NETWORK RESEARCH

Responsible automatically processable regulation

Clement Guitton[1] · Simon Mayer[1] · Aurelia Tamo-Larrieux[2] · Dimitri Van Landuyt[3] · Eduard Fosch-Villaronga[4] · Irene Kamara[5] · Przemysław Pałka[6]

Abstract
Driven by the increasing availability and deployment of ubiquitous computing technologies across our private and professional lives, implementations of automatically processable regulation (APR) have evolved over the past decade from academic projects to real-world implementations by states and companies. There are now pressing issues that such encoded regulation brings about for citizens and society, and strategies to mitigate these issues are required. However, *comprehensive yet practically operationalizable* frameworks to navigate the complex interactions and evaluate the risks of projects that implement APR are not available today. In this paper, and based on related work as well as our own experiences, we propose a framework to support the conceptualization, implementation, and application of *responsible* APR. Our contribution is twofold: we provide a holistic characterization of what responsible APR means; and we provide support to operationalize this in concrete projects, in the form of leading questions, examples, and mitigation strategies. We thereby provide a scientifically backed yet practically applicable way to guide researchers, sponsors, implementers, and regulators toward better outcomes of APR for users and society.

Keywords Encoding regulation · Principles · Automatically processable regulation · Responsible · Systematization of knowledge · Fairness

✉ Clement Guitton
clement.guitton@unisg.ch

Simon Mayer
simon.mayer@unisg.ch

Aurelia Tamo-Larrieux
a.tamo@maastrichtuniversity.nl

Dimitri Van Landuyt
dimitri.vanlanduyt@kuleuven.be

Eduard Fosch-Villaronga
e.fosch.villaronga@law.leidenuniv.nl

Irene Kamara
I.Kamara@tilburguniversity.edu

Przemysław Pałka
przemyslaw1.palka@uj.edu.pl

1   University of St. Gallen, St. Gallen, Switzerland

2   Maastricht University, Maastricht, The Netherlands

3   KU Leuven, Leuven, Belgium

4   Leiden University, Leiden, The Netherlands

5   Tilburg University, Tilburg, The Netherlands

6   Jagiellonian University, Kraków, Poland

1 Introduction

The push for automation has not spared the legal domain, and new breakthroughs in generative artificial intelligence (AI) have also led to excitement among legal researchers; while earlier large language models like GPT-3.5 performed below human average on the *US Uniform Bar Examination* (Bommarito and Katz 2023), current models (such as GPT-4) are able to pass the bar exam and outperform human test-takers (Katz et al. 2023). While such developments have filled headlines, researchers at the intersection of law and technology have already for more than thirty years been seeking to automate the law, with a user-base almost exclusively within academia (Sartor and Branting 1998; Palmirani et al. 2011; Ashley 2017). In recent years, however, we have observed a shift toward different implementations by private and public entities (Guitton et al. 2022a), contributing to the expectation that *automatically processable regulation* (APR) will increasingly permeate people's everyday lives.

APR refers to regulation that is expressed in a form that makes it accessible to be processed automatically and where we see a clear intention of doing so (Guitton et al. 2022a).

APR can seek *efficiency gains*, for instance by legal professionals through automatic summarizing of legal documents (Kanapala and Pamula 2019) or the more problematic risk assessment instruments (Dass et al. 2023; Lagioia et al. 2023), but it also encompasses automated tools that make the law *more accessible* for laypeople, for instance through legal question-answering (Mohun and Roberts 2020a). The field today also includes legal predictions (e.g., on recidivism) (Wang et al. 2023), or the public release of automatic social benefits calculators that are based on different laws (LabPlus 2018; McNaughton 2020; Alauzen 2021; Diver et al. 2022; Guitton et al. 2022a give an in-depth account of these applications). APR, furthermore, does not merely apply to software and services, but also to hardware: Researchers have started to experiment with cyber-physical systems whose program code is directly linked to legal provisions and that are hence able to react to changes in their legal context, for instance in the case of manufacturing robots and their associated industry and safety standards (Hood et al. 2001; Black et al. 2002; Shafei et al. 2018; García et al. 2021), and corresponding ethical approaches have also emerged (Anderson and Fort 2023).

Because of this strong increase in APR applications and since many of the implications of the application of APR remain poorly understood, it has become urgent to clearly identify and alleviate issues which the deployment of APR projects in the real-world begets. Specifically, those involved in the creation of APR need to be able to elaborate the implications of APR implementations *on individuals and on society*. They also need to be able to create solutions tailored to the different nuances that each project brings, which includes an understanding of the project *goals*, of the involvement of different *stakeholders*, of the risks of different used *technologies*, and of how *processes* for oversight, contestability, and transparency are set up.

In this context, we present an *operationalizable* framework that structures the typical issues that should be considered when engaging in an APR project. Our proposed framework draws on state-of-the-art approaches, dozens of interviews with practitioners, and own experience of creating APR; it also draws on frameworks on responsible AI, as well as earlier work focused on identifying the issues triggered by the uptake of APR (Guitton et al. 2022a). Given the expected pervasiveness of APR, our framework has been created to enable a broad range of stakeholders in APR projects—researchers, sponsors, implementers, and regulators—to evaluate a specific APR project from the different relevant viewpoints, including issues at the heart of the considered *legal problem* (e.g., whether it has been considered how the system deals with vagueness), issues that affect *individual users* (e.g., what are possible adverse psychological effects of a system that makes legal decisions in a split second), and issues that affect *society* at large (e.g.,

whether the project's assumptions should be revisited upon cultural evolutions).

In the following Sect. 2, we provide a review of the current state of research, to which we contribute with a user-centered, comprehensive, and practically applicable approach for evaluating risks and responsibility in the adoption of APR. In Sect. 3, we present a review of 13 issues which APR triggers, as well as both lead questions to identify whether a project faces the issue and mitigation strategies to follow when it does. Finally, in Sect. 4, we sketch a path for future work.

2 Current state of research

With advances in AI leveraged in the legal domain (Chalkidis et al. 2020; Bench-Capon 2022), the field of APR has undergone a rapid transformation from early attempts in the 1980s to map the National British Act onto a decision tree (Sergot et al. 1986). In fact, the history of APR is characterized by three distinct waves (Bench-Capon 2022): The first one focused on the creation of logical models of legal knowledge, the second one focused on the creation of legal ontologies, and the third and current wave concentrates on the adoption of machine learning and generative AI approaches. Each of these three waves has given rise to specific issues that are linked to the technology deployed to automate legal processes. Within the first wave, these were problems related to correctly and equivalently transcribing law into mathematical models and technical challenges regarding the limits of logic and reasoning (Sergot et al. 1986). In the second wave, a central challenge has been to develop APR according to different vocabularies that are interoperable and harmonized with one another. A range of ontologies (e.g., the Data Privacy Vocabulary[1]) has since been proposed to remedy this issue and to enable interoperability and, with it, interdisciplinarity across legal domains (Mário et al. 2019). The third wave is currently in full swing, and has recently been boosted with the introduction of powerful generative language models that can for instance summarize legal texts (Bauer et al. 2023) and support the interpretation or classification of legal rules (Liga and Robaldo 2023; Bommarito and Katz 2023).

As a consequence of the close correspondence of APR developments with the symbolic AI field and, especially since the third wave, also with subsymbolic AI, many of the user-centric and societal issues that APR projects face today overlap with ethical issues that have been identified for AI projects more broadly (Fjeld et al. 2020; Loi 2020; Giovanola and Tiribelli 2023). Importantly, the ethics guidelines for trustworthy AI by the High-Level Expert Group of the

---

[1] https://w3c.github.io/dpv/dpv/.

EU brought forward seven key requirements for AI systems to be deemed trustworthy (HLEG 2019).[2] These principles center around human agency (a focus also with its criticism for being narrow, Baum and Owe 2023) and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and environmental well-being, and accountability. Other organizations, such as the Alan Turing Institute (Leslie 2019) as well as researchers (Morley et al. 2020), Floridi and Cowls (2021) provide similar guidance. While these approaches are not necessarily focused on APR projects specifically, the provisions are addressing concerns that are similar to those that have been published regarding the consequences of APR projects (Guitton et al. 2022a). Yet, at this stage, it remains challenging to operationalize existing AI guidelines: A recent meta-review of 106 AI frameworks, criteria, metrics, and checklists shows that the translation of the principles into clear guiding questions and mitigation strategies is still lacking (Prem 2023); it is specifically difficult to extract concrete operationalization *for APR projects*.

Still, APR may draw upon frameworks that are issued by governments on the implementation of automated decision-making systems—including those that apply machine learning and those that do not—that are implemented in the public sector. In the Netherlands for instance, the Ministry of the Interior jointly with a local university developed an instrument to assess impacts of automated decision-making algorithms from a human-rights perspective (Utrecht University 2021) by providing open-ended questions that seek to enable a comprehensive and inclusive discussion among stakeholders. Similarly, the UK government—based on joint work by the Central Digital and Data Office, the Cabinet Office, and the Office for Artificial Intelligence—developed a framework to ensure that automated decision-making systems in the public sector take the needs and interests of *citizens* as a priority (UK Government 2021). Their framework centers around understanding the impacts of such systems, clear responsibility and accountability rules, and ensuring that citizens comprehend the impact of such systems on their lives and rights. While these are only two approaches out of many government-led initiatives within Europe and beyond (Prem 2023), we see a stronger focus on citizens and citizen rights emerging: Research has provided us with methodologies to assess the extent to which AI accommodates human rights (Mantelero and Esposito 2021) and we have recently seen a push toward a stronger inclusion of human rights in AI regulation, for instance by the Office of the United Nations High Commissioner for Human Rights.[3]

Such approaches, hence, do not only stem from the EU, but we see global concerns about the citizen-centered, human-centered, and societal-centered development of technologies that impact our daily lives. Such debates are important because several trade-offs required for APR are the very same ones that are highly relevant in democratic societies (Hoffmann-Riem 2022). For instance, *encoding the law* can either use (intransparent) deep neural networks (Nguyen et al. 2022), or it can use a higher level of human mediation by manually translating legal text into controlled language before using this within APR systems (Höfler and Bünzli 2010; Kowalski et al. 2023) that may in this way feature increased transparency and traceability of decisions. These trade-offs highlight the requirement of making debated, precise, logical, and then published decisions for setting up thresholds of what is desired and accepted, and what is not.

Summarizing the available research and published policies, APR issues overlap with those of AI and of automated decision-making more broadly, but comprehensive yet operationalizable and at the same time citizen-centered guidance for projects falling within the APR domain is lacking. Even within the field of AI where a wealth of research has emerged on the topic of responsibility (Andrada et al. 2023), scholars have recently noted that future research will still "need to grapple with questions of fairness, transparency, accountability", including "the use of AI in criminal justice" (Trotta et al. 2023). In this paper, we address this gap by responding to some of the criticism on the lack of frameworks whose individual concepts can be easily operationalized for APR: We present a framework that is easy to follow (e.g., with questions similar to Utrecht University (2021), UK Government (2021)) and that leaves open choices depending on exact circumstances, yet remains specific and concrete enough to be efficiently usable (e.g., through the provision of mitigation strategies for identified issues). Our framework does not seek to be overly prescriptive, but rather to make implementers aware of possible issues and solutions they might want to consider. Finally, it specifically and actively reflects user- and societally oriented issues of APR projects to permit these to be explored explicitly and not be taken lightly in the form of low-level implementation choices.

3 Responsible APR framework

The list of issues that we address with our framework is based on Guitton et al. (2022a), Guitton et al. (2022b). Guitton et al. (2022a) reviewed the existing literature to create a comprehensive list of APR projects' issues. To demonstrate the viability of this list, Guitton et al. (2022a) applied it to ten real-life cases by gaining insights into these cases through interviews with stakeholders involved in the

---

[2] For a more comprehensive review of factors impacting trust in automation, see Tamò-Larrieux et al. (2023).

[3] See https://www.ohchr.org/en/statements/2023/07/artificial-intelligence-must-be-grounded-human-rights-says-high-commissioner.

implementation, thereby discerning whether the issue was actually relevant in the project or not (Guitton et al. 2022b). In this paper, we adopt the same list of issues and developed questions that will permit stakeholders who are involved in APR projects (as well as outside observers) to explore each issue in a structured way and for a specific project. In this process, we observed that for certain issues, these investigation processes converged, and we, hence, decided to merge the underlying issues. In addition to processing the original list of issues in this way and extending it with specific lead questions, we focused on providing practical mitigation support that can be adopted in concrete projects, and correlating each issue with similar challenges in AI projects.

Table 1 provides an overview of the issues that might undermine the responsible conceptualization, implementation, and usage of APR—and that may be discovered, analyzed, and mitigated through our framework. We have grouped the issues into 13 types and elaborate on each of them in the following sections. To assess the risk of a certain issue arising, we propose leading questions (in Table 1) that should be evaluated during the conceptualization of an APR project, to determine whether and if so how to continue the project. In addition, we provide concrete examples of the discussed issues and how the leading questions enable an informed debate about the APR projects. Across all issues, we argue that a re-evaluation of projects during their life-cycle at regular intervals is central and needs to include post-deployment milestones. In addition, we propose that stakeholders who employ our framework commit themselves to publishing an evaluation about its use, what issues had been correctly identified ex ante and before the APR project was deployed, what aspects had been changed, and what aspects were overlooked (and why).

3.1 Vagueness and balancing of interests

Laws are often formulated in a way that is inherently vague (Endicott 2011) or requires interpretation. This is in part due to the need to remain broadly applicable (Endicott 2011), where courts are called upon to interpret the law and adopt different interpretations (Moses 2020). While natural language may stay vague and is open to different interpretations that require balancing conflicts of interests such as fairness and legal certainty (Radbruch 2006; Hart 2021; Moore 2020), APR projects cannot generally cope with such vagueness, abstractions, and contextualization. Unchecked vagueness when implementing an APR project is however highly prone to lead to ad hoc interpretations by those who create the code that underpins an APR project.

Example Many traffic laws permit overtaking under certain conditions, one of which is commonly that there needs to be "a suitable gap in front of the road user you plan to overtake".[4] While traffic laws typically do not specify technical thresholds (e.g., regarding the specific size of this gap depending on the speed, maximum acceleration, and weather conditions), an APR version of this statement that can be interpreted and utilized fully automatically is required to define such characteristics. Those involved in the creation of the APR project need to be aware of this specific interpretation early in the process since this might readily render an APR project impossible (van Dijck et al. 2023; Emanuilov et al. 2018); further, if deemed feasible, the APR project needs to use an interpretation that remains within the spirit of the law, and the interpretation should be documented appropriately.

Mitigation strategy: We propose that—following the discussion of the given lead questions (see Table 1)—the target regulation undergoes explicit pre-processing with respect to vagueness by several individuals with differing backgrounds and roles in the project. This pre-processing should be carried out independently by these individuals, and they should mark the relevant parts of the regulation and write down their own interpretation; subsequently, the identified vague aspects should be integrated and a common interpretation should be created. For this common interpretation, we propose the use of a controlled natural language—this has the benefit that the resulting (controlled) text is understandable by legal experts as well as by laypeople and at the same time reduces the barrier to implementation of the regulation in a computer system.[5] In this process, it may also occur that no single interpretation can be agreed upon; in this case, we propose to explicit document this and, if the project is still being continued, that the implementation reflects all suggested interpretations. This creates the possibility to (a) evaluate the different interpretations in concrete run-time instances of the problem on the final APR system; and to (b) create transparency in the final system (e.g., by adding, to the user interface, a note that an alternative interpretation is possible). Extending this approach, APR systems could be created so as to remain aware of uncertainty and express the degree of uncertainty in their outputs, for example when communicating a decision that is based upon terminology that has been flagged as being more vague. Systems could furthermore (self-)adaptively or autonomously (De Lemos et al. 2017; Li et al. 2020) take action to address uncertainty at run time, for example by involving a human operator in

---

[4] See https://www.highwaycodeuk.co.uk/using-the-road-overtaking.html.

[5] Some controlled languages, such as Attempto Controlled English (ACE) (Fuchs and Schwitter 1996), are formal languages. These can be directly and unambiguously translated into discourse representation structures.

Table 1 Overview of the 13 types of issues for evaluating APR and the leading questions that can be used to identify and address them

Issue type	Lead questions to evaluate the issue
Vagueness and balancing of interests	Were several possible and valid interpretations of the law implemented? Is there a technique in place to annotate elements of vagueness (e.g., to express probabilistic certainty about the interpretation)? Are vague terms implemented in a way that clearly differentiates them from the rest in the technical implementation of the APR system (to allow flexible modification or configuration)? Does the system evaluate the different outcomes that different interpretations may have, and raise an issue when these different outcomes are fundamentally divergent? Is the human enactment of the same regulation (e.g., in precedents) in line with the APR implementation? Was there sufficient A/B testing, verification and quality assurance before deployment of the APR implementation?
Evolution of Norms and Statutes	Can the system be adjusted to new interpretations of the law? Can the system enact such adjustments automatically? Is, and how is, the evolution of morale and social norms reflected in the implementation? What is the mechanism of how the APR keeps track of changes, both in terms of new interpretation from evolving social norms, and from new statutes and case law?
Lack of Interdisciplinarity	What is the demography and professional training of the individuals involved in the development and implementation of an APR project? Are the different viewpoints sufficiently represented (business, society, citizen)? Was ethical validation performed and ethical approval sought? How were the points of view of those from a professional/academic minority among the involved taken into consideration? Did reviews between fundamentally different expert groups take place? Who gave instructions to developers and how did developers seek advice when in doubt?
Agency	If a human judgment conflicts with the output from the algorithm, is there a process in place to ensure that the human can overrule the algorithm's decision? Is this overruling as well as the motivation behind it appropriately recorded, e.g., in a logfile? Is the implementation sufficiently transparent to show the rationale behind the decision, allowing humans to weigh the arguments against and for breaking the law in an exceptional situation?
Natural Pace	Are users of the system made aware that the system delivers decisions at a "non-human" speed? Does the project team anticipate that such faster processing time would lead to any negative psychological effects on certain users of your system? If applicable, has psychological counsel been sought to verify that this is not an issue?
Workforce Replacement	Who is impacted by work replacement? Does it replace work that some people enjoy doing? Does it replace work which was an essential life-support for those doing it? Does it replace work which offered compensation (financial, status-wise, etc.) that some regarded as either fair or even attractive?
Implementation Transparency	Can individuals, public authorities, and interested stakeholders have access to the implementation code, training datasets, trained models, and information on the APR implementation and deployment? How can different technical implementations enable more transparent and malleable approaches? Are the code, training datasets, and trained models easily accessible for example without burdensome procedures or intermediaries? Does the state play a role in educating its citizens in reading and understanding automatic processable regulation?
Process Transparency	Who verifies that processes are in place to catch errors and to correct any wrongs? Is this process communicated publicly, and clearly? Who verifies that data is collected, retained, and managed appropriately? How does the audit process take place? Should private companies that turn public regulation into APR or leverage APR come under an auditing process, and to what extent?
Affordability	Are the costs to the end user more manageable than through professional support? Should the state support, through subsidies or other means, the development of tools making the law more accessible, hence fostering the rule of law?
Usability	To which extent is the development user-centric? Are there aspects of the projects (micro or macro) that are unclear as to whether there has been a public debate around, and whether or how to bring this debate about? How can we ensure that there is a public debate if the implementation comes from the state?
Responsibility	In case of mistakes in APR, will it be possible to ascribe responsibility to one organizational unit (or a person within that unit), hence guaranteeing clear ownership and associated responsibility which in turn incentivizes developers to take precautions? Is the division of responsibility between encoding, inputting data, project management, and the resulting output clear? Is the division of tasks clear, or is it part of a complex organizational setup prone to hiding a lack of ownership? Is the hierarchy also well established when it comes to decision-making? Or is the culture axed toward group leadership, with groups loosely defined?
Reality	Is it clear to users whether the tool is a simulation or whether it is exactly the same tool that will be used for the official decision-making process? Why can the simulation and the actual decision-making system (not) be the same? Is the justification strong enough? Or, is this just a showcase of the public service's inefficiency? Are the messages displayed to users specific enough on when and how when and how simulation can differ from real usage of decision-making systems?

Table 1 (continued)

Issue type	Lead questions to evaluate the issue
Contestability	Can individuals technically and legally reverse the process by contesting the outcome? How cumbersome is it to appeal to the encoding or the outcome? What cost, if any, to the users does such appeal generate? How can arguments about fairness be brought in during the contestation of decision-making?

cases where the uncertainty about the decision exceeds a threshold.

3.2 Evolution of norms

APR projects should avoid "freezing" (Hildebrandt 2020) the law as well as social norms (Cobbe 2020). There are several ways for how the law might be updated at run time: A statute might be changed, a new statute might change the meaning of an existing statute, or new case law might change the interpretation of a statute. More subtly, a statute might remain in the law but be no longer enforced (Forstmoser and Vogt 2012). Evolving statutes or societal norms are prone to cause system behavior that is—or is perceived as—awkward by its users; or system behavior that is legally wrong after a statute change. It might also lead to dangerous situations, since norms or statutes may have evolved as a result of new insights into security hazards. An APR implementation should not be rigid or brittle in the face of such changes, but rather be designed to co-evolve with such changes (Sacco 1991; Tamò-Larrieux et al. 2021).

Example Regarding *statute evolution*, across many countries, lighting requirements for humans in indoor workplaces are today expressed in standards such as the UK's *BS EN 12464-1:2021* (British Standards Institution 2021). With slight variations, these recommend an illuminance level of 500 lux at a corrected color temperature (CCT) of 4000 K. The standards are formulated clearly enough to permit the implementation of an automated lighting system that encodes these requirements. However, future versions of the standards might not only prescribe different thresholds, but they might furthermore increase the level of granularity of the thresholds (e.g., CCT might be dependent on the time of day) and they might require taking into account human heterogeneity in aspects such as chronotypes, age, emotional states, and pathology. This example illustrates that norms (or other regulation) that have been used as a firm basis for implementing an APR system might not remain stable enough throughout the lifetime of that system. Another real-case example illustrates the problem of freezing *societal norms* in APR and involves the evolution of how society has regarded cohabitation of non-married couples: In Switzerland, a law that forbade such cohabitation was still officially valid well into the 1970s, but was not enforced due to evolving societal norms (Forstmoser and Vogt 2012). Fictive

APR-enabled locks that would only open to married couples would hence need to be updated to conform to the new social interpretation to the law. This is specifically tricky due to the lack of formal repelling of the law, hence requiring an APR system to interpret morality.

Mitigation strategy: Designing for evolution and change involves the modular implementation of an APR project that permits separating fast-evolving and slow-evolving parts of the underlying regulation. This design contains implications from statute or societal norm evolution to those (software) modules that are expected to be affected by this evolution while limiting the spread of such changes to other parts of the system through encapsulation. This can be accomplished by ensuring first that developers of APR projects understand at least the three different facets underpinning legal interpretation (Emanuilov et al. 2018): *legal formants* (i.e., the elements that constitute the "living law" of a state including the legislation but also the links between the provisions, case law, and legal doctrine), *cryptotypes* (i.e., the principles, ethical values, and assumptions underpinning the legal norms), and *synecdoche* (i.e., the fact that not all rules are fully articulated and that unexpressed general rules may be referred to by special rules). Furthermore, developers of APR projects should emphasize non-functional attributes of modifiability and evolvability by adopting software design patterns that facilitate change (Gamma 2002). In addition, the development process itself should be structured to remain flexible and iterative, e.g., by adopting processes such as agile or continuous development which are predicated upon continuous evolution and involve flexible update strategies (Beck et al. 2001; Fowler and Foemmel 2006). This process should specifically integrate periodical re-assessment of the validity of assumptions, which involves examining whether the assumptions made while encoding the law are still valid. During the development of an APR system, appropriate assumption management mechanisms should be adopted (Kruchten et al. 2006; Van Landuyt et al. 2012; Yang et al. 2017) to keep track of and model assumptions, for example to allow assessing the impact if assumptions were to be invalidated over time; this could be implemented as a form of test-driven development (Beck 2002). If updates cannot be done automatically, an approach should be identified that leads to

regular manual updates without the requirement to re-start the APR translation process from scratch.

3.3 Interdisciplinarity

The conceptualization, implementation, and deployment of an APR project requires the interaction of individuals who join the project in different roles, typically at least *domain experts*, *legal professionals*, and *software engineers* (incl. developers, architects, testers (Mohun and Roberts 2020b)). Bringing together the expertise of these individuals is challenging—this is due to professionally induced misalignments that range from the use of different vocabularies to ignorance regarding each other's assumptions, formation, and professional motives, and it furthermore includes the adherence to different and often incompatible problem-solving methodologies (Emanuilov 2018; Bisconti et al. 2023). Stakeholders of APR projects should expect this issue to be more prevalent when parts of the creation of APR is outsourced, as well as in teams that collaborate in settings that emphasize modularity, such as in agile approaches (Beck et al. 2001).

Example A software engineering firm develops a tool to automate the management of employee working time. Time-recording and over-time regulation are identified as ideal candidates to turn the underlying legal text into APR: These regulations are expressed in clear terms and specify clear numerical thresholds that can be easily interpreted and converted into program code (McNaughton 2020). Yet, a case from 2018 where plaintiffs disputed the interpretation of a (missing) comma in over-time legislation demonstrates that such assumptions are not as straightforward as suspected (Victor 2018): The legal text specified that there were exemptions to paying over-time for "packing for shipment or distribution". Truckers argued that the exemption included "packing for distribution", but that distribution itself is not exempt of over-time; and the court sided with this interpretation. Involvement of a legal professional in the process of creating the APR in this case would decrease the possible negative consequences of a falsely straightforward interpretation and likely would ensure that different options are shown to users in face of uncertainty.

Mitigation strategy: To mitigate issues that stem from the inevitable interdisciplinarity in an APR project, we propose that stakeholders should strive for the early sensitization of all project participants on the organizational level. This can be accomplished efficiently by having the project team train the translation of regulation into APR already during the project conceptualization phase, where we propose to make use of surprising and edge cases (see

the aforementioned examples) to illustrate the fallacies of believing that one's own interpretation of a legal clause is, indeed, clear. Early research notably points out that there is considerable variation in how legal experts and programmers understand and encode the law, the resources they draw upon, and their confidence in implementing legal tools (Escher et al. 2022).

3.4 Agency

APR systems that make decisions automatically without considering individual agency (i.e., capacity to make own decisions) may result in a significant loss of control for their users (Gill 2020). While developing AI systems that comply with legal standards and ethical principles is essential, an overly rigid approach can stifle individual discretion (Vladeck 2014; Müller 2020; Cervantes et al. 2020). For instance, an APR system designed to stop illegal content dissemination might automatically censor or report content, even in cases where there might be valid reasons to do so (Oversight Board 2020). Taking human agency into account is, thus, critical when designing and implementing APR, as systems should not automatically dismiss a person's decision in all cases, but rather include a way to handle a person's wish to disobey the law even when given clear warnings about it (Greenstein 2022). More generally, to have such mechanisms in place matter as the possibility to challenge the law via disobedience is crucial in liberal societies as a way to initiate moral and legal changes (Thoreau 2021). The alternative would be to over-emphasize instruments and procedures over human control and comprehension (Danaher 2016), a phenomenon that has been termed *algocracy*.

Example An autonomous vehicle's software is based on an APR version of traffic law, and has sensors that monitor the vehicle's environment. When the car detects a red light, it stops and prevents manual override by the driver (Thadeshwar et al. 2020; Zhang et al. 2021). The car's design, hence, preempts the human from taking a decision that is against the law. In case of exceptional situations such as a medical emergency, however, it might be required that the human judgment of the situation takes prime over the technology's APR-based enforcement of the law; in the case above, for instance, the driver should be able to cross a red light after weighing the risk that is associated with this action.

Mitigation strategy: In addition to introducing an explicit process step where legal professionals survey the regulation that is to be turned into APR in a specific project with respect to agency issues, we propose an agency-preserving and transparent user interface design to mitigate this issue (Almada 2019). APR systems should be implemented

so that the system *typically* preserves the human's agency with respect to breaking the law. The system in this case should make sure that the user explicitly accepts this agency, for instance through an explicit confirmation or a (possibly even physical) override switch that may function similarly to the well-known safety seals that void product warranty when broken (Lyons et al. 2021).

3.5 Natural pace

APR conflates the law and the application of the law (Hildebrandt and Koops 2010), with repercussions at both the individual and societal level. At the individual level, there is a psychological role at play when humans have the opportunity of being heard by another human being; emotions and empathy are a part of legal proceedings (Ranchordàs 2022). At the societal level, the application of the law has been a time-consuming endeavor notably to ensure a level of scrutiny and care. The effort to make processes more efficient (Zheng et al. 2022) can result in collapsing timelines when applying the law, with consequences however highly uncertain, as most people are not accustomed to receiving an answer to their legal queries delivered immediately (Robinson 2020) and might suffer from negative psychological consequences in effect. Potentially, individuals might start to question the validity of the answer obtained and the care in the due process, which ultimately could impact the trust placed in institutions and their decision-making (Ahn and Chen 2022).

Example Consider a scenario where a robot-judge streamlines the legal process, making it more efficient by quickly analyzing legal cases, evaluating evidence, and rendering decisions. However, a psychological issue arises when these decisions are delivered instantly. Imagine a couple going through a tough divorce. They are in a custody battle over their child, the family household needs to be sold, and emotions are running high. They appear before the robot-judge, present their arguments, and within seconds, the robot-judge issues a custody decision, granting primary custody to one of the parents. While the efficiency of the robot-judge is commendable, the immediate delivery of the decision may have a profound psychological impact on the couple since they receive the life-altering decision in a matter of seconds, leaving them potentially shocked, emotionally unprepared, and unable to process the outcome adequately.

Mitigation strategy: Even if a decision can be taken automatically and with very low delay, it does not have to be delivered right away after being reached. Stakeholders of APR projects should consider delaying the delivery of the response to allow users to process their experience, which is a straightforward implementable approach to alleviate such concerns. How much time the answer needs to be delayed before delivery depends on the exact context and may, in some cases, even be left to the user of the system.

3.6 Replacement of workforce

APR impacts current gatekeepers to the law: legislators, judges, lawyers, and a wealth of para-legal professions (Susskind and Susskind 2015). It is conceivable that new forms of cooperation will emerge, for instance between those with the legal know-how and those with the technical one, yet it is also likely that APR will give rise to similar debates as witnessed during the digitalization of government services (Hanschke and Hanschke 2021; Bucher et al. 2021; Halvorsen et al. 2021) and to which extent replacement will take place (Tobar and González 2022). These debates focused on worker protection, adequate compensation, social status, and more. While replacement of repetitive and unpleasant tasks might be socially welcomed (Frey and Osborne 2017), oftentimes the deployment of automated systems occurs without taking the opinion of individual workers impacted by the systems into account (Rigotti et al. 2023).

Example A start-up seeks to cut the costs of legal firms by developing tools for para-legals professionals. Those in the para-legal profession may have good reasons to be in it: Legal schools can be exorbitantly expensive (e.g., in the US), and the type of work can be very different with taking on less responsibility and low(er) level of stress (Moran 2020). The start-up is developing many tools around information research and e-discovery which could raise the prospects of replacing a certain number of para-legals assistants. The start-up and their future clients may have to balance and contrast their vision of reducing costs by replacing employees with one of supporting employees performing better at their job. The nuance in marketing the tool would have repercussions on many lives.

Mitigation strategy: The first step to mitigate this issue is to clearly indicate and analyze what the APR solution can create in terms of costs/benefits. From interviews with stakeholders of APR projects, we observe that many such projects have started without a model of how much investment will be required to generate a return (Guitton et al. 2022a). Furthermore, a distinction must be made between private-led projects and public ones: For projects driven by private companies, the companies will likely have surveyed the market to realize the business potential of their investment; only specific regulations or subsidies could prevent the deployment of solutions in such a way that it would wipe out a whole class of employees. On the other hand, whenever

a state seeks to develop a solution that could replace a large number of employees, the role of the state and the prevailing social contract should come back to the fore, and if needed, debates should occur (Kochan and Dyer 2021).

3.7 Implementation transparency

Transparency is important because it helps evaluate whether decisions are lawful and therefore justified (Yeung and Weller 2018; Hollanek 2023). Following the same line of argumentation, since laws are public to enable the rule of law and review of lawfulness, when legislation becomes encoded, the relevant code, training datasets, and trained models should be made accessible to the broader public as well. Access to this relevant implementation details constitutes an aspect of the right to access to and receive information (Tamò-Larrieux et al. 2021). On top of this, public review of implementation can serve as an additional check on both specific interpretation or error which were introduced into the code, datasets, or models (implicit or explicit, malign or benign) (Tamò-Larrieux et al. 2021). Those responsible for encoding should be able to justify choices made throughout implementation, and the publication of details should automatically come with the publication of such rationale (Malgieri 2021).

Example Many countries have developed supporting software solutions for citizens or businesses to fill out their taxes, such as the Tax Authority in the Netherlands using the Standard Business Reporting software.[6] Although the tax code could be *a priori* straightforward for encoding, there are many questions as to what constitutes revenue, deductible spending, and how certain assets ought to be declared (Lawsky 2013). Without an independent party outside of employees of the state reviewing the quality and suitability of the code, choices around situations, which do not have a straightforward answer and cannot be prodded. As a result, citizens are often left with the assumption, dangerously, that the solution the state has provided is correct. Besides, whether such a solution with a public function should come from a state or from a private commercial entity raises the transparency and legitimacy questions of delegating public functions to private actors, without proper checks and balances in place (Yeung 2023).

Mitigation strategy: Several issues persist with the publishing of the code and rationale behind APR in practice (Guitton et al. 2022a). First for state institutions: Despite existing laws mandating the publication of code, there have

been cases with push-back where state institutions have contested and refused to deliver their code (Cluzel-Métayer 2020; Alauzen 2021). Parliaments enacting laws that mandate the publication of code when state institutions engage in APR should, hence, be a basic requirement, but this can still be insufficient; further steps should include the possible creation of an ombudsman position overseeing the proper release of implementation details, and civil societies continuing to apply pressure for laws to be enforced as enacted. The situation is more delicate when it comes to companies that create APR. Companies would rightly see their development giving them a competitive advantage in their market, and hence as proprietary. In this case, companies could still publish details of the choices they had to make while encoding laws. They could also contract audit firms that verify equivalences between the regulation and the APR, with only the result from the audit being published.

3.8 Process transparency

Transparency is also required when it comes to correcting mistakes (Descampe et al. 2022; Walmsley 2021; Andrada et al. 2023). While we classify correcting wrongs as falling within contestability (see below), there should be a process upstream to ensure that mistakes are caught in the design of the implementation and test phases. This process should be clear and transparent: For APR, the risk could be that the process falls between established channels because of its nature. For those working in private companies, it could even be that there is an underlying assumption that users which are not clients are not given the possibility to contest.

Example A state regulator outsources to a private entity to develop an APR version of a public law impacting financial companies, which specifically deals with the modeling to calculate the risk-bearing capital of insurers. By publishing the APR version, the regulator hopes that it makes it easier to compare and accepts insurers' models, while insurers hope that it also makes it easier for them to understand how the regulator would accept or reject the model. The regulator makes the code public. One insurer, however, disagrees with how the contractor encoded a part of the regulation. With the asymmetry of power between the insurer and the public regulator, the company is unsure of the repercussion of engaging in trying to correct the mistakes. Without the establishment of a proper neutral channel, the insurer is concerned that it would have to raise the issue during an on-site visit by the regulator, entangling the issue with many unrelated ones and thereby making it more difficult for it to be heard.

Mitigation strategy: Users of an APR should easily be able to find how to initiate a contest process, and those in

---

[6] See https://www.belastingdienst.nl/wps/wcm/connect/bldcontenten/belastingdienst/business/tax_return/filing_digital_tax_returns/filing-tax-returns-using-accounting-software.

charge of APR should ensure that this process is appropriately staffed to handle queries. The different steps should be understandable and users should be able to track the evolution of their queries. Those attending to users' queries should have the authority not only to escalate them but also to take real action to remedy them. To limit conflict of interest, those assessing the queries should also differ from the ones who were involved in the implementation.

3.9 Affordability

A review of current APR projects has shown that many implementations are geared toward making the law more accessible (Guitton et al. 2022b). This is a welcome step considering the gap that exists between the concept that ignorance of the law is no excuse (Brooke 1992), and laypeople's actual understanding of law (van Rooij 2020). APR can, therefore, provide a starting point for a legal self-assessment; yet only if it is affordable. In light of this, a key question is what the role of the state in supporting everyone to improve their knowledge of the law should be—a debate going back centuries (Herzog 2018). States can decide to leverage public funds to promote the development of APR for accessibility by either engaging in development themselves, or by subsidizing companies' offers, or with a mixture of both. Whatever the policy is, the use of public money for such a goal should be debated and made explicit, acknowledging the shortfalls of the current assumption that everyone knows the law, as well as acknowledging the difficulty of carrying out a cost/ebenfit analysis of this type of envisioned public service.

Example A company develops a tool that allows citizens to ask for the penalty that they might incur if they commit a criminal deed; the tool takes into consideration statutory limits and case law, and can be tasked with complex queries. The tool could potentially contribute to deterrence of violations of legislation. One of the weakness of deterrence theory however is that not all criminals or people who violate the law necessarily think rationally weighing out costs/benefits as the costs are too difficult to gauge.

Mitigation strategy: In case the market does not develop toward affordable solutions fostered by competition, the state could subsidize certain solutions that are proven to be of broader societal use or make them available without cost by integrating them in the public governance toolset via public procurement. In the case of subsidies, appropriate safeguards and post-market surveillance should be in place to ensure that the subsidized solution continues to be used for public good as originally intended. Those safeguards could take

the form of certification, and be subject to the jurisdiction of the audit courts.

3.10 Usability

The law is complex, and when attempting to make it more accessible, issues may occur, most notably two of interest within this section: Either the law is made more simple than it really is, what some have called "simplexity" (Blank and Osofsky 2020), or the APR is so cumbersome to use that it does not bring support to laypeople. Designing usable software has a long history, going as far back as to the 1980s (Gould 1988). Nowadays, there is a wealth of literature on how to design usable user interfaces (Li and Nielsen 2019; Göransson et al. 2004). Furthermore, several studies have linked the interface design choices with acceptance, being part of instigating structural changes, or conversely, with fostering inequalities between those with the appropriate background and those without (Hadfield-Menell et al. 2016). Developers of APR should, therefore, be aware that many apparently meaningless choices are not as they appear to be—and should take into consideration different users' views.

Example A tool is developed to answer legal questions for laypeople. When given a situation, it provides with relevant case law to compare. Delving into case law can be cumbersome and intricate though: Extracting which ones are still valid opinions, which ones apply to the very specific case at hand and which ones do not, and more, can require professional training. A tool that makes it difficult for users to navigate through this without taking the user experience into account would defeat the point of creating a tool in the first place.

Mitigation strategy: The usual best practice to test user interface and user experience should be adopted in APR projects. This means conducting several surveys on user groups that are representative of the population on the type of problems people are experiencing even prior to developing, leveraging commonly used techniques from the field of user-centered design such as feature fakes or Wizard-of-Oz prototyping to test whether the features could be useful before development. Post-development usability testing, feedback follow-ups, and assessing different variants with split/run tests and blind product tests should be conducted.

3.11 Responsibility

Assigning responsibility has two purposes in law and technology governance (Schwartz 1996)—although we note that others have identified different break-down of responsibility within the field (Dastani and Yazdanpanah 2023). On the

one hand, there is a need to guarantee that *someone* will pay if another party suffers damages; this condition is necessary to sustain trust in society (Pagallo 2013; Tamò-Larrieux et al. 2023). On the other, assigning of responsibility aims to incentivize an appropriate level of care, in the sense that it deters possible mistakes by increasing the perception of the likelihood of negative consequences (Shuman 1993). In other words, assigning responsibility to specific agents—either through tort liability (Pałka 2021) or administrative sanctions (Pahlka 2023)—aims to ensure both that the citizens trust the systems they interact with, and that the persons to-be-responsible (and who aim to avoid financial or reputational sanctions) strive to avoid mistakes. This is no different in the specific case of APR: Mistakes are likely to be made, hence a responsible framework should assign responsibility to enable trust while incentivizing a high level of care.

Example A local small-claims court receives many lawsuits that are inadmissible on formal grounds; its employees lose a significant amount of time verifying the formalities and instructing the claimants to correct their paperwork. The president of the court hence decides to hire a consulting firm that is supposed to help design an automatic system for the formal verification of lawsuits and that automatically responds to the claimants whose documents are lacking. The consulting firm proposes near-full automation (with a human having to click an *Accept* button before the decision is sent out), designs the specifications of the system, and outsources its creation to a software engineering firm. This company implements the system according to the requirements, and the judges delegate to the clerks the task of accepting the automated decisions. The clerks cannot access the reasoning of the system but, trusting it, end up accepting nearly everything. After a while, it turns out that the number of false positives (i.e., lawsuits that were formally correct but have been rejected by the system) is high, while the instructions on how to amend the claims are actually wrong. An angry citizen approaches the president of the court who, enraged, confronts a judge, who summons a clerk; the clerk is left wondering where to complain further.

Mitigation strategy: Agile development processes that stretch from the discovery of the requirements to the continuous deployment and testing are today the standard approach for the development of software (and, to a lesser extent, hardware) systems. However, it does not promote a clear responsibility for when things start going wrong. A possibility would be to have a clear *product owner* or equivalent hierarchical decision-taker per division line encoding/input data/implementation/outcome. This person would be charged, within their division, with primary responsibility for devising objectives for what it means to be mistake-free, and for ensuring that these are followed through correctly.

Further, both at the design stage and the implementation stage, intra-organizational responsibility should be clear.

3.12 Reality

Governments might deploy APR systems for individuals to determine their eligibility to specific benefits, but these systems might be created to only issue "simulations" that are not legally binding (Guitton et al. 2022b). The reasoning behind such simulations rather than decision-making tools is multifold, such as the fear of a backlash due to badly encoded legislation or that the publication of code could facilitate fraud (Guitton et al. 2022b). Regardless of the merit of the concerns, users may mistake the simulation for a real legally binding tool, leading to disappointed refused applicants turning their anger and frustration to clerks left with more work to explain the difference in output (Alauzen 2021). Furthermore, if the differences between the simulation and the actual decision-making system are too large, it begs the question of the added value of the simulation in the first place.

Example A company develops a tool for determining how much in social benefits a person is entitled, all on the basis of current in-force legislation (at the time of the tool's development). The company sells the tool to the government that further uses it for its final decision-making. Later, the company refactors the code with tweaks to make it easier to use and publishes a simulation for everyone to use online. However, as a result of the simplification, the outputs turn out to be different between the versions used by officials and the free online one.

Mitigation strategy: A low-hanging fruit should be the clear labeling of an APR product as a simulation, with significant warnings on how, as a simulation, it differs from the tool used for (legally binding) decision-making. In general, however, we advocate to minimize the use of simulations and to promote rather the use by end users of the same application that is also used for legal decision-making. When too many factors make this impossible, we recommend pointing out cases where a deviation can be expected and possibly even restricting the use of the application only for cases that have been thoroughly tested for equivalence with the outcome from the real automated decision-making algorithm so as to minimize any chance of deviation.

3.13 Contestability

Procedural rights are one of the cornerstones of liberal democracy (Krajewski 2021). This explains why in the context of automated decisions, contestability has received significant attention in legal and computer science

scholarship (Almada 2019; Tubella et al. 2020; Lyons et al. 2021; Bayamlıoğlu 2022; Fanni et al. 2023). While process transparency ensures that users are able to find the correct procedure to follow to file their objection, contestability ensures that they can effectively follow through with it. For APR, contestability translates into two distinct rights: the right to contest the encoding and training data of the APR, and the right to contest a specific decision. Central to APR is notably the difficulty of bringing together legal and technical expertise to understand the APR implementation and how the legal interpretation might be problematic. Arguably, the issue of contestability will be much more salient in projects geared toward more efficient decision-making than those geared at offering a greater accessibility to the law (Henin and Le Métayer 2022). The overarching goal of contestability, in both cases, should however remain to ensure that there is a means—be it via a due process, or via a technical implementation—that humans can overrule either the implementation or the outcome of the algorithm (Beutel et al. 2019).

Example Reusing the example of the automatic decision of rejection of admissibility of a case to a small-claims court, this begs the question regarding the recourse that a person could have to appeal to the rejection (especially given that, in fact, the lawsuit was correct on formal grounds), and whether the person could contest the encoding directly with the implementing firm and/or the consultancy which managed the project. There are, therefore, at least two distinct aspects involved: those concerning the encoding and those concerning the outcome.

Mitigation strategy: Although regulation such as GDPR has created the right to correct one's own personal data, have it deleted, or even not being subject to fully automated decision-making (Kuśmierczyk 2022; Bayamlıoğlu 2022) a channel should be established for users to flag and seek redress when they think that they have been treated wrong. In case companies and the state do not show enough proactivity in setting up such channels and making sure that they are staffed properly, legislation should fill this void by introducing a requirement for *contestability by design*.

4 Conclusion and future work

The presented framework, which is based on the state-of-the-art literature in the fields of computer science and law, provides a foundation to deliberate about the responsible conceptualization, implementation, and usage of APR

projects. Aside from the useful categorization of issues into 13 types, we provide explanations as well as examples on what the core of the issue is centered around and how, through the use of guiding questions (see Table 1) we can identify those. Our approach combining descriptions, examples, and guiding questions help to make informed decisions about whether and how to proceed with an ongoing APR project; ideally, such an informed discussion occurs within the initial phases of the conceptualization of an APR implementation and continues throughout its implementation and beyond, through the entire lifetime of the APR. In addition, and addressing the identified gap within the literature on making ethical guidelines operationalizable, we proposed generic mitigation strategies that can be adapted to different contexts.

Future research will need to evaluate further, how in practice the responsible APR framework is operationalized and used, and how it impacts the final design of APR projects. We would like to recommend that stakeholders using our framework commit to sharing evaluations of its effectiveness, including what issues were accurately identified beforehand, what changes were made, and what aspects may have been overlooked and why, in an effort to continuously improve the responsible development of APR projects. Through a workshop, we have already conducted an expert evaluation but plan to go beyond this, notably by replicating how others have been suggesting testing frameworks for responsible AI (Amershi et al. 2019; Morales-Forero et al. 2023). We plan as well to follow a multi-stage approach, with: (1) a "modified heuristic evaluation" (Amershi et al. 2019) during which participants are asked to provide examples of both application and violation of the issue; (2) a user study during which participants will evaluate one APR project against one issue of the framework; (3) a vignette study during which participants will be asked to design an approach to turn existing statute into APR while heeding the framework. This validation should contribute to further strengthening our understanding, development, and education of responsible APR, a necessary step to seize on the potential of APR to shape societies into ones where people have a better understanding of regulation, and where regulatory processes are run more efficiently.

Declarations

References

Ahn MJ, Chen Y-C (2022) Digital transformation toward AI-augmented public administration: the perception of government employees and the willingness to use AI in government. Gov Inform Q 39(2):101664

Alauzen M (2021) Splendeurs et misères d'une start-up d'Etat: Dispute dans la lutte contre le non-recours aux droits sociaux en France (2013–2020). Réseaux 225(2021):121–150

Almada M (2019) Human intervention in automated decision-making: toward the construction of contestable systems. In: Proceedings of the Seventeenth International Conference on artificial intelligence and law, pp 2–11

Amershi S, Weld D, Vorvoreanu M, Fourney A, Nushi B, Collisson P, Suh J, Iqbal S, Bennett PN, Inkpen K, Teevan J, Kikin-Gil R, Horvitz E (2019) Guidelines for human-AI interaction. In: CHI

Anderson MM, Fort K (2023) From the ground up: developing a practical ethical methodology for integrating AI into industry. AI & Soc 38(2):631–645. https://doi.org/10.1007/s00146-022-01531-x

Andrada G, Clowes RW, Smart PR (2023) Varieties of transparency: exploring agency within AI systems. AI & Soc 38(4):1321–1331. https://doi.org/10.1007/s00146-021-01326-6

Ashley KD (2017) Artificial intelligence and legal analytics: new tools for law practice in the digital age. Cambridge University Press, Cambridge

Bauer E, Stammbach D, Gu N, Ash E (2023) Legal extractive summarization of US court opinions. arXiv preprint arXiv:2305.08428

Baum SD, Owe A (2023) From AI for people to AI for the world and the universe. AI & Soc 38(2):679–680. https://doi.org/10.1007/s00146-022-01402-5

Bayamlıoğlu E (2022) The right to contest automated decisions under the General Data Protection Regulation: beyond the so-called "right to explanation''. Regul Gov 16(4):1058–1078

Beck K (2002) Test driven development. By example (Addison-Wesley Signature). Addison-Wesley Longman, Amsterdam

Beck K, Beedle M, Bennekum AV, Cockburn A, Cunningham W, Fowler M, Grenning J, Highsmith J, Hunt A, Jeffries R, Kern J, Marick B, Martin RC, Mellor S, Schwaber K, Sutherland J, Thomas D (2001) Manifesto for Agile Software Development. https://agilemanifesto.org/

Bench-Capon T (2022) Thirty years of Artificial Intelligence and Law: editor's introduction. Artif Intell Law 30:475–479

Beutel A, Chen J, Doshi T, Qian H, Woodruff A, Luu C, Kreitmann P, Bischof J, Chi EH (2019) Putting fairness principles into practice: challenges, metrics, and improvements. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. ACM, pp 453–459

Bisconti P, Orsitto D, Fedorczyk F, Brau F, Capasso M, De Marinis L, Eken H, Merenda F, Forti M, Pacini M, Schettini C (2023) Maximizing team synergy in AI-related interdisciplinary groups: an interdisciplinary-by-design iterative methodology. AI & Soc 38(4):1443–1452. https://doi.org/10.1007/s00146-022-01518-8

Black J, Kingsford DS (2002) Critical reflections on regulation [Plus a reply by Dimity Kingsford Smith]. Australas J Leg Philos 27(2002):1–46

Blank JD, Osofsky L (2020) Automated legal guidance. Cornell Law Rev 106(2020):179–246

Bommarito MJ, Katz DM (2023) GPT takes the bar exam. https://doi.org/10.2139/ssrn.4314839

British Standards Institution (2021) Light and lighting-lighting of work places–part 1: indoor work places. BSI

Brooke A (1992) When ignorance of the law became an excuse: lambert its progency. Am J Crim law 19(2):279–312

Bucher E, Fieseler C, Lutz C (2021) Mattering in digital labor. J Manag Psychol 34(4):307–324

Cervantes J-A, López S, Rodríguez L-F, Cervantes S, Cervantes F, Ramos F (2020) Artificial moral agents: a survey of the current status. Sci Eng Ethics 26(2020):501–532

Chalkidis I, Fergadiotis M, Malakasiotis P, Aletras N, Androutsopoulos I (2020) LEGAL-BERT: the muppets straight out of law school. EMNLP, In Findings of the Association for Computational Linguistics, pp 2898–2904

Cluzel-Métayer L (2020) The judicial review of the automated administrative act. Eur Rev Digit Admin Law 1(1–2):101–103

Cobbe J (2020) Legal singularity and the reflexivity of law. Hart Publishing, Oxford

Danaher J (2016) The threat of algocracy: reality, resistance and accommodation. Philos Technol 29(3):245–268

Dass RK, Petersen N, Omori M, Lave TR, Visser U (2023) Detecting racial inequalities in criminal justice: towards an equitable deep learning approach for generating and interpreting racial categories using mugshots. AI & Soc 38(2):897–918. https://doi.org/10.1007/s00146-022-01440-z

Dastani M, Yazdanpanah V (2023) Responsibility of AI Systems. AI & Soc 38(2):843–852. https://doi.org/10.1007/s00146-022-01481-4

De Lemos R, Garlan D, Ghezzi C, Giese H, Andersson J, Litoiu M, Schmerl B, Weyns D, Baresi L, Bencomo N et al (2017) Software engineering for self-adaptive systems: research challenges in the provision of assurances. In: Software Engineering for Self-Adaptive Systems III. Assurances: International Seminar, Dagstuhl Castle, Germany, December 15-19, 2013, Revised Selected and Invited Papers. Springer, pp 3–30

Descampe A, Massart C, Poelman S, Standaert F-X, Standaert O (2022) Automated news recommendation in front of adversarial examples and the technical limits of transparency in algorithmic accountability. AI & Soc 37(1):67–80. https://doi.org/10.1007/s00146-021-01159-3

Diver L, McBride P, Medvedeva M, Banerjee Arjun B, D'hondt E, Nicolau Tatiana D, Dushi D, Gori G, Van Den Hoven E, Meessen P, Hildebrandt M (2022) Typology of legal technologies. https://researchportal.vub.be/en/publications/typology-of-legal-technologies. Accessed 14 Dec 2023

Emanuilov I (2018) Navigating law and software engineering towards privacy by design: stepping stones for bridging the gap. In: Computers, Privacy and Data Protection Conference 2018, Date: 2018/01/24-2018/01/26, Location: Brussels

Emanuilov I, Wuyts K, Van Landuyt D, Bertels N, Coudert F, Valcke P, Joosen W (2018) Navigating law and software engineering towards privacy by design: stepping stones for bridging the gap. In: Leenes R, van Brakel R, Gutwirth S, De Hert P (eds) Data protection and privacy: The internet of bodies, pp 123–140

Endicott T (2011) The value of vagueness. Oxford University Press, Oxford, pp 14–30

Escher N, Bilik J, Miller A, Huseby JJ, Ramesh D, Liu A, Mikell S, Cahill N, Green B, Banovic N (2022) Cod(e)ifying the law. In: Programming languages and the law (ProLaLa), Philadelphia

Fanni R, Steinkogler VE, Zampedri G, Pierson J (2023) Enhancing human agency through redress in Artificial Intelligence Systems. AI & Soc 38(2):537–547. https://doi.org/10.1007/s00146-022-01454-7

Fjeld J, Achten N, Hilligoss H, Nagy A, Srikumar M (2020) Principled artificial intelligence: mapping consensus in ethical and rights-based approaches to principles for AI. Berkman Klein Center Research Publication

Floridi L, Cowls J (2021) A unified framework of five principles for AI in society, vol 144. Springer, Cham, pp 5–6

Forstmoser P, Vogt HU (2012) Einführung in das Recht. Stämpfli. https://books.google.ch/books?id=NpKoMAEACAAJ. Accessed 14 Dec 2023

Fowler M, Foemmel M (2006) Continuous integration. https://www.martinfowler.com/articles/continuousIntegration.html

Frey CB, Osborne MA (2017) The future of employment: how susceptible are jobs to computerisation? Technol Forecast Soc Change 114(2017):254–280

Fuchs NE, Schwitter R (1996) Attempto Controlled English (ACE). In: CLAW 96, First International Workshop on Controlled Language Applications

Gamma E (2002) Design patterns-ten years later. In: Broy M, Denert E (eds) Software pioneers: contributions to software engineering, Springer, Berlin, Heidelberg, pp 688–700. https://doi.org/10.1007/978-3-642-59412-0-39

García K, Zihlmann Z, Mayer S, Tamò-Larrieux A, Hooss J (2021) Towards privacy-friendly smart products. In: 18th International Conference on Privacy, Security and Trust (PST)

Gill KS (2020) AI &Society: editorial volume 35.2: the trappings of AI Agency. AI & Soc 35(2):289–296. https://doi.org/10.1007/s00146-020-00961-9

Giovanola B, Tiribelli S (2023) Beyond bias and discrimination: redefining the AI ethics principle of fairness in healthcare machine-learning algorithms. AI & Soc 38(2):549–563. https://doi.org/10.1007/s00146-022-01455-6

Göransson B, Gulliksen J, Boivie I (2004) The usability design process-integrating user-centered systems design in the software development process. Softw Process Improv Pract 8(2):111–131

Gould JD (1988) How to design usable systems. Elsevier

Greenstein S (2022) Preserving the rule of law in the era of artificial intelligence (AI). Artif Intell Law 30(3):291–323

Guitton C, Tamò-Larrieux A, Mayer S (2022a) Mapping the issues of automated legal systems: why worry about automatically processable regulation? Artif Intell Law 31:571–599

Guitton C, Tamò-Larrieux A, Mayer S (2022b) A Typology of Automatically Processable Regulation. Law Innov Technol 14:2

Hadfield-Menell D, Dragan A, Abbeel P, Russell S (2016) Cooperative inverse reinforcement learning. In: 30th Conference on Neural Information Processing Systems (NIPS 2016)

Halvorsen T, Lutz C, Barstad J (2021) The collaborative economy in action: European perspectives. In: Klimczuk A, Česnuityte V, Avram G (eds), The collaborative economy in action: European perspectives, pp 224–235

Hanschke V, Hanschke Y (2021) Clickworkers: problems and solutions for the future of AI labour

Hart HLA 1961 (2012) The concept of law. Clarendon Press, Oxford

Henin C, Le Métayer D (2022) Beyond explainability: justifiability and contestability of algorithmic decision systems. AI & Soc 37(4):1397–1410. https://doi.org/10.1007/s00146-021-01251-8

Herzog T (2018) A short history of European Law: the last two and a half millennia. Harvard University Press, Cambridge, MA

Hildebrandt M (2020) Code-driven law: freezing the future and scaling the past. Hart Publishing, Oxford

Hildebrandt M, Koops B-J (2010) The challenges of ambient law and legal protection in the profiling era. Mod Law Rev 73(3):428–460

HLEG AI (2019) Ethics guidelines for trustworthy AI. Report. European Commission. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Hoffmann-Riem W (2022) Legal technology/computational law: preconditions, opportunities and risks. Cross-Discipl Res Comput Law 1:1

Höfler S, Bünzli A (2010) Designing a controlled natural language for the representation of legal norms. In: Second Workshop on Controlled Natural Languages. https://doi.org/10.5167/uzh-35842

Hollanek T (2023) AI transparency: a matter of reconciling design with critique. AI & Soc 38(5):2071–2079. https://doi.org/10.1007/s00146-020-01110-y

Hood C, Rothstein H, Baldwin R (2001) The government of risk: understanding risk regulation regimes. Oxford University Press, Oxford

Kanapala PSA, Pamula R (2019) Text summarization from legal documents: a survey. Artif Intell Rev 51(2019):371–402

Katz DM, Bommarito MJ, Shang G, Arredondo P (2023) GPT takes the bar exam. https://doi.org/10.2139/ssrn.4389233

Kochan TA, Dyer L (2021) Shaping the future of work: a handbook for building a new social contract. Routledge, London

Kowalski R, Dávila J, Sartor G, Calejo M (2023) Logical English for law and education. Springer, Cham, pp 287–299

Krajewski M (2021) Relative authority of judicial and extra-judicial review: EU courts, boards of appeal, ombudsman, vol 105. Bloomsbury Publishing

Kruchten P, Lago P, Van Vliet H (2006) Building up and reasoning about architectural knowledge. In: International Conference on the quality of software architectures. Springer, pp 43–58

Kuśmierczyk M (2022) Algorithmic bias in the light of the GDPR and the proposed AI act. In: Equality. Faces of modern Europe", Wydawnictwo Centrum Studiów Niemieckich i Europejskich im. Willy'ego Brandta, Wrocław

LabPlus (2018) Better Rules for Government: Discovery Report. Report. New Zealand Government. https://www.digital.govt.nz/dmsdocument/95-better-rules-for-government-discovery-report/html

Lagioia F, Rovatti R, Sartor G (2023) Algorithmic fairness through group parities? The case of COMPAS-SAPMOC. AI & Soc 38(2):459–478. https://doi.org/10.1007/s00146-022-01441-y

Lawsky SB (2013) Modeling uncertainty in tax law. Stanf Law Rev 65(2013):241

Leslie D (2019) Understanding artificial intelligence ethics and safety: a guide for the responsible design and implementation of AI systems in the public sector. Report. The Alan Turing Institute. https://www.turing.ac.uk/sites/default/files/2019-08/understanding_artificial_intelligence_ethics_and_safety.pdf

Li M, Nielsen P (2019) Making usable generic software. A matter of global or local design?. In: Tenth Scandinavian Conference on information systems

Li N, Cámara J, Garlan D, Schmerl B (2020) Reasoning about when to provide explanation for human-in-the-loop self-adaptive systems. In: Proceedings of the 2020 IEEE Conference on autonomic computing and self-organizing systems (ACSOS), Washington, DC. pp 19–23

Liga D, Robaldo L (2023) Fine-tuning GPT-3 for legal rule classification. Comput Law Secur Rev 51:105864

Loi M (2020) People Analytics must benefit the people. An ethical analysis of data-driven algorithmic systems in human resources management. Report. Algorithmwatch

Lyons H, Velloso E, Miller T (2021) Conceptualising contestability: perspectives on contesting algorithmic decisions. In: Proceedings of the ACM on Human-Computer Interaction 5(CSCW1):1–25

Malgieri G (2021) "Just" Algorithms: justification (beyond explanation) of automated decisions under the GDPR. Law Bus 1:16–28

Mantelero A, Esposito S (2021) An evidence-based methodology for human rights impact assessment (HRIA) in the development of AI data-intensive systems. Comput Law Secur Rev Forthcoming

Mário C, de Rodrigues O, de Freitas FLG, Barreirosa EFS, de Azevedoc RR, de Almeida Filho AT (2019) Legal ontologies over time: a systematic mapping study. Expert Syst Appl 130(2019):12–30

McNaughton S (2020) Innovate on Demand episode 7: regulatory artificial intelligence. https://csps-efpc.gc.ca/podcasts/innovate7-eng.aspx. Accessed 14 Dec 2023

Mohun J, Roberts A (2020a) Cracking the code: rulemaking for humans and machines. Report. OECD Working Papers on Public Governance No. 42

Mohun J, Roberts A (2020b) Cracking the code: rulemaking for humans and machines. Report. OECD Working Papers on Public Governance No. 42

Moore JG (2020) Hart, Radbruch and the necessary connection between law and morals. Law Philos 39(6):691–761

Morales-Forero A, Bassetto S, Coatanea E (2023) Toward safe AI. AI & Soc 38(2):685–696. https://doi.org/10.1007/s00146-022-01591-z

Moran M (2020) A new frontier facing attorneys and paralegals: the promise challenges of artificial intelligence as applied to law legal decision-making. Leg Educ, Fall/Winter, pp 6–15

Morley J, Floridi L, Kinsey L, Elhalal A (2020) From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. Sci Eng Ethics 26(2020):2141–2168

Moses LB (2020) Not a single singularity. Hart Publishing, Oxford

Müller VC (2020) Ethics of artificial intelligence and robotics. Stanf Encycl Philos. https://plato.stanford.edu/entries/ethics-ai/

Nguyen HT, Phi MK, Ngo XB, Tran V, Nguyen LM, Tu MP (2022) Attentive deep neural networks for legal document retrieval. Artif Intell Law 2022:1–30

Oversight Board (2020) Breast cancer symptoms and nudity. Report. Meta. https://www.oversightboard.com/decision/IG-7THR3SI1. Accessed 14 Dec 2023

Pagallo U (2013) The laws of robots: crimes, contracts, and torts, vol 10. Springer

Pahlka J (2023) Recoding America: why government is failing in the digital age and how we can do better, pp 336

Pałka P (2021) Private law and cognitive science. Cambridge University Press, pp 217–248. https://doi.org/10.1017/9781108623056.011

Palmirani M, Governatori G, Rotolo A, Tabet S, Boley H, Paschke A (2011) LegalRuleML: XML-based rules and norms. In: International Workshop on Rules and Rule Markup Languages for the Semantic Web (Lecture Notes in Computer Science book series, Vol. 7018). Springer, pp 298–312

Prem E (2023) From ethical AI frameworks to tools: a review of approaches. AI and Ethics. https://doi.org/10.1007/s43681-023-00258-9

Radbruch G (2006) Five minutes of legal philosophy (1945). Oxf J Leg StudiesGustav Radbruch 26(1):13–15

Ranchordàs S (2022) Empathy in the digital administrative state. Duke Law J Forthcoming

Rigotti C, Puttick A, Fosch-Villaronga E, Kurpicz-Briki M (2023) Mitigating diversity biases of AI in the labor market

Robinson SC (2020) Trust, transparency, and openness: how inclusion of cultural values shapes Nordic national public policy strategies for artificial intelligence (AI). Technol Soc 63(2020):101421

Sacco R (1991) Legal formants: a dynamic approach to comparative law (Installment I of II). Am J Comp Law 39(1):1–34

Sartor G, Branting K (1998) Judicial applications of artificial intelligence. Springer-Science+Business Media, B.V., Dordrecht

Schwartz GT (1996) Mixed theories of tort law: affirming both deterrence and corrective justice. Tex L Rev 75(1996):1801

Sergot MJ, Sadri F, Kowalski RA, Kriwaczek F, Hammond P, Cory HT (1986) The British Nationality Act as a logic program. Commun ACM 29(5):370–386

Shafei A, Hodges J, Mayer S (2018) Ensuring workplace safety in goal-based industrial manufacturing systems

Shuman DW (1993) The psychology of deterrence in tort law. U Kan L Rev 42(1993):115

Susskind R, Susskind D (2015) The future of the professions: how technology will transform the work of human experts. Oxford University Press, Oxford

Tamò-Larrieux A, Mayer S, Zihlmann Z (2021) Not hardcoding but Softcoding data protection principles. Technol Regul 2021:17–34

Tamò-Larrieux A, Guitton C, Mayer S, Lutz C (2023) Regulating for trust: can law establish trust in artificial intelligence? Regulation & Governance. https://doi.org/10.1111/rego.12568

Thadeshwar H, Shah V, Jain M, Chaudhari R, Badgujar V (2020) Artificial intelligence based self-driving car. In: 2020 4th International Conference on computer, communication and signal processing (ICCCSP). IEEE, pp 1–5

Thoreau H 1849 (2021) On the duty of civil disobedience. Antiquarius, La Vergne

Tobar F, González R (2022) On machine learning and the replacement of human labour: anti-Cartesianism versus Babbage's path. AI & Soc 37(4):1459–1471. https://doi.org/10.1007/s00146-021-01264-3

Trotta A, Ziosi M, Lomonaco V (2023) The future of ethics in AI: challenges and opportunities. AI & Soc 38(2):439–441

Tubella AA, Theodorou A, Dignum V, Michael L (2020) Contestable black boxes. In: Rules and reasoning: 4th International Joint Conference, RuleML+ RR 2020, Oslo, Norway, June 29–July 1, 2020, Proceedings 4. Springer, pp 159–167

UK Government (2021) Ethics, Transparency and Accountability Framework for Automated Decision-Making. Report. https://www.gov.uk/government/publications/ethics-transparency-and-accountability-framework-for-automated-decision-making/ethics-transparency-and-accountability-framework-for-automated-decision-making. Accessed 14 Dec 2023

Utrecht University (2021) IAMA makes careful decision about deployment of algorithms possible

van Dijck G, Alexandru-Daniel, Snel J, Nanda R (2023) Retrieving relevant EU drone legislation with citation analysis. Drones 7(8):490

van Rooij B (2020) Do people know the law? Empirical evidence about legal knowledge and its implications for compliance. Cambridge University Press, Cambridge

Van Landuyt D, Truyen E, Joosen W (2012) Documenting early architectural assumptions in scenario-based requirements. In: 2012 Joint Working IEEE/IFIP Conference on Software Architecture and European Conference on Software Architecture. IEEE, pp 329–333

Victor D (2018) Oxford comma dispute is settled as maine drivers get \$5 million

Vladeck DC (2014) Machines without principals: liability rules and artificial intelligence. Wash L Rev 89(2014):117

Walmsley J (2021) Artificial intelligence and the value of transparency. AI & Soc 36(2):585–595. https://doi.org/10.1007/s00146-020-01066-z

Wang C, Han B, Patel B, Rudin C (2023) In pursuit of interpretable, fair and accurate machine learning for criminal recidivism prediction. J Quant Criminol 39(2023):519–581. https://doi.org/10.1007/s10940-022-09545-w

Yang C, Liang P, Avgeriou P, Eliasson U, Heldal R, Pelliccione P (2017) Architectural assumptions and their management in industry–an exploratory study. In: Software Architecture: 11th European Conference, ECSA 2017, Canterbury, UK, September 11-15, 2017, Proceedings 11. Springer, pp 191–207

Yeung K (2023) The new public analytics as an emerging paradigm in public sector administration 27(2):1–32. https://doi.org/10.5334/tilr.303

Yeung K, Weller A (2018) How is 'Transparency' understood by legal scholars and the machine learning community? Amsterdam University Press, Amsterdam. https://doi.org/10.1515/9789048550180

Zhang Q, Hong DK, Zhang Z, Chen QA, Mahlke S, Mao ZM (2021) A systematic framework to identify violations of scenario-dependent driving rules in autonomous vehicle software. In: Proceedings of the ACM on measurement and analysis of computing systems 5(2):1–25

Zheng S, Trott A, Srinivasa S, Parkes DC, Socher R (2022) The AI Economist: taxation policy design via two-level deep multiagent reinforcement learning. Sci Adv 8(18):eabk2607