EJLT European Journal of Law and Technology

The Challenges of Personal Data Processing in Developing AI-Driven Tools for Judicial Proceedings in the EU: The Example of CREA2

Lana K. Gotvan, Daša Tičar, Katarina Zajc[*]

Abstract

This article examines the challenges of processing personal data in the development of tools driven by artificial intelligence (AI) for judicial proceedings in the European Union (EU), as encountered during the CREA2 project. The article underscores the importance of access to judicial decisions as essential inputs for such tools, while also addressing concerns related to the processing of personal data contained within these inputs. Following the introduction, Section 2 provides a brief overview of the relevant EU data protection regulatory framework, with a focus on key provisions of the General Data Protection Regulation (GDPR). Section 3 further explores the GDPR by examining the concepts of data anonymisation and pseudonymisation, particularly in the context of privacy protection. Section 4 analyses the intersection of data protection and AI, highlighting the specific challenges encountered during the data collection phase of the CREA2 project. Finally, Section 5 discusses future developments and Section 6 summarises the key issues discussed and presents concluding remarks.

Keywords: data protection, GDPR, AI-driven tools, anonymisation, pseudonymisation.

[*] At the time of writing, all three authors were affiliated with the Faculty of Law, University of Ljubljana.

1. Introduction

This article addresses the challenges of processing personal data when developing tools, driven by AI,[1] for judicial proceedings in the EU. The use of AI has expanded across various sectors, including the judiciary, sparking vibrant discussions by both proponents and sceptics.[2] The European Commission actively promotes the development and use of AI through funding various projects. One example is the recent CREA2 project,[3] which aims to introduce algorithms to assist natural and legal persons in resolving disputes, particularly in inheritance and divorce cases when dividing assets.

To ensure that AI-driven tools produce relevant and accurate outputs for judicial decision-making, they must be fed reliable inputs that fully capture the details of the case and the applicable legal framework. The model developed for the CREA2 project was designed to accurately predict court decisions on asset division in inheritance and divorce cases. Its primary goal was to support settlements by providing a benchmark that reflects the likely judicial outcome in the case at hand. Additionally, the model aimed to estimate the costs and duration of the proceedings. The authors contributed to the model's development by extracting key inputs from Slovenian judicial decisions, including case outcomes (i.e., asset division and court allocation), legal grounds for the rulings, as well as data on costs and timelines.

---

[1] For a more detailed explanation of AI, see for example ISO, 'ISO/IEC 22989:2022 – Information Technology – Artificial Intelligence – Artificial Intelligence Concepts and Terminology' <https://www.iso.org/standard/74296.html> accessed 21 June 2024; Yongjun Xu and others, 'Artificial Intelligence: A Powerful Paradigm for Scientific Research' (2021) 2 *The Innovation* 100179.

[2] Alfonso Renato Vargas-Murillo and others, 'Transforming Justice: Implications of Artificial Intelligence in Legal Systems' (2024) 13 *Academic Journal of Interdisciplinary Studies* 433; Dovilė Barysė and Roee Sarel, 'Algorithms in the Court: Does It Matter Which Part of the Judicial Decision-Making Is Automated?' (2024) 32 *Artificial Intelligence and Law* 117; Vasiliy A Laptev and Daria R Feyzrakhmanova, 'Application of Artificial Intelligence in Justice: Current Trends and Future Prospects' (2024) H*uman-Centric Intelligent System*s 394 <https://link.springer.com/10.1007/s44230-024-00074-2> accessed 21 June 2024; Valentina Aleksandrovna Rodikova, 'Artificial Intelligence vs. Judicial Discretion: Prospects and Risks of Judicial Practice Automation' (2023) 4 *Legal Issues in the Digital Age* 59; Francesco Contini, 'Artificial Intelligence and the Transformation of Humans, Law and Technology Interactions in Judicial Proceedings' (2020) 2 *Law, Technology and Humans* 4; Yadong Cui, *Application of AI in Judicial Practice* (Springer Nature, 2020) 21; Dorottya Papp, Bernadett Krausz and Franciska Gyuranecz, 'The AI Is Now in Session – The Impact of Digitalisation on Courts' (2022) 7 *Cybersecurity and Law* 272; AV Makutchev, 'Modern Possibilities and Limits of Artificial Intelligence Introduction into the Justice System' (2022) 17 *Actual Problems of Russian Law* 47; Paweł Marcin Nowotko, 'AI in Judicial Application of Law and the Right to a Court' (2021) 192 *Procedia Computer Science* 2220.

[3] More about the CREA2 project can be found at European Commission, 'Conflict Resolution with Equitative Algorithms 2 (CREA2)' (EU Funding & Tenders Portal) <https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/how-to-participate/org-details/999999999/project/101046629/program/43252386/details> accessed 21 June 2024.

This article critically examines the legal and technical tensions between data utility and privacy in AI-supported judicial tools, using the CREA2 project in Slovenia as a case study. Specifically, the challenge lies in determining which data can be processed without breaching regulatory standards while still enabling AI-driven tools to produce useful outputs. The processing of personal data in the EU must adhere to the GDPR[4] (the exact material scope is given in Article 2 GDPR) and any relevant national legislation. The GDPR applies to AI systems whenever personal data is involved (including storage), due to its broad definition of 'processing'. Thus, in some instances, complying with data protection laws may limit the full development and usefulness of AI-driven tools.

Following the introduction, Section 2 of this article presents the EU data protection framework (which is also applicable in Slovenia). We also explore the specific Slovenian legal landscape, where we conducted our research, focusing on the issue of the inaccessibility of judicial decisions. Within the international CREA2 project, a comparative perspective revealed that the Slovenian inaccessibility of first-instance judicial decisions poses a relatively unique barrier to the development of AI-driven tools. Thus, this article also serves as a case study, highlighting a key obstacle that developers of AI-based solutions may encounter.

Section 3 focuses on the key aspects of the GDPR, including what constitutes personal data under the GDPR and explores the principles of privacy by design and privacy by default. We then discuss data anonymisation and pseudonymisation, two tools that have different implications for data protection.

Section 4 discusses the intersection between data protection and AI, as well as the data protection challenges encountered during the data collection phase of the CREA2 project. While we argue that all judicial decisions should be anonymised and made public, achieving effective anonymisation can be challenging, especially in a small country like Slovenia. Thus, other measures have to be taken to ensure compliance with the GDPR, for example, pseudonymisation. Section 5 discusses possible future developments, while Section 6 concludes and summarises the most pertinent issues raised.

2. Protection of Personal Data in Judicial Proceedings

In developing the AI-driven tool for the CREA2 project, it was essential to feed the model with both the applicable legislation and relevant case law related to asset division in inheritance cases and divorce cases. To enable the AI-driven tool to generate meaningful and autonomous outcomes after training, it first had to be

[4] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons regarding the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

provided with judicial decisions and the underlying case facts, in addition to the relevant legal provisions governing the case.

Thus, when collecting input data from relevant first-instance judicial decisions in Slovenia, we were faced with the question of the nature of the data to be processed by the AI-driven tool. This issue was raised because first-instance judicial decisions are not readily available or published on the internet in Slovenia (as is further discussed in Section 2.2). First-instance decisions were required as inputs because they contain data on the facts of the case, i.e. the description of the assets to be divided as well as their allocation by the court. In contrast, appeal decisions, which are publicly available online in Slovenia, rarely disclose the facts of the case as they generally focus on questions of the law.

The following subsections first briefly present the relevant data protection framework in the EU (measures for ensuring data protection are further discussed in Section 3). We then discuss the Slovenian landscape and the issue of the inaccessibility of judicial decisions. These two sections provide the context for the challenges we encountered during the data collection phase of the CREA2 project.

2.1 Framework and Definitions

Protection of personal data is a fundamental right, embedded in Article 8 of the Charter of Fundamental Rights of the European Union (CFR).[5] This right is closely connected to the right to respect for private and family life, codified in Article 7 CFR.[6] Additionally, the Treaty on the Functioning of the European Union (TFEU) safeguards this right in Article 16(1). The EU has established a wide regulatory framework to ensure the protection of personal data,[7] with the GDPR leading this effort. Additionally, national legislation governs the processing of personal data within

[5] '1. Everyone has the right to the protection of personal data concerning him or her. 2. Such data must be processed fairly for specified purposes and based on the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified. 3. Compliance with these rules shall be subject to control by an independent authority.'

[6] Juliane Kokott and Christoph Sobotta, 'The Distinction between Privacy and Data Protection in the Jurisprudence of the CJEU and the ECtHR' (2013) 3 *International Data Privacy Law* 222.

[7] For example, Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications); Directive 2002/58/EC was amended by Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC; 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 and others.

individual Member States. As the Slovenian legal framework[8] is based on the GDPR, we mostly focus on the EU framework.[9]

Personal data are defined in Article 4(1) GDPR as any information relating to an identified or identifiable natural person (one who can be identified, directly or indirectly). Thus, any data clearly associated with a particular person are considered personal, i.e., anything regarding their name, appearance, identity, etc.[10] This broad definition ensures comprehensive protection of individuals' privacy rights, reflecting the Regulation's robust approach to data protection. The general rules outlined in Article 6 GDPR delineate six lawful bases for the processing[11] of personal data. Thus, judicial decisions can be published based on the public interest (Article 6(1)(e) GDPR) in enhancing public trust in the justice system and improving the quality of judicial proceedings and decisions.

2.2 Publicity of Judicial Decisions in Slovenia

For the development of the AI-driven tool as part of the CREA2 project, we needed to collect data from Slovenian judicial decisions relating to inheritance and divorce proceedings. In Slovenia, judicial decisions are considered public information.[12] However, there are hurdles in accessing first-instance judicial decisions, which are not published (while second- and third-instance decisions are available online). Thus, for unavailable decisions, one has to request access based on Article 12 of the Slovenian Access to Public Information Act.[13] Therefore, while information of a public nature (which includes first-instance judicial decisions) is freely accessible to any legal or natural person, it must be requested. Such a request takes time[14] and effort both for

[8] The main legal act governing data protection in Slovenia is the Personal Data Protection Act (Zakon o varstvu osebnih podatkov (ZVOP-2), Official Journal of the Republic of Slovenia 163/22).

[9] Unless otherwise specified, the terms used throughout this article carry the meanings defined in the GDPR.

[10] Full definition: '"personal data" means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.'

[11] Article 4(2) GDPR defines processing as: 'any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction'.

[12] Article 4 of the Slovenian Access to Public Information Act (Zakon o dostopu do informacij javnega značaja (ZDIJZ), Official Journal of the Republic of Slovenia 51/06).

[13] Zakon o dostopu do informacij javnega značaja (ZDIJZ), Official Journal of the Republic of Slovenia 51/06.

[14] According to Articles 23 and 24 of the Slovenian Access to Public Information Act, the institution (the court in this case) must decide on the applicant's request for access to public information without delay, and at the latest within 20 working days from the date of receipt of

the requesting party as well as for the courts. Scholars and the general public have criticised this system for years since it not only hinders transparency and the public's overview of the functioning of the courts, but is also detrimental for practising lawyers who have difficulties keeping up with legal developments in the interpretation and application of the law. Furthermore, it leads to inconsistent case law across the country.[15]

A more desirable approach to the accessibility of judicial decisions can, for example, be seen in Germany, where the Federal Court (*Bundesgerichtshof*) has firmly held that courts are typically obliged to publish their decisions in civil cases at least in an anonymised form. The Court has held that this is part of the judiciary's public duty in disseminating case law.[16] Similarly, the EU court system has an established database of reported decisions dating back to 1954.[17] We argue that a similar approach should be taken in Slovenia. While our research under the CREA2 project was encumbered due to the inaccessibility of judicial decisions,[18] we make a broader argument here in support of the wide availability of judicial decisions. All judicial decisions should be published online,[19] as this significantly enhances public access to information, ensures consistent case law and creates a level playing field.[20]

Despite the difficulties in accessing first-instance judicial decisions in Slovenia, there have been some positive developments in this field. The Slovenian Ministry of Justice has recently proposed a piece of legislation mandating the online publication of all decisions, including those of first-instance courts, in an anonymised form. Exceptions

---

the complete request. This deadline may be extended by a maximum of 30 working days if the court needs more time to provide the requested information in order to carry out partial access to public information (see Article 7 of the same Act). According to Article 25 of the Slovenian Access to Public Information Act, the court must make the content of the requested information available to the applicant without delay if it grants the request. It does so either by making it available for inspection or by providing a copy, photocopy or electronic record of it.

[15] Gregor Zagozda, 'Vrhovno sodišče RS v čudežni deželi' (IUS-INFO, 17 November 2020) <https://www.iusinfo.si/medijsko-sredisce/v-srediscu/274133> accessed 22 June 2024; Nina Betetto, 'Usklajenost sodne prakse na višjih sodiščih v civilnih zadevah' (IUS-INFO) <https://www.iusinfo.si/literatura/L030Y2004V7P1072N1> accessed 22 June 2024.

[16] See decision of the German Federal Court (Bundesgerichtshof) Az. IV AR (VZ) 2/16 of 5 April 2017.

[17] Court of Justice of the European Union, Case Law Search (Curia, 2025) <https://curia.europa.eu/juris/recherche.jsf?cid=4557189> accessed 26 March 2025.

[18] We would like to thank the District and Local Courts of Ljubljana for their willingness to assist us in our research.

[19] With exceptions to be made, for example, for the protection of minors.

[20] Lord Justice Brooke, 'Publishing the Courts: Judgments and Public Information on the Internet - Lord Justice Brooke (2003)' (ICLR, 4 February 2018) <https://www.iclr.co.uk/blog/archive/publishing-the-courts-judgments-and-public-information-on-the-internet-lord-justice-brooke-2003/> accessed 6 June 2024.

to this rule, such as the protection of minors, would be defined in statute.[21] Nevertheless, it is unclear when the duty of online publication will be adopted and how long it will take to set up a comprehensive database of all judicial decisions. The online publication of all first-instance judicial decisions will enhance transparency, ensure consistent case law and facilitate the development of AI-driven tools for judicial proceedings.

3. Data Protection Regulation in the EU

The CREA2 project required careful application of data protection principles to enable the lawful use of judicial decisions for training the AI-driven tool. This section explores the legal and technical mechanisms employed to achieve this.

We explore data protection through the principles of privacy by design and privacy by default, and discuss data anonymisation and pseudonymisation. We focus on anonymisation and pseudonymisation because these were the two appropriate measures utilised by the courts during the data-gathering stage of the CREA2 project.

3.1 Data Protection by Design and Data Protection by Default

According to the preamble of the GDPR, the rights and freedoms of natural persons require that appropriate technical and organisational measures be taken to ensure compliance with the Regulation's requirements.[22] Therefore, controllers[23] and processors[24] must implement measures that ensure data protection by design and data protection by default – complementary concepts that mutually reinforce one another.[25] The notions of data protection by design and data protection by default are briefly reviewed below. Both notions were relevant for developing the AI-driven tool for the CREA2 project, to make sure it complied with data protection regulations. For our data collection, anonymisation and pseudonymisation were the most relevant appropriate measures to ensure data protection by design, as is further discussed in Sub-sections 3.2 and 3.3.

---

[21] Ministrstvo za pravosodje, 'Prenova sodniške zakonodaje in dostop javnosti do sodnih odločb' (Portal GOV.SI, 14 November 2023) <https://www.gov.si/novice/2023-11-14-prenova-sodniske-zakonodaje-in-dostop-javnosti-do-sodnih-odlocb/> accessed 22 June 2024.

[22] Recital 78 GDPR.

[23] Article 4(7) GDPR defines a 'controller' as: 'natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law.'

[24] Article 4(8) GDPR defines processor as: 'a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller'.

[25] European Data Protection Board, 'Guidelines 4/2019 on Article 25 Data Protection by Design and by Default Version 2.0' 6 <https://www.edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_201904_dataprotection_by_design_and_by_default_v2.0_en.pdf> accessed 22 June 2024.

Data protection by design is governed by Article 25(1) GDPR and mandates that controllers implement appropriate technical and organisational measures as well as integrate necessary safeguards into their processing activities to protect the rights of data subjects. Appropriate measures and necessary safeguards should be interpreted broadly, encompassing any method the controller may use in data processing. A measure is deemed appropriate if it implements data protection principles effectively.[26] Data protection principles are provided in Article 5 GDPR and include lawfulness, fairness and transparency, purpose limitation, data minimisation, etc. Examples of appropriate measures and necessary safeguards may include pseudonymisation, technical solutions, personnel training, etc.[27]

Article 25(1) GDPR also outlines the elements the controllers must consider when determining appropriate measures and safeguards. Controllers must consider current technological advancements to meet the 'state of the art' obligation. Additionally, the cost of implementation must be taken into account, ensuring that the controllers do not face disproportionate expenses if alternative measures exist. The controllers must also consider the nature, scope, context and purpose of processing, and conduct a risk analysis beforehand to inform their decisions.[28]

Data protection by default, on the other hand, is governed by Article 25(2) GDPR and mandates that the controllers implement appropriate technical and organisational measures for ensuring that, by default, only necessary personal data are processed, and that such data are not accessible to an indefinite number of people. The term 'by default' refers to the default configuration choices or processing options, such as those in software applications, services, devices or manual processing procedures, that impact the amount of personal data collected, the extent of processing, the period of storage, and the accessibility of the data. The fundamental requirement of Article 25(2) of the GDPR is that data protection is inherently built into the processing activities by default.[29]

The principle of **data minimisation**,[30] which is referred to in Article 25(1)–(2) GDPR, mandates that controllers should not collect more data than necessary to achieve the intended processing purpose. Controllers need to assess the amount and types of personal data collected. Processing should be limited to what is strictly necessary, and data should not be retained longer than needed, with any retention being objectively justifiable by the data controller. Furthermore, controllers must limit access to the data to ensure it is not accessible to unauthorised individuals.[31]

---

[26] Effectiveness is a crucial consideration in developing data protection by design and will depend on the context of the processing in question. Controllers can determine appropriate key performance indicators (to demonstrate this effectiveness. ibid 7.
[27] ibid 5–11.
[28] ibid 7–11.
[29] ibid 11–14.
[30] Article 5(1)(c) GDPR.
[31] European Data Protection Board (n 25) 11–14, 21.

Important design and default data minimisation elements may include pseudonymisation and anonymisation.[32] Both data pseudonymisation and anonymisation are de-identification techniques,[33] which are frequently used to comply with data protection rules. These two tools are especially prevalent in academic research, where personal data are often not necessary for achieving research purposes. Nevertheless, the technical distinctions and legal consequences of the two concepts are often misunderstood; thus, we discuss them further in Sub-sections 3.2. and 3.3 below.

Aside from anonymisation and pseudonymisation, there exist many other appropriate measures and necessary safeguards which may ensure data protection, such as technical solutions or personnel training. Nevertheless, we focus on anonymisation and pseudonymisation as judicial decisions that needed to be anonymised during their collection under the CREA2 project, and courts, performing this anonymisation, were faced with difficulties (discussed further in Section 4). Thus, pseudonymisation presented itself as an alternative feasible de-identification technique for balancing data protection with access to judicial decisions.

3.2 Data Anonymisation

'Anonymisation' is the process of rendering personal data anonymous.[34] According to Recital 26 GDPR, anonymous data is 'information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable'. Importantly, Recital 26 also states that the processing of anonymous information falls outside the scope of the GDPR, including processing for statistical or research purposes. Thus, the GDPR does not have to be considered when processing anonymous data. Anonymisation serves as an alternative to deletion if all relevant contextual factors are considered and the likelihood and severity of risks, including the risk of reidentification, are regularly evaluated.[35]

For the CREA2 project, the courts sought to provide us with the collected data in an anonymised form to ensure that the inputs fed to the AI-driven tool would not contain any personal data, thus rendering the GDPR inapplicable. However, challenges were encountered in fully anonymising the collected data. To discuss this further, we must

---

[32] ibid 21.

[33] Mike Hintze and Khaled El Emam, 'Comparing the Benefits of Pseudonymisation and Anonymisation under the GDPR' (2018) *Journal of Data Protection & Privacy* 146 <https://hstalks.com/article/246/comparing-the-benefits-of-pseudonymisation-and-ano/> accessed 2 July 2024.

[34] Article 29 Working Party on Data Protection, 'Opinion 05/2014 on Anonymisation Techniques' 1–37 <ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf> accessed 22 June 2024.

[35] Article 29 Working Party, 'Opinion 4/2007 on the Concept of Personal Data' <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf> accessed 22 June 2024; European Data Protection Board (n 25) 13.

first turn to determining when data can, in fact, be considered anonymous. The anonymisation threshold, which defines the point at which the processing of data falls outside the scope of the GDPR according to Recital 26, must be clearly defined. This is the subject of the identifiability test discussed in the following sub-section.

3.2.1 The Identifiability Test

To determine whether data are anonymous, one must consider when an individual is considered identified or identifiable. Recital 26 of the GDPR stresses that when determining if an individual is identifiable, 'all means reasonably likely to be used, such as singling out, … to identify the natural person directly or indirectly' should be considered. The Article 29 Working Party on Data Protection[36] previously suggested the following test for determining the identifiability of an individual: 'In general terms, a natural person can be considered as "identified" when, within a group of persons, he or she is "distinguished" from all other members of the group. Accordingly, the natural person is "identifiable" when, although the person has not been identified yet, it is *possible* to do it…'.[37]

According to this test, a person does not need to be named to be identified or identifiable. An individual is considered identifiable if, based on the specific circumstances and the state of technology,[38] the provided information allows them to be connected to the data in a way that does not apply to anyone else in the group. This is referred to as 'singling out', which must be prevented for anonymisation to be effective. Consequently, it is crucial to consider which identifiers, i.e. any pieces of information closely connected to an individual that can be used to single them out, either directly or indirectly, [39] are contained in the data.[40]

Hence, the mere removal of direct identifiers, such as a name,[41] does not render the data anonymous as indirect identifiers[42] (for example, a series of location data,

---

[36] The Article 29 Working Party on Data Protection has been replaced by the European Data Protection Board.
[37] Article 29 Working Party (n 35) 12.
[38] Data Protection Commission, 'Guidance Note: Guidance on Anonymisation and Pseudonymisation' 5 <https://www.dataprotection.ie/sites/default/files/uploads/2022-04/Anonymisation%20and%20Pseudonymisation%20-%20latest%20April%202022.pdf> accessed 22 June 2024.
[39] Data Protection Commission, 'Anonymisation and Pseudonymisation' <https://www.dataprotection.ie/dpc-guidance/anonymisation-pseudonymisation> accessed 20 June 2024.
[40] ibid.
[41] 'A direct identifier is specific information that references to an individual, such as name or an identification number.' European Data Protection Supervisor and Agencia Española de Protección de Datos, '10 Misunderstandings Related To Anonymisation'.
[42] 'An indirect identifier (also called quasi-identifier) is any piece of information (e.g. a geographical position in a certain moment or an opinion about a certain topic) that could be used, either individually or in combination with other quasi-identifiers, by someone that has knowledge about that individual with the purpose of re-identifying an individual in the dataset.' ibid.

history, or a compilation of personal life details) might still lead to the identification or distinction of individuals. Furthermore, even data stripped of all identifiers can still be linked to an individual if it is combined with other information related to that individual, for example, by linking different data sets.[43] Thus, if a compilation of data enables the identification of an individual, the data are not considered anonymised and, hence, their processing is governed by the GDPR.

3.3 Data Pseudonymisation

We now turn to pseudonymisation, the legal implications of which are very different to those of anonymisation as pseudonymisation does not exclude the application of the GDPR. As the data collected for the CREA2 project could not always be anonymised – a challenge discussed in more detail in Section 4 – pseudonymisation was used to ensure compliance with the GDPR.

Under Article 4(5) GDPR pseudonymisation is '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, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.' Pseudonymisation is a de-identification technique designed to protect personal data by replacing it with random values or by using cryptographic techniques. Typically, the pseudonym is generated without a direct connection to the individual, although it does not usually guarantee anonymity.[44] Pseudonymisation ensures both referential integrity and statistical accuracy.

Pseudonymisation removes direct identifiers (for example, name, phone number, identification number, etc.) from the data set, but may leave indirect identifiers (for example, a geographical position at a certain moment) in place.[45] In addition to replacing direct identifiers, technical and organisational measures must be implemented to comply with the GDPR definition of pseudonymisation. While pseudonymised data remains personal data, pseudonymisation helps fulfil GDPR requirements and is explicitly mentioned as an example of technical and organisational measures ensuring data protection (Article 25 GDPR).[46]

Pseudonymisation is a broad concept which can be further divided into two categories:

(1) Basic pseudonymisation which transforms the direct identifiers and puts appropriate controls in place to ensure the integrity of cryptographic keys.

[43] Data Protection Commission (n 38) 4, 6.
[44] Data Protection Commission (n 38); 'What Is Pseudonymization' (Imperva) <https://www.imperva.com/learn/data-security/pseudonymization/> accessed 20 June 2024.
[45] Hintze and El Emam (n 33) 146.
[46] ibid 147.

(2) Strong pseudonymisation, which also influences indirect identifiers and destroys any cryptographic keys, making pseudonymisation irreversible.[47]

Recital 28 of the GDPR explicitly states that pseudonymisation 'can reduce the risks to the data subjects concerned and help controllers and processors to meet their data-protection obligations'. Nevertheless, explicit referral to pseudonymisation in the GDPR does not 'preclude any other measures of data protection'. This disclaimer is in place because the GDPR frequently refers to pseudonymisation as an example of an appropriate measure or safeguard for ensuring data protection principles (see Recital 156 and Articles 6, 25, 32 and 89 GDPR). Article 6(4) GDPR even suggests that pseudonymisation may serve as an appropriate safeguard for processing personal data for purposes other than those for which they were originally collected, depending on the circumstances.

Therefore, Recital 28 of the GDPR is crucial for understanding that the mere use of pseudonymisation will not automatically lead to compliance with the GDPR. All circumstances of the case, like the purposes of processing and the collection of measures and safeguards in place, must be considered to determine compliance. Nevertheless, pseudonymisation is one of the key measures that can contribute to the lawfulness of personal data processing under the GDPR. Furthermore, the Court of Justice of the European Union has held that pseudonymised data is not considered personal data under the GDPR if the recipient of the data cannot reidentify the individuals.[48] This article focuses on pseudonymisation specifically, because it was a key tool in preparation of the data for the CREA2 project when data could not be anonymised sufficiently.

3.4 Anonymisation and Pseudonymisation of Judicial Decisions to Ensure Data Protection

As stated, if documents are sufficiently anonymised, based on the identifiability test (discussed in Sub-section 3.2.1), they do not fall under the scope of the GDPR.[49] Thus, publishing anonymised judicial decisions does not raise any data protection concerns under the GDPR and can safely be used in the development of AI-driven tools.

While anonymisation is the ideal solution as it removes all notions of 'personal' from personal data, achieving effective anonymisation can be demanding, particularly in fully automated processes. To ensure efficient anonymisation and robust data protection one should not rely only on AI-driven tools to perform anonymisation, but include a review by a human expert.[50] A human review is necessary, as AI-driven tools

[47] ibid.
[48] T-557/20 *SRB v EDPS* ECLI:EU:T:2023:219.
[49] Recital 26 GDPR.
[50] European Data Protection Supervisor and Agencia Española de Protección de Datos (n 41).

are still susceptible to errors[51] and human oversight enhances the legitimacy and trustworthiness of the process.

When anonymisation of judicial decisions is not feasible or is too costly, pseudonymisation may act as an appropriate measure or necessary safeguard for protecting the rights of data subjects and implementing data protection principles effectively.[52] Thus, pseudonymisation serves as a useful mechanism in striking a balance between safeguarding data protection rights and ensuring public access to judicial decisions and transparency. Nevertheless, it is important to note that unlike anonymised documents pseudonymised documents are governed by the GDPR.

When anonymising or pseudonymising court decisions, several key factors must be considered. First, determining the personal data sets to anonymise or pseudonymise, including identifiers such as first names, last names, addresses and ID numbers, is paramount. Second, determining which non-personal data sets, such as legal entity information, business secrets and state secrets, should also be subject to anonymisation or pseudonymisation where applicable. Third, consideration must be given to whether special categories of data warrant anonymisation or pseudonymisation before publication. Lastly, the identification of whose personal data should undergo anonymisation or pseudonymisation, ranging from defendants and witnesses to judges, lawyers, expert witnesses and third parties, necessitates careful consideration.[53] This comprehensive approach ensures that data protection principles are upheld while facilitating access to judicial information.

4. Data Protection and AI

4.1 Background

Building on the regulatory tools discussed above, this section explores the challenges of implementing these principles during the development of AI models in judicial contexts.

[51] Although humans are also susceptible to errors, the likelihood of mistakes is reduced by having both types of verification to ensure that personal data has been properly anonymised.
[52] Recitals 28 and 156, as well as Articles 6, 25, 32 and 89 GDPR.
[53] Council of Europe, 'Publication of Judicial Decisions the Council of Europe's Points for Consideration' 50 <https://rm.coe.int/publication-of-judicial-decisions-the-council-of-europe-s-points-for-c/1680aeb36d> accessed 22 June 2024.

The rise of AI has introduced a new era of challenges,[54] including privacy concerns.[55] The EU has partially responded to these challenges by introducing the AI Act[56] as the first global comprehensive legal framework governing AI. While a detailed presentation of the EU AI Act and its relationship to the GDPR falls beyond the scope of this article, there is a significant overlap between many data protection principles included in the GDPR and the requirements set by the EU AI Act for the safe development and use of AI systems.[57] The EU AI Act even specifies that it does not override the GDPR.[58]

Nevertheless, some key GDPR principles – such as purpose limitation, data minimisation, the special treatment of 'sensitive data', and restrictions on automated decisions – are challenged by the new methods of processing personal data enabled by AI. This tension can be mitigated by interpreting and developing the data protection principles in a way that facilitates the beneficial uses of AI. For example, the principle of data minimisation can in certain contexts be understood as reducing the 'personality' of the data available and not necessarily as reducing the quantity of data. This could involve measures such as pseudonymisation, which limits the ease with which data can be linked to individuals.[59]

Importantly, in December 2024 the European Data Protection Board (EDPB) issued an opinion on the use of personal data for the development and deployment of AI

---

[54] On the flipside AI can also be leveraged to automate data privacy and security processes. Siva Karthik Devineni, 'AI in Data Privacy and Security.' (2024) 3 *International Journal of Artificial Intelligence and Machine Learning* 35.

[55] Technical solutions addressing privacy and data protection challenges arising from Big Data and AI advancements have also been discussed. These solutions include establishing regulatory sandboxes; sustaining research, innovation, and implementation of privacy-preserving or privacy-enhancing technologies; and promoting and contributing to the development of technical standards for privacy protection. See, for example, Tjerk Timan and Zoltan Mann, 'Data Protection in the Era of Artificial Intelligence: Trends, Existing Solutions and Recommendations for Privacy-Preserving Technologies' in Edward Curry and others (eds), *The Elements of Big Data Value: Foundations of the Research and Innovation Ecosystem* (Springer International Publishing 2021) <https://doi.org/10.1007/978-3-030-68176-0_7> accessed 29 January 2025; Soumia Zohra El Mestari, Gabriele Lenzini and Huseyin Demirci, 'Preserving Data Privacy in Machine Learning Systems' (2024) 137 *Computers & Security* 103605.

[56] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act).

[57] James Clark, Muhammed Demircan and Kalyna Kettas, 'Europe: The EU AI Act's Relationship with Data Protection Law: Key Takeaways' (Privacy Matters, 25 April 2024) <https://privacymatters.dlapiper.com/2024/04/europe-the-eu-ai-acts-relationship-with-data-protection-law-key-takeaways/> accessed 29 January 2025.

[58] Article 2(7) EU AI Act.

[59] European Parliament. Directorate General for Parliamentary Research Services, *The Impact of the General Data Protection Regulation on Artificial Intelligence* (Publications Office 2020) II <https://data.europa.eu/doi/10.2861/293> accessed 29 January 2025.

models.[60] The opinion addresses the circumstances under which AI models can be considered anonymous, stating that the anonymity of an AI model should be evaluated on a case-by-case basis. For a model to be considered anonymous, it must be highly unlikely (1) to directly or indirectly identify individuals whose data was used in its creation, and (2) to allow personal data to be extracted through queries.

While we've stressed the importance of access to data in the interests of transparency, the question of access to data in the interests of developing AI-driven tools remains. However, this is not the focus of our article, since Section 2 merely argued for easier access to *public* data. First-instance judicial decisions are public information in Slovenia; they are just not readily available and need to be requested through a specific procedure.

Additionally, an important question arises in this context: can individuals reasonably anticipate their personal data being used for the development of AI-driven tools? The EDPB, in its opinion,[61] highlighted that one criterion for assessing reasonable expectations is whether the data was publicly available. Therefore, we see no issue with using public judicial decisions for AI-driven tools, provided their publication complies with GDPR standards – an issue we address in the following section.

4.2 The Identifiability Threshold and its Implications on CREA2

The CREA2 project was aimed at developing a game-theoretical algorithm to assist natural and legal persons in resolving disputes regarding the division of assets, particularly in inheritance[62] and divorce cases. Thus, in its development phase, the AI-driven tool had to be provided with anonymised judicial decisions[63] relating to the division of assets. It is important to note that people often refer to documents as 'anonymised' even though they do not pass the identifiability test (discussed in Sub-section 3.2.1.) and are thus not considered anonymous under the GDPR. Therefore, the anonymisation of the inputs needed for the CREA2 project presented significant challenges, which are discussed in the following section.

The controller (the lead of the research team who determined the purposes and means of processing personal data) as well as the processors (all members of the

[60] 'Opinion 28/2024 on Certain Data Protection Aspects Related to the Processing of Personal Data in the Context of AI Models | European Data Protection Board' <https://www.edpb.europa.eu/our-work-tools/our-documents/opinion-board-art-64/opinion-282024-certain-data-protection-aspects_en> accessed 29 January 2025. This opinion addresses three key issues: 1) the circumstances under which AI models can be considered anonymous, 2) the potential use of legitimate interest as a legal basis for developing or deploying AI models, and 3) the implications of developing an AI model using personal data processed unlawfully.

[61] ibid.

[62] According to Recital 27 GDPR, the GDPR does not apply to the deceased. However, it can still apply to the deceased's heirs.

[63] While this article focuses on judicial decisions in inheritance and divorce cases, it is important to note that Article 10 GDPR sets down specific conditions for processing of personal data relating to criminal convictions and offences.

research team who processed personal data on behalf of the controller) had to adhere to the GDPR when processing personal data.[64] While the judicial decisions that were provided in the data gathering phase were stripped of all direct identifiers (for example, name of the deceased and their heirs, personal identification numbers, bank account numbers, etc.), some data had to be included as it was a necessary input for the AI-driven tool. This follows the principle of data minimisation (Article 25(1)–(2) GDPR), according to which controllers should not collect more data than necessary to achieve the intended processing purpose. As already mentioned, in the context of AI-driven tools, the principle of data minimisation should be understood as reducing the 'personality' of the data available (for example, through pseudonymisation).[65]

For example, the provided judicial decisions in inheritance cases had to include the number of heirs, descriptions of assets (e.g., number of properties, the amount of money in bank accounts, number and type of cars, jewellery, etc.) and their division. Without these data, the AI-driven tool would not be able to make any valuable predictions regarding the division of assets in future, unsolved cases. Consequently, it is important to understand that anonymisation is not always possible whilst retaining a useful dataset for specific processing.[66]

Thus, despite removing direct identifiers from the judicial decisions, the indirect identifiers, the combination of non-personal data or the specific circumstances provided in court decisions, and especially the compilation of all of the included data, could narrow down or single out the data subject, potentially compromising anonymity. This is especially true when dealing with big data analytics and it underscores the limitations of many existing pseudonymisation and anonymisation techniques.[67] Furthermore, the more specific the data (for example, assets of the deceased, like intellectual property rights), the bigger the possibility of identification.[68] This problem is aggravated when one considers the size of a country like Slovenia, as its population of merely two million people poses a unique challenge (there are a limited number of people that hold such specific rights).

4.2.1. Case Study: Data Collection Challenges in Slovenia

Since first-instance court decisions are not available online in Slovenia, we needed to request access to the relevant pool of decisions from specific courts. As mentioned above, this process takes time and increases the workload of the court staff. Once the requests for accessing the data were approved, the researchers were provided with the anonymised and pseudonymised court decisions, which allowed them to feed the relevant data to the AI-driven tool's knowledge base.

[64] See also Article 28 GDPR regarding the obligations of the processor.
[65] European Parliament. Directorate General for Parliamentary Research Services (n 59) II.
[66] European Data Protection Supervisor and Agencia Española de Protección de Datos (n 41).
[67] Timan and Mann (n 55) 156.
[68] European Data Protection Supervisor and Agencia Española de Protección de Datos (n 41).

Discussions with the courts revealed several issues in the data anonymisation process, which required further effort and time from the court staff to address. Although the court staff aimed to fully anonymise the decisions by removing personal information such as names, addresses and children's names, as well as dates of birth and death, the remaining data—combined with the context in which the data were collected—still posed a risk of reidentification of the relevant individuals. Specifically, since the decisions were obtained from a district court handling inheritance cases for individuals who pass away in that region, it was not possible to hide the fact that the deceased resided within the district. Additionally, the timing of the decision could still be inferred from the case number or decision date.

In a small population such as Slovenia's, relatively few individuals pass away within the same period and district. While this alone did not pose a significant risk – given that multiple decisions from the same period were available, making identification based solely on timing unlikely – the potential for reidentification increased when additional details were included. Of particular concern were references to intellectual property rights or specific types of artwork. For instance, a description of assets such as 'copyright in four novels' or even the descriptions of physical assets such as '14 paintings and 80 statues' could plausibly point to a single individual in the region – an artist with a distinctive body of work who passed away during that particular period (e.g., early in the year).

Additional data also posed significant challenges for anonymisation, for example, data on the number, brand and model of vehicles owned; descriptions of rare and specific types of weaponry; and detailed information on real-estate ownership, such as ownership of houses in certain areas of Slovenia and Croatia, with references to small cadastral municipalities. As previously noted, the decisions also specified the number of heirs and their respective shares, and in cases where an heir had predeceased the decedent, this was explicitly mentioned – thereby potentially revealing aspects of the family structure.

Despite the removal of explicit personal identifiers such as names, addresses and dates, the combination of such granular details can enable reidentification of the deceased. Consequently, making the decisions available to the public also risks exposing the identities of living individuals, particularly heirs and descendants, thus failing to adequately safeguard their personal data.

Since the initial goal of input anonymisation under the CREA2 project could not always be achieved, pseudonymisation (which enhances data protection) became increasingly important. Since pseudonymised data falls under the GDPR, a lawful basis for processing is needed under Article 6 GDPR (or Article 9 GDPR for processing and managing special categories of personal data). Article 6(4) GDPR sets out conditions for secondary data processing (for example, for research and analysis purposes) that 'is compatible with the purpose for which the personal data are initially collected'. One of the key criteria for discerning the lawfulness of processing under Article 6(4) of the GDPR is 'the existence of appropriate safeguards, which may include encryption or pseudonymisation'. Thus, while pseudonymisation does not in itself

guarantee compliance with the GDPR, it increases the possibility that data processing for secondary purposes complies with the GDPR.[69]

Thus, when preparing the data for the AI-driven tool personal data were replaced with random values, ensuring both referential integrity and statistical accuracy. Slovenia's small population required a conservative approach, particularly for highly specific assets. In these instances, the data provided by the AI-driven tool had to be modified to an even higher degree to prevent the identification of individuals involved. Furthermore, court decisions deemed to carry a residual risk of reidentification were excluded from the knowledge base altogether. These practical challenges underscore the difficulty of applying anonymisation standards in jurisdictions with small populations and highly specific judicial data.

5. The Path Forward

The issue of sufficient anonymisation will also arise in the future, when (hopefully) first-instance judicial decisions will begin to be published online in Slovenia. While such publication would significantly enhance transparency, access to legal information, and consistency in case law, it remains questionable whether all judicial decisions will be able to pass the identifiability test as the aggregation of all included data may single out the data subject. Nevertheless, even if sufficient anonymisation is not possible, other appropriate measures can be taken to ensure compliance with the GDPR, while still retaining the benefits of transparency, consistent case law, etc.

Consequently, while giving us access to the data for the CREA2 project, the courts had to ensure its compliance with the GDPR before we were allowed to feed the data to the AI-driven tool. In contrast, other project partners worked with judicial decisions that were already publicly available online, which were presumed to comply with data protection laws. In Slovenia, the courts realised that a large number of judicial decisions could not in fact be anonymised according to GDPR standards due to the indirect identifiers that were inevitably left in the decisions. This is especially problematic for countries with a small population, such as Slovenia.

Whether a decision is sufficiently de-identified to be considered anonymous, must, of course, be established on a case-by-case basis, as stressed also by the EDPB.[70] It is merely our purpose to flag the importance of understanding the exact definition of anonymisation, especially in light of specific circumstances in a small country. When processing data that has not in fact been anonymised, it is essential to uphold the rights guaranteed by the CFR and the European Charter of Human Rights, and to ensure compliance with the GDPR.

[69] Hintze and El Emam (n 33) 151.

[70] 'Opinion 28/2024 on Certain Data Protection Aspects Related to the Processing of Personal Data in the Context of AI Models | European Data Protection Board' (n 60).

Going forward, policymakers should support the development of technical standards and regulatory guidance that enable the responsible use of judicial data. This includes establishing a national anonymisation protocol for judicial decisions, mandating human oversight in any automated anonymisation processes, creating a publicly accessible, centralised database of judicial decisions, with layered access based on data sensitivity, and clarifying the role of pseudonymisation in lawful secondary processing, particularly in research and AI development contexts. These steps would help reconcile the principles of innovation and privacy protection. They would also provide legal certainty to courts, researchers and developers working at the intersection of law and AI.

6. Conclusion

This article examined the intersection of data protection and the development of AI-driven tools in the context of judicial proceedings within the EU, using the CREA2 project and the Slovenian legal environment as a case study. It highlighted the legal and practical challenges of processing personal data – particularly using anonymisation and pseudonymisation techniques – under the GDPR when using judicial decisions as inputs for AI tools.

Our findings underscore several core tensions. On the one hand, AI-driven tools for judicial proceedings require access to detailed, structured and legally relevant case data to generate accurate, useful outcomes. On the other hand, GDPR compliance, especially the requirement to avoid reidentification risks, places strict limits on the nature and scope of data that can be processed. These constraints were particularly acute in Slovenia due to the inaccessibility of first-instance court decisions and the heightened reidentification risks associated with a small population.

Consequently, this study demonstrates the importance of designing AI systems in accordance with privacy by design and privacy by default principles. Nevertheless, this article calls for improved access to judicial decisions across the EU. Future policy and legislative initiatives must consider the practical difficulties encountered during projects like CREA2, particularly in smaller jurisdictions, and support the development of robust anonymisation standards, legal clarity for secondary data use, and access to judicial decisions.