Automated Anonymization of Court Decisions

Facilitating the Publication of Court Decisions through Algorithmic Systems

Kalliopi Terzidou
Doctoral Researcher (LL.M.)
Faculty of Law, Economics and Finance
Department of Law
University of Luxembourg
Luxembourg
kalliopi.terzidou@uni.lu

ABSTRACT

The practice of anonymization of court decisions has been further systematized by EU Member States' courts, after the entry into force of the General Data Protection Regulation and its transposition into national laws. Anonymization of the parties' personal information protects their privacy during the publication of judgments, which is necessary for the scrutiny of the judiciary's reasoning in a given case and the filing of an appeal whenever a party disagrees with the court's reasoning and/or order. European courts have recently resorted to algorithmic approaches to automate the process of anonymization, which can bestow prompt and consistent application of anonymization rules for court administrations to comply with the applicable personal data protection legislation. These automated solutions can also encompass technical and administrative challenges, ranging from re-ιdentification risks that compromise the protection of the parties' personal data to the lack of acceptance of the algorithmic system by court staff during their daily work routine. The present paper reviews current anonymization practices conducted through algorithmic techniques by, first, explaining the legal framework underlying the publication and anonymization of court decisions, second, examining three algorithmic solutions for the anonymization of court decisions by different EU Member States, and third, reflecting on their efficiencies and challenges for court administrations.

CCS CONCEPTS

• **Security and privacy** → *Human and societal aspects of security and privacy* → *Privacy protections*; • **Applied computing** → *Law, social and behavioral sciences* → *Law*; • **Social and professional topics** → *Computing / technology policy* → *Privacy policies*; • **Computing methodologies** → *Artificial intelligence* → *Natural language processing*.

KEYWORDS

Algorithms, Anonymization, Courts, Court Decisions, Data Protection, European Union, General Data Protection Regulation, Pseudonymization, Publication of Court Decisions

1 Introduction

Personal data protection rules force court administrations to anonymize the personal data of, at least, the parties to the case, as a precondition for their analogue or digitalized publication.

Anonymization is intrinsically linked to the notion of privacy and private life, which can entail, among others, the "inner circle in which the individual may live his own personal life as he chooses and exclude the outside world[1]." The conduct of individuals' personal life without unwarranted external influences presupposes that they are in control of their personal data, meaning "any information relating to an identified or identifiable natural person," including their choice to publicly release them through published court decisions[2].

Anonymization might seem, at first, to impede the full enjoyment of freedom of information, as part of the personal information in the document is, to some extent, obscured. Freedom of information requires the release of court decisions to the public, so that parties have the opportunity to apply for appeal proceedings, the public can scrutinize the activities of court officials, and legal practitioners can be informed of recent updates in (national) case law. Nevertheless, anonymization can be observed as a compromise between the rights of privacy and freedom of information, since it manages to safeguard the parties' identities without impeding access to court decisions, the main interest of which lies in the reasoning and decision of the judge(s)[3].

The effectiveness of anonymization practices might be complemented by algorithmic solutions that (semi-) automate previously manual efforts by court clerks. AI-based systems using Natural Language Processing (NLP) techniques (perhaps consisted of or combined with Machine Learning techniques) attract the attention of, among others, public administrations due to their partial autonomy from human operators in implementing anonymization rules on marked entities within the court decision's text[4]. These applications promise to aid court staff by minimizing the time and effort spent in the manual processing of judgments for anonymization purposes, while consistently applying the given anonymization rules throughout the court decision. Still, in this preliminary stage of the development and use of these algorithmic systems, there are several issues to be confronted, such as the insufficient number of machine-readable judgment documents to serve as input for the training of the model and the absence of consistent and homogenous anonymization rules among national courts to achieve a consistent level of personal data protection within the judicial system of a single EU Member State.

The present paper explores the implementation of algorithmic systems for the automation of anonymization practices among EU Member States' courts for the streamline of court administration. The paper initially addresses the regulation of anonymization of the parties' personal information, as a prerequisite for the publication of court decisions in compliance with data protection rules. It then proceeds to the presentation of three algorithmic systems for the automation of the anonymization of court decisions from court administrations in Austria, Finland, and Luxembourg. Information on these three algorithmic solutions has been sourced through desktop research, except for the JUANO system for which additional information was provided through a public presentation of the system given by officials of the Public Prosecutor's Office

of Luxembourg (Parquet general) at University of Luxembourg in 2023. The paper concludes with a discussion over the technical and administrative advantages emerging during the use of these algorithmic systems and the potential issues hindering their successful implementation in court administrations.

2 The Requirement of the Anonymization of Court Decisions

With the advent of the General Data Protection Regulation (GDPR) and its transposition into national laws, national courts across the EU must comply with the requirement of anonymization of personal data within the judgment's document, having a certain margin of appreciation in defining specific rules on the categories of personal data to be anonymized, as well as on the extent, methods, and tools of anonymization. In order to delineate the requirement of anonymization of court decisions, a review of the regulation of the accessibility and (online) publication of court decisions is firstly attempted, in the framework of which anonymization rules and practices are developed by European court administrations, as analyzed later in the section.

2.1 The Regulation of the Publication of Court Decisions

The practice of publication of court decisions has its foundation on the freedom of expression and information, a fundamental right prescribed in several international treaties[5]. Persons must be able to receive and impart information with no interference by public authorities, exceptionally justified under certain conditions, such as when this is prescribed in law and is necessary in a democratic society[6]. European states prescribe the freedom of expression and information in national legislation (primarily in constitutional law), while its scope might be extended to the right of access to public documents. All European states recognize the right to access documents pertaining to an administrative procedure, so that concerned individuals have the opportunity to challenge an administrative act against them according to the principle of due process; however, in certain cases, interested parties might need to justify their interest in order to access official documents in the absence of a Freedom of Information Act or equivalent legislation[7].

By analogy, court documents must be accessible to the wider public, in accordance with the given national legislation(s) prescribing the freedom of information. Yet it is often the case that only participants to the case or legal professionals might have access to the document. From a theoretical perspective, the institutional requirements of the independence and impartiality of the tribunal might require the protection of judges from outside pressures, which can manifest in the indirect influence exerted by the public when endorsing a specific line of reasoning that might induce judges to adopt it during different stages of the trial or different instances of the proceedings. However, the contrary claim

could be argued as well, namely that publicity of court documents might enhance the independence and impartiality of the judiciary, since a transparent court administration can deter court officials from engaging to activities that diverge from their official duties. From a practical perspective, the restriction of the accessibility of court documents might be attributed to organizational reasons, such as the overload of court officials with back-office duties and the lack of an efficient filing system to expedite the processing of judgments for purposes such as publication. These reasons could further explain why few European states have adopted legislative acts regulating the right to access court documents, including Finland (Act 370/2007 on the Publicity of Court Proceedings in General Courts) and Slovenia (Act on the Access to Information of Public Character)[8]. One main difference between the two legislative acts is that the former grants all members of the public the right to access trial documents, unless these are characterized as secret either because they contain sensitive information or because their revelation could harm the interests protected under the relevant secrecy obligation provisions, while the latter requires the establishment of a legitimate interest by the interested person to grant access to the document.

The publication of court decisions as such is not always regulated. On a national level, the requirement of publication is usually foreseen in legislation or guidelines, or it is observed in practice, with higher courts (supreme, constitutional, and higher administrative courts) having a better record in publishing decisions[9]. There is no transnational legislative framework establishing an obligation to publish court decisions, either in analogue or digital format; however, Article 6 (1) of the European Convention of Human Rights prescribes the requirement of the public pronouncement of the judgment as a means against arbitrariness of the judiciary's administration of justice in a specific case, increasing confidence in courts as an institution[10, 11]. Nevertheless, the Court does not literally interpret the term 'public pronouncement' and gives contracting states a margin of appreciation to determine the means of pronouncement according to the "special features of the proceedings in question," which might include publication of the judgment in a publicly accessible (online) repository[12].

Online publication of court decisions is sometimes regulated explicitly by national courts, applying either negative criteria for the selection of decisions of higher courts (all decisions are published with some exceptions) or positive criteria for the selection of decisions of lower courts (decisions are published if they are of interest to the development of law), while some states leave the selection of cases in the discretion of the judiciary[13]. In the absence of an international or regional regulation, the Committee of Ministers of the Council of Europe has issued a recommendation on the processing of court decisions in online databases, proposing, among others, the dissemination of national jurisprudence through automated systems, if necessary through an objective selection of decisions, with regular updating of the database to cover newly pronounced case law (including decisions

on appeals)[14]. The European Commission places similar importance to online interconnected repositories for legislation and case law, as can be demonstrated by public information tools, such as EUR-Lex[15]. In 2020, the Commission issued a Communication on the digitalization of justice in the EU, where it highlights the importance of online registries and databases in facilitating prompt and affordable access to jurisprudential information to the wider public and to legal experts, allowing interconnection with other national and European justice portals, and securing compliance with the 'once only' principle, stating that citizens should have standardized information about a single public authority for consistency and simplification reasons[17, 18]. The Commission also recognized the need for single identifiers regarding a specific case across different national and transnational registries and databases to facilitate case law research and citation, thus releasing the European Case Law Identifier (ECLI) with a recognizable format by all EU Member States[19].

2.2 The Regulation of the Anonymization of Court Decisions

Pseudonymization is described by the GDPR as "the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information," further specifying that the additional information leading to the identification of the person is stored separately in a (technically) safe location (Article 4, para. 5 GDPR). The difference between anonymization and pseudonymization is that the former completely obscures the personal information in an irreversible way, so there is no chance of recovering it, while the latter partially de-identifies a personal datum, with a possibility of tracing it back to its original form. This difference is implied in Recital 26 GDPR, stating that data protection rules do not apply to anonymous information that can no longer identify a natural person, taking into consideration "…all the means reasonably likely to be used…either by the controller or by another person to identify the natural person directly or indirectly," objectively determined according to the cost and time resources required for the identification process. As it is demonstrated in Section 3.1, pseudonymization rather than anonymization is the current practice among EU Member States' courts because the result is reportedly reversible. In this sense, anonymization is used as a buzzword, which might be attributed to the etymology of the Greek word 'ανώνυμος,' meaning the one without a name. For the purposes of the present paper, the term anonymization is used to cover both current practices of pseudonymization and potential practices of anonymization of court decisions.

The regulation of the protection of personal data has imposed requirements on their processing by public and private authorities, such as the existence of specific legal bases for their processing (Article 6, para. 1 GDPR). The processing of the personal data of parties involved to a case can be viewed both as a legal obligation and as a public task, since it is performed by judicial authorities in the context of their official duties and in accordance with

procedural laws and/or (legal) frameworks governing court administration. However, the extent of processing of the personal data can be minimized to exclude their sharing at the time of the publication of the court decision, since the purpose of their processing is primarily the conduct and management of court proceedings which are concluded with the pronouncement of the judgment, notwithstanding a possible limited retention period for administrative purposes, such as the transcription of the trial (Article 5 para. 1 b, c, e GDPR). The minimization of processing can be achieved through the anonymization process, which contributes to the wider transparency and accountability of judicial institutions through the compliant publication of court decision documents. The scrutiny of the judges' professional activity, especially of their reasoning in the application of the relevant law, enables the filing of an appeal against a decision, when there is perceived misapplication of procedural rules by court officials, notwithstanding disciplinary actions against them. The right to be forgotten (Article 17 GDPR) can be viewed as an additional basis for the anonymization of judgments, especially in cases where the subjects to the court proceedings have the right to explicitly state their desire not to have their personal data published along with the rest of the body of the decision. It is worth noting that the Proposal for an AI Act (Preamble 40) refers to the algorithmic anonymization or pseudonymization of court decisions as an example of "purely ancillary administrative activities that do not affect the actual administration of justice in individual cases," thus not classified as 'high-risk' AI systems that would result in the application of a set of requirements pertaining to their development and use[20].

Anonymization of court decisions is treated differently, depending on the EU Member State in question. National jurisdictions sometimes regard anonymization as an obligation under national law regulating the accessibility of court documents. This is the case for states such as Austria and Belgium[21]. Section § 15 (4) of the OGHG - OGH (Austrian Supreme Court) Act characteristically states that "In the judicial decision documentation, names, addresses and, if necessary, other names of places and territories which allow conclusions to be drawn about the case in question shall be anonymised by letters, numbers or abbreviations in such a way that the comprehensibility of the decision is not lost[22]." [unofficial translation] Certain states, such as France[23], Latvia[24], and Slovakia[25], further specify rules on anonymization under the respective legislative documents. However, rules on the anonymization of court decisions are more often included in guideline documents that concern either all or specific judicial institutions, usually higher courts. Czech Republic[26], Denmark[27], and the Netherlands[28] have published rules on the anonymization of court decisions that generally apply to all judicial bodies. Cyprus[29] and Finland[30] are examples of states where higher courts (supreme and higher administrative) have their own, custom rules for the anonymization of their decisions. In its publication policy for decisions, the Supreme Administrative Court of Finland states that "As a rule,

solution descriptions and decisions are published in anonymized form. The names of individuals and other identifying information (customer number, registration number, address) will be removed from them…If necessary, the names are replaced by letter designations or in another way…The names of companies or other entities will not be removed from the material to be published..." [unofficial translation] Finally, Data Protection Authorities in France[31] and Greece[32] have been active in releasing decisions on the protection of personal data in the framework of the publication of court decisions, stressing the obligation of anonymization of personal data and enumerating certain rules to be followed during this process.

As can be observed in the diverse regulation of the anonymization process of court decisions among EU Member States' courts, differences can exist in: (i) the categories of natural or legal persons that must be anonymized, an often-observed practice being the anonymization of the litigants' and witnesses' personal details, unless they are public figures, and not those of legal experts involved in the case (e.g. judges, prosecutors, court clerks, lawyers); (ii) the extent of anonymization, constituting either a rule or an exception and concerning either all judicial institutions or certain types of courts, with possible qualifications depending on the type of case (e.g. cases involving minors or asylum seekers)[33]; (iii) the method of anonymization, with options extending from complete obscurity of the personal data to the use of initials, fake data, or labels, and; (iv) the tools of anonymization, ranging from completely manual to NLP-based software (with additional monitoring by the users)[34].

The lack or variety of legislative instruments for anonymization of court decisions leads to diverse practices, some of which may be more consistent and adequate for personal data protection than others. A harmonized legislative framework for the anonymization of court decisions, at least on a national level given the distinct legal cultures in the EU, could lead to the uniform application of personal data protection provisions and enhance the trust in courts when they treat all cases equally in preparation for their publication, eliminating any doubt regarding the independence and impartiality of the judiciary. Another important advantage of uniform rules is the standardization of oversight practices by court authorities over the processing of personal data within their institutions. According to Article 55 (3) of the GDPR, "supervisory authorities shall not be competent to supervise processing operations of courts acting in their judicial capacity." Preamble 20 of the Regulation specifies that specific judicial bodies should instead be entrusted with the supervision of personal data processing by courts when the latter implement judicial tasks, such as decision-making. Anonymization of court decisions for later publication could be covered by this provision as an activity carried out during the management of court administration, so that the existence of standard anonymization practices and of the respective supervisory procedures across national judicial bodies would guarantee a consistent level of personal data protection compliance in each EU Member State.

3 The Automation of the Anonymization of Court Decisions through Algorithmic Systems

The practice of anonymizing court decisions before their publication has been particularly present among of European courts after the entry into force of the GDPR in 2016. According to the findings of a 2020 study realized under the Directorate-General for Justice and Consumers of the European Commission, 11 out of the 21 EU Member States are pursuing anonymization or pseudonymization projects based on AI technology, targeting not only court decisions but also other court documents pertaining to the organization of the judiciary[35]. The following section highlights three practical cases of algorithmic applications for anonymization of court documents from the judicial systems of Austria, Finland, and Luxembourg, briefly describing their common characteristics, and envisioning their likely efficiencies to court administrations, as well as the challenges to their development and systematic use by court staff.

3.1   Practical Examples from EU Member States

In Austria, the Federal Ministry of Justice, in cooperation with the Federal Computing Center, launched the "Use of artificial intelligence in the anonymization of court decisions" ("Einsatz von künstlicher Intelligenz bei der Anonymisierung von Gerichtsentscheidungen") project to address the burden of clerks when removing personal data from court decisions (mainly Supreme Court decisions) published in the Federal Legal Information System[36]. The solution is based on Machine Learning (ML) models that can identify individuals, organizations, locations, and other metadata by processing a number of manually marked judgments, extracting the relevant metadata, and anonymizing them according to the applicable data protection legislation[37].

In Finland, the ANOPPI project was initiated in 2018 by the Ministry of Justice for the implementation of an AI application with the goal of performing the anonymization and content description of court decisions, among other, and ultimately increase the availability of their digital versions[38]. Regarding the anonymization function, the Ministry explains that the self-learning algorithm automatically identifies and marks personal data in the form of words or phrases that are referring to the same person and suggests a reference for the anonymization of personal information.

In July 2022, the Public Prosecutor's Office (Parquet général) of Luxembourg announced the release of the JUANO application for the pseudonymization of court decisions, realized through ML methods to ensure consistent pseudonymization results of judicial decisions and their greater availability to legal professionals and the public[39]. JUANO is part of the JUPAL - Paperless Justice strategy of the Public Prosecutor's Office, which aims at the efficiency and simplification of professionals' work and, subsequently, to the improvement of the transparency and accessibility of the justice system[40]. As described in the announcement, JUANO uses ML methods that perform a series of interconnected tasks, including the recognition of personal data (entities) in the text and the proposal of pseudonyms (categories) for their replacement, while the court clerk is only required to upload the text of the decision in the dedicated platform and verify the algorithm's recommendation. Following a public presentation of the JUANO system by officials from the Public Prosecutor's Office of Luxembourg, it was reported that the algorithmic system follows specific rules regarding which categories of personal data must be pseudonymized, including the names of defendants, witnesses, and victims, email addresses, and IBAN numbers, and excluding entities such as the State, public and social security institutions, judges, prosecutors, and court clerks.

It can be deduced from the above descriptions that these algorithmic systems for the anonymization or pseudonymization of court decisions have as a goal to identify entities, such as the litigants, and their personal data, such as their names, throughout the court decision document to replace them with generic, non-identifiable terms. To achieve this goal, the texts of judgments are collected and manually tagged to act as input for the algorithmic system, that is processed by the selected technique, reportedly ML[41]. Through the ongoing training of the model with case law texts and the evaluation of its outputs by human operators, with the necessary interventions in the input data or its functioning parameters, the application improves its functioning and can eventually provide more accurate outputs, consisting in anonymization or pseudonymization recommendations which can be verified by court clerks through the court's filing system so the decision can be published in compliance with personal data protection rules.

3.2 The Efficiencies of Automated Anonymization for Court Administrations

Algorithmic anonymization tools can contribute to a better organized and more privacy compliant court administration, encouraging the re-use of court decision documents by interested parties through open data initiatives.

Concerning the administration of courts, the expected advantage is that court clerks could spend less time in reviewing court decisions to mark personal data and anonymize them according to the established rules, permitting them to focus on other administrative tasks. Of course, they would still need to spend some time to verify the recommendations of the algorithm for anonymization of the marked personal data. Furthermore, the processing and storage of anonymized decisions in the e-filing system of the court, if existing, would allow court clerks to easily access and promptly publish them or, if their publication is impeded for certain reasons (for example, the protection of minors), to efficiently manage relevant information requests.

Concerning compliance matters, a sufficiently trained algorithmic system should be capable of consistently and homogenously applying anonymization rules, as prescribed in national laws,

guidelines, or relevant documents. These rules should regulate the categories of personal data to be anonymized, the types of decisions that merit anonymization (according to the type of court and instance of proceedings), the parts of a decision that need to be treated (facts, applicable law, reasoning, decision, and order), and the method of anonymization. The verification of the system's outputs as to the effective obscurity of the targeted personal data ensures that no personal data are wrongly anonymized or left de-anonymized before the publication of the judgment, which would lead to a serious compromise of the protection of the parties' privacy. As analyzed below, certain algorithmic methods can consistently apply an anonymization technique over a limited number of clearly defined variables and in specific contexts, therefore being unable to effectively detect indirect and quasi-identifiers of data subjects throughout different court decision and relevant documents[42].

Closely related to data protection compliance, AI-based anonymization tools can further facilitate compliance with open data policies that require the availability of public data, including court decisions, in online databases, so interested parties are able to download and re-use them (upon request) without or with minimum intellectual property restrictions. AI applications implement the anonymization process in a speedy and consistent manner, thus supplying databases with not only data protection compliant decisions but also decisions whose text is machine-readable (e.g. XML, HTML formats), enabling the re-use of the whole body of the decision or of extracts including specific metadata. In that way, both members of the legal profession and of the public can benefit by the automated solutions, the former by researching past case law to inform their decision-making or strategy in a given case, and the latter by remaining informed on cases that directly or indirectly concern them. The publication of case law metadata in public open data repositories could also benefit legal tech companies in developing their own applications, such as legal search engines which enable professionals to find past case law based on the insertion of key words and other filters in the search engine[43].

3.3 Challenges to the Development and Implementation of Anonymization Systems

Notwithstanding the administrative and compliance efficiencies that AI anonymization applications directly or indirectly afford to court administrations, there are challenges of a technical and administrative nature that need to be addressed for their effective development and implementation.

Technical challenges cover issues inherent to the development of the algorithmic systems, from their training to the production and evaluation of their outputs. Concerning the training of the system, a sufficient number of court decisions should be available for the gradual improvement of the accuracy of the model's anonymization recommendations. The decisions should further have a machine-readable format (for example, XML markup language), which is important for the system to be trained with domain-specific

language, given the legal terminology and the characteristic structure of court decisions[44]. A machine-readable format further enables developers to focus the training of the system to specific parts of the decisions that contain personal data, excluding sections that are deprived from personal information, such as that of the applicable law. Scanned versions of court decisions stored in a PDF, Word, or similar format do not suffice for the processing of the text by the algorithm.

The ineffective training of the system due to the above-mentioned issues can lead to further technical malfunctions, an important one being its inability to correctly anonymize a personal datum, depending on the applicable anonymization rules. One way in which this can manifest is when the system cannot recognize the role that the identified entity has during the court proceedings. The Federal Computing Center of Austria gives the example of the name of the judge that does not merit anonymization as per the applicable rules, even though it also constitutes personal information[45].

A different manifestation can be the recognition of a single entity as two different ones due to the diverse presentation of its name across the document. For example, if the applicant John Smith is presented in one part of the document as 'JOHN SMITH' and in another as 'SMITH JOHN' or with his initials 'J.S.' or 'S.J,' the system might recognize the different entries as different entities, applying different anonymization methods to the same entity. During the presentation of the JUANO system by officials from the Public Prosecutor's Office of Luxembourg, several points of improvement were suggested regarding the preparation of the decision's texts, including checking spaces between words and their spelling, ensuring the consistent order of people's first and last names throughout the document, and limiting the use of uncommon symbols and special characters. Another reported solution by representatives of the Federal Ministry of Justice of Austria is the training of ML models to process specific language resources, combined with the use of a "quantitative benchmarking tool" to assess the results of the re-training of the model[46].

A final technical consideration is the risk of re-identification of personal data, which defeats the purpose of anonymization in protecting the parties' personal information. Re-identification can occur in different ways, including: the linking of two or more records (information pieces) within the document regarding the same data subjects, such as their name, date of birth, and address[47]; the linking of an anonymized with a de-anonymized dataset or with publicly available information, even outside the court decision, or; the exploitation of the low uniqueness of the record in a single dataset, meaning its higher occurrence within the court decision[48]. Officials from the Public Prosecutor's Office of Luxembourg stated that obscuring metadata from the pseudonymized court decision's file that is uploaded in the court's website minimizes the risk of re-identification of personal data within its corpus. Alternative approaches to the mentioned pseudonymization techniques that might prevent re-identification incidents include encryption and tokenization, as well the

modification of the veracity of data to severe any reference to a data subject (randomization), which can be combined with generalization of the subjects' attributes by altering "the respective scale or order of magnitude[49]."

As it was stated in reference to Recital 26 GDPR, the irreversibility of anonymization should be objectively assessed according to the means reasonably likely to be used by the controller or third parties for the identification of natural persons, which is difficult to determine in a single point in time given the exponential evolution of (AI) technology. The Article 29 Data Protection Working Party lists several key factors to the irreversibility of anonymization, which can be summarized in that the data controller, as the entity that "determines the purposes and means of the processing of personal data" (Article 4, point 7 GDPR), should balance the available resources for anonymization (time, effort, costs, know-how) with the available technical means and public datasets contributing to the identification of natural persons[50]. The anonymization process should not be limited to directly identifiable elements but should also extend to indirect ones to achieve an irreversible result, depending on the context and purposes of the processing, as well as on the existing regulations on personal data protection. In the context of court administration and regarding the purpose of anonymization for the compliant publication of judgments, courts, as the controllers of the parties' personal data, should assess their resources (which are not always available due to the increased backlog, limited funding, and lack of personnel) and design technological solutions for the automation of the anonymization process, consulting technical experts within or outside judicial institutions. They should, additionally, draft mitigation plans to develop long-term strategies on how to combat risks to the system, relating not only to the re-identification of the anonymized personal data but also to broader cybersecurity aspects.

Given that the above issues are addressed by national court administrations and the respective IT departments, further technical arrangements must follow for the successful implementation of algorithmic systems for the anonymization of court decisions. One of these would be the standardization of the format and of the (symbolic) language of judgments to enable the interoperability of online public databases where their anonymized versions can be found (including court e-filing systems)[51]. This solution would enable court decisions to advance through the anonymization and publication interfaces more efficiently and would maximize the uniform application of anonymization rules and the production of consistent outputs.

A second arrangement would be the implementation of linked data protocols, alluding to the cross-reference of data across different (legal) sources, for instance across different court e-filing interfaces. One perceived benefit of this linking process is the tracking of the flow of cases throughout the different stages of court proceedings, so as to derive conclusions regarding the quality of justice and inform policymaking on matters such as the reduction of reoffending[52]. Furthermore, the linking of machine-readable legal documents and included metadata across databases allows users to efficiently search for legal information, without having to navigate through diverse sources, and to receive highly accurate results as per their queries[53]. Finally, linked data can facilitate the re-use of annotation methods implemented on court documents and increase the efficiency of the anonymization model that can now replicate these methods to tag and anonymize relevant entities within the court decision more quickly and consistently.

Administrative challenges concern the mode and rate of implementation of AI-based anonymization solutions in court administrations. An initial issue would be the acceptance of the system by court staff in the sense of its systematic use during their everyday tasks, for reasons such as the lack of familiarization with the system's functioning. The Federal Ministry of Justice of Austria addressed this issue by directly integrating the system in the digital court filing system and enabling clerks to receive an automatically anonymized version of the judgment after uploading the court decision in the digital anonymization tool, the results of which they can then manually verify[54]. After checking the results, a final version of the anonymized document is produced that can be published directly through the court's e-filing system. Similarly, officials from Public Prosecutor's Office of Luxembourg report that court clerks can automatically generate a Word file of the pseudonymized decision through the JUANO system that is further transformed to a PDF file, an extract of which is later published to the online platform of JUDOC. This online database hosts pseudonymized extracts of decisions from ordinary courts and from the Supreme Court (Cour Supérieure de Justice) that pose a particular legal interest and are pre-selected by magistrates[55]. Regardless of the mode of integration of the system, training sessions could be offered to court staff by technical experts within or outside EU Member States' courts to explain the functioning of the system, so as to render its use effortless on an everyday basis.

A second administrative challenge would be the effective cooperation between legal and technical experts while developing the anonymization tool. It is important that court officials in charge of the administration of the respective court can initiate a productive dialogue with contracted IT experts to communicate basic information regarding the structure of court decisions and key legal terms recurring in these documents, along with the designated anonymization. In turn, IT experts, having gained the necessary domain knowledge, can suggest the most appropriate methods and tools for the efficient anonymization of court decisions, while being available to the court staff to address any emerging issue concerning the operation of the algorithmic system.

4   Final Remarks

The digitalization of anonymization of court decisions is a necessary step towards the improvement of publicity rates of case law in compliance with data protection rules, to widen the citizens' access to information regarding court proceedings and to provide legal professionals with more resources to inform their work. The

use of NLP techniques as such can expedite the process of anonymization of a court decision, albeit court clerks having to still spend some time for the verification of the algorithmic outputs, further guaranteeing the consistent application of the designated anonymization rules throughout all judgments. The machine-readable format of the relevant documents enables their effective processing by the algorithmic system and their release as open data, so interested individuals can re-use them to inform their actions.

The integration of such systems in court administration must be conducted in a safe and controlled manner to guarantee their efficient use by court staff. Court decisions must be rendered machine-readable so developers can train the algorithm with enough input data to induce the production of accurate outputs, in the form of detection and labeling, or other method of anonymization, of personal information of the parties to the proceedings. Cases of re-identification of the anonymized data should also be addressed through the choice of state-of-art software applying robust anonymization methods and the drafting of prevention and mitigation plans to safeguard the parties' personal information from privacy breaches. Court staff should become familiarized with the algorithmic system, which can be achieved through training sessions on the use of the system and through its integration in the court's e-filing system to render it more user-friendly. Clerks should also be in contact with technical experts in charge of the algorithmic system to communicate the necessary domain knowledge during the system's development and any issues they encounter during its use.

ACKNOWLEDGMENTS

REFERENCES

[1] Application no. 76639/11. *Denisov v. Ukraine*. Grand Chamber. 25 September 2018. Retrieved from https://hudoc.echr.coe.int/eng?i=001-186216. para. 96.

[2] European Parliament and Council of the EU. 2016. *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. Retrieved November 28, 2022 from https://eur-lex.europa.eu/eli/reg/2016/679/oj. Article 4 (1).

[3] Council of Europe. 1953. *European Convention on Human Rights - Official texts, Convention and Protocols*. Retrieved January 22, 2022 from https://www.echr.coe.int/Documents/Convention_ENG.pdf. Articles 8 (2) and 10 (2).

[4] Alan Akbik. The Flair NLP Framework. Institut für Informatik. Retrieved July 11, 2022 from https://www.informatik.hu-berlin.de/en/forschung-en/gebiete/ml-en/Flair.

[5] European Commission. 2000. *Charter of Fundamental Rights of the European Union*. Retrieved October 15, 2022 from http://data.europa.eu/eli/treaty/char_2012/oj/eng. Article 11.

[6] Council of Europe. 1953. *European Convention on Human Rights - Official texts, Convention and Protocols*. Retrieved January 22, 2022 from https://www.echr.coe.int/Documents/Convention_ENG.pdf. Article 10.

[7] OECD. 2010. *The Right to Open Public Administrations in Europe: Emerging Legal Standards*. OECD, Paris. DOI:https://doi.org/10.1787/5km4g0zfqt27-en. 10-11.

[8] Policy Department C - Citizens' Rights and Constitutional Affairs European Parliament and Vesna Naglič. 2013. *National Practices with Regard to the Accessibility of Court Documents*. Retrieved February 4, 2023 from https://www.europarl.europa.eu/thinktank/en/document/IPOL-JURI_ET(2013)474406. 21-29.

[9] Marc van Opijnen, Ginevra Peruginelli, Eleni Kefali, and Monica Palmirani. 2017. On-Line Publication of Court Decisions in the EU: Report of the Policy Group of the Project 'Building on the European Case Law Identifier.' *Rochester, NY*. Retrieved February 3, 2023 from https://papers.ssrn.com/abstract=3088495. 11-13.

[10] Application no. 40908/05. *Fazliyski v. Bulgaria*. Fourth Section. 16 April 2013. Retrieved from https://hudoc.echr.coe.int/eng?i=001-118573. para. 69.

[11] Application no. 7984/77. *Pretto and Others v. Italy*. Court (Plenary). 08 December 1983. Retrieved from https://hudoc.echr.coe.int/eng?i=001-57561. para. 21.

[12] Application no. 7984/77. *Pretto and Others v. Italy*. Court (Plenary). 08 December 1983. Retrieved from https://hudoc.echr.coe.int/eng?i=001-57561. para. 26.

[13] Marc van Opijnen. 2016. Court Decisions on the Internet: Development of a Legal Framework in Europe. *Journal of Law, Information and Science* (2016). Retrieved February 7, 2023 from http://classic.austlii.edu.au/au/journals/JlLawInfoSci/2016/3.html. 5-7.

[14] Council of Europe Committee of Ministers. 1995. *Recommendation No. R (95) 11 concerning the selection, processing, presentation and archiving of court decisions in legal information retrieval systems*. Retrieved February 6, 2023 from https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=09000016804f120c. Appendix I. Section III.

[15] EUR-Lex website. Retrieved from https://eur-lex.europa.eu/homepage.html.

[16] European e-Justice Portal website. Retrieved from https://e-justice.europa.eu/home?action=home&plang=en.

[17] European Parliament. 2016. *Regulation (EU) 2018/1724 of the European Parliament and of the Council of 2 October 2018 establishing a single digital gateway to provide access to information, to procedures and to assistance and problem-solving services and amending Regulation (EU) No 1024/2012 (Text with EEA relevance.)*. Retrieved February 7, 2023 from http://data.europa.eu/eli/reg/2018/1724/oj/eng. Article 1 (1) (a).

[18] European Commission. 2020. *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions l Digitalisation of justice in the European Union - A toolbox of opportunities*. Retrieved February 7, 2023 from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0710. 13-14.

[19] European Commission. European e-Justice Portal - European Case Law Identifier (ECLI). Retrieved February 7, 2023 from https://e-justice.europa.eu/content_european_case_law_identifier_ecli-175-en.do.

[20] European Commission. 2021. *Proposal for a Regulation of the European Parliament and the Council laying down harmonised rules of Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*. Retrieved October 19, 2022 from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206. Articles 8-15.

[21] *Loi n° 2005009655 du 10/08/2005 instituant le système d'information Phenix (1)*. Retrieved from http://www.ejustice.just.fgov.be/cgi_loi/loi_a1.pl?language=fr&la=F&cn=2005081057&table_name=loi&&caller=list&F&fromtab=loi&tri=dd+AS+RANK&rech=1&numero=1&sql=(text+contains+(%27%27))#Art.9. Article 9.

[22] Bundesminister für Justiz. 1968. *Bundesgesetz vom 19. Juni 1968 über den Obersten Gerichtsho*. Retrieved from https://www.jusline.at/gesetz/oghg/paragraf/15. § 15 (4) OGHG Entscheidungsdokumentation Justiz; Service Public Federal Justice. 2005.

[23] L'Assemblée nationale et le Sénat. 2019. *Loi n° 2019-222 du 23 mars 2019 de programmation 2018-2022 et de réforme pour la justice (1)*. Retrieved from https://www.legifrance.gouv.fr/jorf/article_jo/JORFARTI000038261761. Article 33.

[24] Ministru kabineta noteikumi Nr.123. 2009. *Noteikumi par tiesu informācijas publicēšanu mājaslapā internetā un tiesu nolēmumu apstrādi pirms to izsniegšanas*. Retrieved from https://likumi.lv/ta/id/187832-noteikumi-par-tiesu-informacijas-publicesanu-majaslapa-interneta-un-tiesu-nolemumu-apstradi-pirms-to-izsniegsanas. Article 12.

[25] Ministerstva spravodlivosti Slovenskej republiky. 2011. *Vyhláška č. 482/2011 Z. z. Vyhláška Ministerstva spravodlivosti Slovenskej republiky o zverejňovaní súdnych rozhodnutí*. Retrieved from https://www.zakonypreludi.sk/zz/2011-482#p4. Articles 4-5.

[26] Ministerstva spravedlnosti. 2002. *(1/2002 MS) Instrukce, Ministerstva spravedlnostize dne 3. prosince 2001, č. j. 505/2001–Org, kterou se vydává*

vnitřní a kancelářský řád pro okresní, krajské a vrchní soudy. Retrieved from https://www.nkcr.cz/data/predpisy/VKR_k_12_3_2018.pdf.

[27] Justitsministeriet. 1988. *CIR nr 85 af 08/07/1988. Cirkulære om indlæggelse af afgørelser i Retsinformation.* https://www.retsinformation.dk/eli/mt/1988/85. Articles 2-7.

[28] de Rechtspraak, *Anonimiseringsrichtlijnen*. Retrieved from https://www.rechtspraak.nl/Uitspraken/Paginas/Anonimiseringsrichtlijnen.a spx.

[29] Ανώτατο Δικαστήριο. 2018. *Εγκύκλιος-Οδηγίες/Ανωνυμοποίηση δικαστικών αποφάσεων.* Retrieved from http://www.supremecourt.gov.cy/Judicial/sc.nsf/All/B450E5061E647886C2 25837C00309575?OpenDocument.

[30] Korkein hallinto-oikeus. 2019. *Päätösten julkaisukäytäntö*. Retrieved from https://www.kho.fi/fi/index/paatokset/julkaisuohje.html.

[31] Commission Nationale de l'Informatique et des Libertés. 2001. *Délibération 01-057 du 29 novembre 2001.* Retrieved from https://www.legifrance.gouv.fr/cnil/id/CNILTEXT000017653503/#:~:text= Une%20pr%C3%A9caution%20minimale%20%C3%A0%20l,rendus%20li brement%20accessibles%20sur%20Internet.

[32] Αρχή Προστασίας Δεδομένων. 2000. *Ανακοίνωση αποφάσεων του ΣτΕ σε τρίτους.* Retrieved from https://www.dpa.gr/el/enimerwtiko/prakseisArxis/anakoinosi-apofaseon- toy-ste-se-tritoys.

[33] Directorate-General for Library, Research and Documentation. Research Note. 2017. *Anonymity of the parties on the publication of court decisions*. Retrieved from https://curia.europa.eu/jcms/upload/docs/application/pdf/2021-02/ndr_2017- 002_neutralisee-en.pdf. 1-7.

[34] Marc van Opijnen, Ginevra Peruginelli, Eleni Kefali, and Monica Palmirani. 2017. *On-Line Publication of Court Decisions in the EU: Report of the Policy Group of the Project 'Building on the European Case Law Identifier.'* Rochester, NY. Retrieved February 3, 2023 from https://papers.ssrn.com/abstract=3088495. 24-26.

[35] Directorate-General for Justice and Consumers (European Commission) and Trasys International. 2020. *Study on the use of innovative technologies in the justice field: final report*. Publications Office of the European Union, LU. Retrieved January 24, 2022 from https://data.europa.eu/doi/10.2838/585101. 111-142.

[36] Das Bundesrechenzentrum Kompetenzzentrum für Digitalisierung. 2022. Einsatz von künstlicher Intelligenz bei der Anonymisierung von Gerichtsentscheidungen: Justiz gewinnt eAward 2022. Retrieved January 30, 2023 from https://www.brz.gv.at/presse/eaward2022.html.

[37] Das Bundesrechenzentrum Kompetenzzentrum für Digitalisierung. 2022. Mein Name ist Nym, Ano Nym. Bundesrechenzentrum. Retrieved January 30, 2023 from https://www.brz.gv.at/was-wir-tun/Innovationen/Mein-Name- ist-Nym--Ano-Nym.html.

[38] Ministry of Justice of Finland. 2018. Anoppi project. Retrieved January 25, 2022 from https://oikeusministerio.fi/en/project?tunnus=OM042:00/2018.

[39] Communiqué du Parquet général. 2022. Lancement d'une application pour faciliter la pseudonymisation de décisions judiciaires. Retrieved July 11, 2022 from http://justice.public.lu/fr/actualites/2022/07/communique-parquet- general-juano.html.

[40] Portail des marchés publics du Grand-Duché de Luxembourg. 2019. *Détails de la consultation, Paperless Justice, JUPAL Rapport Final.* Retrieved from https://pmp.b2g.etat.lu/index.php?page=entreprise.EntrepriseDemandeTelec hargementDce&id=517006&orgAcronyme=t5y. 1.

[41] Ministry of Justice of Austria. Martin Hackl and David Steinbauer. 2021. Anonymization of court decisions in Austria. Webinar on the use of AI in the justice field. 29 March 2021. Retrieved from https://commission.europa.eu/system/files/2021- 04/anonymisation_webinar_29032021_austria.pdf.

[42] European Data Protection Supervisor and Agencia Española de Protección de Datos. 2021. *10 Misunderstandings Related to Anonymisation,* Retrieved from https://edps.europa.eu/system/files/2021-04/21-04-27_aepd- edps_anonymisation_en_5.pdf. 6.

[43] Predictice. Predictice pour les avocats individuels. Retrieved February 4, 2023 from https://blog.predictice.com/avocat-individuel.

[44] Marie-Francine Moens and RIK De busser. 2002. First steps in building a model for the retrieval of court decisions. *International Journal of Human- Computer Studies* 57, 5 (November 2002), 429–446. DOI:https://doi.org/10.1006/ijhc.2002.1029. 434-435.

[45] Das Bundesrechenzentrum Kompetenzzentrum für Digitalisierung. 2022. Mein Name ist Nym, Ano Nym. Bundesrechenzentrum. Retrieved January 30, 2023 from https://www.brz.gv.at/was-wir-tun/Innovationen/Mein-Name- ist-Nym--Ano-Nym.html.

[46] Ministry of Justice of Austria. Martin Hackl and David Steinbauer. 2021. Anonymization of court decisions in Austria. Webinar on the use of AI in the justice field. 29 March 2021. Retrieved from

https://commission.europa.eu/system/files/2021- 04/anonymisation_webinar_29032021_austria.pdf.

[47] Alex Hern. 2019. "Anonymised" Data Can Never Be Totally Anonymous, Says Study. The Guardian. 23 July 2019. sec. Technology. Retrieved from https://www.theguardian.com/technology/2019/jul/23/anonymised-data- never-be-anonymous-enough-study-finds.

[48] P.L.M Kelani Bandara, HMN Dilum Bandara, and Shantha Fernando. 2020. Evaluation of Re-identification Risks in Data Anonymization Techniques Based on Population Uniqueness. In *2020 5th International Conference on Information Technology Research (ICITR)*. 1–5. DOI:https://doi.org/10.1109/ICITR51448.2020.9310884. 2.

[49] Article 29 Data Protection Working Party. 2014. *Opinion 05/2014 on Anonymisation Techniques.* 0829/14/EN WP216. 10 April 2014. Retrieved from https://ec.europa.eu/justice/article-29/documentation/opinion- recommendation/files/2014/wp216_en.pdf. 12-23.

[50] Article 29 Data Protection Working Party. 2014. *Opinion 05/2014 on Anonymisation Techniques.* 0829/14/EN WP216. 10 April 2014. Retrieved from https://ec.europa.eu/justice/article-29/documentation/opinion- recommendation/files/2014/wp216_en.pdf. 8-10.

[51] European Commission. 2012. *Regulation (EU) No 1025/2012 of the European Parliament and of the Council of 25 October 2012 on European standardisation, amending Council Directives 89/686/EEC and 93/15/EEC and Directives 94/9/EC, 94/25/EC, 95/16/EC, 97/23/EC, 98/34/EC, 2004/22/EC, 2007/23/EC, 2009/23/EC and 2009/105/EC of the European Parliament and of the Council and repealing Council Decision 87/95/EEC and Decision No 1673/2006/EC of the European Parliament and of the Council Text with EEA relevance*. Retrieved February 4, 2023 from http://data.europa.eu/eli/reg/2012/1025/oj/eng. Preambles 1, 33, and 34.

[52] Ministry of Justice of the United Kingdom. 2022. Data First: Criminal Courts Linked Data. 10 March 2022. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads /attachment_data/file/1058219/data-first-criminal-courts-linked-data.pdf. 14-23.

[53] Erwin Filtz, Sabrina Kirrane, and Axel Polleres. 2021. The linked legal data landscape: linking legal data across different countries. *Artif Intell Law* 29, 4 (December 2021), 485–539. DOI:https://doi.org/10.1007/s10506-021- 09282-8. 486-487.

[54] Das Bundesrechenzentrum Kompetenzzentrum für Digitalisierung. 2022. Mein Name ist Nym, Ano Nym. Bundesrechenzentrum. Retrieved January 30, 2023 from https://www.brz.gv.at/was-wir-tun/Innovationen/Mein-Name- ist-Nym--Ano-Nym.html.

[55] Parquet général de Luxembourg. La Justice du Grand-Duché de Luxembourg - Base de Jurisprudence JUDOC. https://justice.public.lu/fr/jurisprudence/jurisprudence-judoc.html.