto: Privacy ontology for legal reasoning'. In: *Electronic Government and the Information Systems Perspective: 7th International Conference, EGOVIS 2018, Regensburg, Germany, September 3–5, 2018, Proceedings 7*. Springer, 139–152.

Monica Palmirani and Fabio Vitali. 2011. 'Akoma-Ntoso for legal documents'. *Legislative XML for the Semantic Web: Principles, Models, Standards for Document Management*, 75–100.

Silvio Peroni, Monica Palmirani, and Fabio Vitali. 2017. 'UNDO: The United Nations system document ontology'. In: *The Semantic Web–ISWC 2017: 16th International Semantic Web Conference, Vienna, Austria, October 21-25, 2017, Proceedings, Part II 16*. Springer, 175–183.

Ursula Plesner and Lise Justesen. 2022. 'The Double Darkness of Digitalization: Shaping Digital-ready Legislation to Reshape the Conditions for Public-sector Digitalization'. *Science, Technology, & Human Values*, 47, 1, 146–173. eprint: `https://doi.org/10.1177/0162243921999715`.

Joseph Potvin. 2023. 'Data with direction: design research leading to a system specification for 'an internet of rules''. PhD thesis. Université du Québec en Outaouais.

Joseph Potvin et al.. 2021. 'Oughtomation: Practical Normative Data'.

Eric S Raymond. 2003. *The art of Unix programming*. Addison-Wesley Professional.

Edwina L Rissland, Kevin D Ashley, and Ronald Prescott Loui. 2003. 'AI and Law: A fruitful synergy'. *Artificial Intelligence*, 150, 1-2, 1–15.

Giovanni Sartor et al.. Dec. 2022. 'Thirty years of Artificial Intelligence and Law: the second decade'. *Artificial Intelligence and Law*, 30, 4, (Dec. 2022), 521–557.

James C. Scott. 1998. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press.

M. J. Sergot, A. S. Kamble, and K. K. Bajaj. 1991. 'Indian Central Civil Service Pension Rules: A Case Study in Logic Programming Applied to Regulations'. In: *Proceedings of the 3rd International Conference on Artificial Intelligence and Law* (ICAIL '91). Association for Computing Machinery, Oxford, England, 118–127.

M. J. Sergot, F. Sadri, R. A. Kowalski, F. Kriwaczek, P. Hammond, and H. T. Cory. May 1986. 'The British Nationality Act As a Logic Program'. *Commun. ACM*, 29, 5, (May 1986), 370–386.

Giovanni Sileno. 2016. 'Aligning Law and Action'. PhD thesis. University of Amsterdam.

Didier Torny. 2005. 'L'administration sanitaire entre contraintes techniques et contraintes juridiques. L'exemple des maladies émergentes'. *Revue générale de droit médical*. La protection de la santé publique 6.

André Valente, Thomas Russ, Robert MacGregor, and William Swartout. 1999. 'Building and (re) using an ontology of air campaign planning'. *IEEE Intelligent Systems and their Applications*, 14, 1, 27–36.

Serena Villata, Michal Araszkiewicz, Kevin Ashley, Trevor Bench-Capon, L. Karl Branting, Jack G. Conrad, and Adam Wyner. Dec. 2022. 'Thirty years of artificial intelligence and law: the third decade'. *Artificial Intelligence and Law*, 30, 4, (Dec. 2022), 561–591.

Joseph Weizenbaum. 1976. 'Computer power and human reason: From judgment to calculation.'

Jan Wielemaker, Tom Schrijvers, Markus Triska, and Torbjörn Lager. 2012. 'Swi-prolog'. *Theory and Practice of Logic Programming*, 12, 1-2, 67–96.