Lawful Grounds to Share Justice Data for Lawtech Innovation in the UK

Stergios Aidinlis,[*] Hannah Smith,[+] John Armour[§] and Jeremias Adams-Prassl[¶]

(2024) 140 Law Quarterly Review

forthcoming

Abstract

The digitalisation of the UK's justice system provides new opportunities to develop insights into its efficiency and fairness. Both the UK government and commercial entities in the Lawtech Innovation space are keen to capitalise on the opportunities presented by these developments. But does the processing of justice data by lawtech innovation actors comply with UK data protection law? We respond to this question by presenting three possible constellations of relying on "public interest" and "legitimate interests" as grounds for processing in this context. We argue that lawtech analytics can potentially contribute to the "public interest" by enhancing access to justice through making bespoke legal insights much more accessible compared to the cost of specialised legal advice. Nonetheless, in the light of cautious regulatory interpretations of the first data protection principle, we proceed with a further exploration of "legitimate interests" as an alternative for public bodies and commercial entities, provided that certain requirements are satisfied. Our analysis raises broader questions about the wider regulation of the legal services sector, considering the increasing influence and participation of commercial entities.

---

[*] Assistant Professor in Artificial Intelligence Law, Durham University, corresponding author. We would like to acknowledge generous support from the UKRI (AI FOR ENGLISH LAW UKRI GRANT No ES/S010424/1 – PI: John Armour) and the ERC (iMANAGE - Grant agreement No. 947806 – PI: Jeremias Adams-Prassl) in conducting the research that led to this publication. We would also like to thank Dr Natalie Byrom and Professor Joe Tomlinson for acting as discussants in a workshop where an earlier version of this work was presented, as well as an anonymous reviewer and the editor of the L.Q.R. for their insightful suggestions and help in preparing this work for publication. All errors and omissions remain our own.

[+] Research Fellow, The University of Western Australia.

[§] Dean of the Faculty of Law and Professor of Law and Finance, University of Oxford.

[¶] Professor of Law and Associate Dean (Research), Fellow of Magdalen College, University of Oxford.

I. Introduction: digital justice, access to justice, and legal analytics

In 2016, HM Courts and Tribunals Service (HMCTS) outlined its initial plans to reform the justice system to improve its accessibility and efficiency, committing nearly £1bn to the reforms.[1] The HMCTS reform programme sought to introduce "digital technology and modern ways of working to support public and professional users…improving access to justice and efficiency of the system for all".[2] In the intervening years, progress has been made in opening up online services to more people and introducing new online services in the areas of public family law, as well as the immigration and asylum tribunals. The impact of Covid-19 further catalysed the digitalisation of the justice system: 426,000 individuals have now used the online services provided by HMCTS.[3]

The intent to "free the courts from the constraints of storing, transmitting, and communicating information on paper"[4] serves not only to make the legal processes more efficient and streamlined. Alongside these benefits is the potential to enhance key constitutional principles promulgated and protected by the justice system. Of particular interest to this article is the impact of this reform programme on citizens' right of access to justice, notably elucidated in *UNISON*[5] as "inherent in the rule of law". Without access to the courts, "laws are liable to become a dead letter, the work done by Parliament may be rendered nugatory, and the democratic election of Members of Parliament may become a meaningless charade".[6] The benefits from this

[1] The Lord Chancellor, the Lord Chief Justice and the Senior President of Tribunals, ''Transforming Our Justice System'' (2016) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/553261/joint-vision-statement.pdf>.

[2] HM Courts & Tribunals Service, ''Reform Update: Summer 2019'' (2019) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/806959/HMCTS_Reform_Update_Summer_19.pdf> 3.

[3] HM Courts & Tribunals Service, ''The HMCTS reform programme'' (2021) <https://www.gov.uk/guidance/the-hmcts-reform-programme>.

[4] J. Rozenberg, ''HMCTS Reform - the Online Court: Will IT Work?'' (2018) <https://long-reads.thelegaleducationfoundation.org/hmcts-reform/>.

[5] *R (UNISON) v Lord Chancellor [2017] UKSC 51; [2020] A.C. 869* at [66] per Lord Reed.

[6] *R (UNISON) v Lord Chancellor [2017] UKSC 51; [2020] A.C. 869* at [68] per Lord Reed.

right accrue not just to individuals but to society more broadly as the knowledge that rights are enforceable "underpins everyday economic and social relations".[7]

Multiple start-ups have been established in the past few years whose commercial interests and businesses aims intersect with the right of access to justice. There is significant promise in the legal analytics offered by these start-ups. Their utilisation of advances in data analytics promises the ability to provide citizens with a plethora of information that enhances their rights in a cost-effective way. The insights gained from the data now collected as part of HMCTS' reform programme may enable such start-ups to provide citizens with data and effective advice on their legal rights, probabilities of success, and potential sources of funding. To best fulfil the potential of their products, these lawtech start-ups require access to the relevant justice data now generated as part of the reform programme.

This discussion has become increasingly topical, not least given the rise of generative AI systems, including large language models such as ChatGPT. On the one hand, the use of generative AI for legal purposes may be quite problematic. Writing about the limits of computational law, Mireille Hildebrandt cautions against a law that does not treat people as human agents but as "subject to a statistical, machinic logic".[8] While we share this normative concern, there is no doubt that the use of AI for legal purposes is already here: in 2023, a litigant in person provided false citations before a civil court in Manchester after consulting ChatGPT.[9] With UK judges starting to admit to the use of ChatGPT,[10] the risk of inaccurate suggestions by the chatbot influencing critical legal decisions becomes apparent. Furthermore, considering that AI is trained on historic data,[11] it might be unable to capture developments in the common law, potentially resulting in stagnant jurisprudence. On

---

[7] *R (UNISON) v Lord Chancellor [2017] UKSC 51; [2020] A.C. 869* at [71] per Lord Reed.
[8] M. Hildebrandt, ''Boundary Work between Computational 'Law' and 'Law-as-We-Know-it'' in *D. Curtin and M. Catanzariti (eds.), Data at the Boundaries of European Law (Oxford: Oxford University Press, 2023), at p. 30*.
[9] J. Hyde, ''LiP presents false citations to court after asking ChatGPT'' <https://www.lawgazette.co.uk/news/lip-presents-false-citations-to-court-after-asking-chatgpt/5116143.article>.
[10] G. Corfield, ''British judge uses 'jolly useful' ChatGPT to write ruling'' (The Telegraph) <https://www.telegraph.co.uk/business/2023/09/14/british-judge-uses-jolly-useful-chatgpt-to-write-ruling/>.
[11] P. Hacker, ''A legal framework for AI training data—from first principles to the Artificial Intelligence Act'' (2021) 13(2) L.I.T. 257 at 270.

the other hand, assuming that generative AI systems are adapted to the legal system and fine-tuned with the use of legal data,[12] such risks may be mitigated and there might be significant promise of offering cost-effective legal advice to citizens, especially considering the rising number of litigants in person after severe cuts in UK legal aid.[13]

However, the effective deployment of data analytics in the legal sector requires data as an input. "Justice data"—the data associated with legal processes and outcomes—contains many pieces of personal data about litigants and others. While its publication as part of the justice system is expressly permitted under data protection law, matters are less clear-cut as regards secondary processing—that is, further processing for purposes beyond those expressly covered by the justice exemption. The purpose of this article is to explore the current regulatory landscape and the potential avenues for rendering what we refer to as "secondary uses of justice data" for lawtech innovation purposes lawful. While we offer a comprehensive definition of "justice data" in section III of the article, one can envisage that beyond judgments, many of which are easily accessible,[14] such case documents as pleadings, briefs, exhibits and deposition transcripts might be useful to extract key information about cases. This structured data might then be used by a legal-tech start-up to build products like a database of criminal charges and typical sentencing ranges, or analytics on judge's tendencies and lawyers' track records.[15] Determining lawfulness, however, is a prerequisite for realising the potential social and economic benefits from lawtech innovation, since the absence of a robust legal basis for data processing may render providers of legal

[12] Fine-tuning allows us to refine pre-trained models such as GPT-3.5 or GPT-4 for ''precise applications by subjecting them to a more targeted dataset that closely aligns with the specific task at hand'', see P. Junco, ''The Power Of Fine-Tuning In Generative AI'' <https://www.forbes.com/sites/forbestechcouncil/2023/10/10/the-power-of-fine-tuning-in-generative-ai/>.

[13] The Law Society, ''A decade of cuts: Legal aid in tatters'' <https://www.lawsociety.org.uk/contact-or-visit-us/press-office/press-releases/a-decade-of-cuts-legal-aid-in-tatters>.

[14] Although it shall be noted that there is no comprehensive database of English case law, recent government efforts resulted in the ''Find Case Law'' archive which seeks to gather most published judgments, especially from higher courts. For concerns about the increasing availability of such data, see Z Adams, A Adams-Prassl, and J Adams-Prassl, 'Online Tribunal Judgments and the Limits of Open Justice' (2022) 42 Legal Studies

[15] D.L. Chen, ''Judicial analytics and the great transformation of American Law'' (2019) 27(1) A.I.L. 15 at 30.

technologies liable to sanctions for unlawful data processing,[16] thus jeopardising their business models.

Our key argument is that there is scope under UK data protection law, and under the EU General Data Protection Regulation (hereafter GDPR), lawfully to provide access to justice data for lawtech entities. We argue that the better alternative for public bodies is to rely on the "public interest" ground, considering the potential of lawtech innovation to enhance access to justice by making bespoke legal insights much more accessible compared to the cost of specialised legal advice. Yet, acknowledging the scepticism in current regulatory guidance against the invocation of the "public interest" for data processing involving commercial entities as a key actor, we argue that public bodies should also consider relying on "legitimate interests" as a legal basis. With both parts of the argument, we provide an extensive consideration of the relevant legal provisions, the context-specific considerations in the particular case of data sharing, and the applicable regulatory guidance in the UK and the EU.

Discussion is structured as follows. The article commences by introducing key considerations relating to data protection law, as well as the terminology central to the challenge under scrutiny. It then provides a background of the relevant context, defining "justice data", introducing key actors, and outlining the relevant law governing the reuse of justice data. The next section provides a closer analysis of the potential avenues for lawful data sharing in this context, presenting a "constellation" of different grounds for legitimising the reuse of justice data by lawtech start-ups and analysing their implications. The article highlights the potential for the "public interest" and the "legitimate interests" grounds to justify the processing of justice data for lawtech innovation purposes, highlighting their respective opportunities and constraints. In doing so, the article draws upon existing guidance issued by the I.C.O.[17] on the use of the legitimate interest grounds to demonstrate its continued applicability to novel data sharing processing. The intent is to demonstrate that such principles can continue to guide data processing in a rapidly developing context that both capitalises upon

[16] B. Fiten and G. Somers, ''AI-Based Legal Tech Solutions: Discover The Legal Pitfalls'' (timelex) <https://www.timelex.eu/en/blog/ai-based-legal-tech-solutions-discover-legal-pitfalls>.
[17] I.C.O., ''Legitimate Interests'' <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/legitimate-interests/>.

new opportunities whilst promoting a healthy respect for individuals' rights and interests. A concluding section reflects upon these arguments to consider the broader implications for the regulation of legal services.

II. Data protection law: logic and limits

Data protection law covers the processing of *personal* information by both legal and natural subjects.[18] The I.C.O. broadly define "personal" data as information "about a particular living individual";[19] such information may lie in the public realm or could be privately held. A more specific legal definition is provided under article 4 of the EU GDPR, confining personal information to information *relating to*[20] natural persons who "can be identified or who are identifiable, directly from the information in question; or who can be indirectly identified from that information in combination with other information".[21] Identifiability, as prescribed in the law, is not static. Recital 26 of the GDPR stipulates that the question of whether such information is identifiable, and thus personal, hinges on the specific context of processing and "all the means reasonably likely to be used" by the data controller or a third person shall be assessed before reaching a conclusion (Recital 26, EU GDPR). As some of the present authors have argued elsewhere,[22] the "data environment", i.e., a set of factors around data processing and the circumstances under which it takes place, determine whether the same information will be personal or non-personal or anonymous. When data are anonymous, data protection law does not apply, as its primary aim is to "enhance the exercise of individual control over personal data" while at the same time facilitating necessary data flows for the operation of the economy.[23]

---

[18] *L. Bygrave, Data protection law (Amsterdam: Wolters Kluwer, 2002) at p. 10.*
[19] L. Dalla Corte, ''Scoping personal data: towards a nuanced interpretation of the material scope of EU data protection law'' (2019) 10(1) E.J.L.T..
[20] Data "relating to" natural persons is a broader category than data "about" natural persons, since the former can also include information whose "purpose" or "effect" is associated with the natural person, than merely its "content". See the CJEU judgment in *Nowak v Data Protection Commissioner* (C-434/16) EU:C: 2017:994; [2018] 1 W.L.R. 3505 at [35].
[21] I.C.O., "What is personal data?" <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/key-definitions/what-is-personal-data/>.
[22] M. Mourby et al, ''Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK'' (2018) 34(2) C.L.S.R. 222 at 224.
[23] *O. Lynskey, Foundations of EU Data Protection Law (Oxford: Oxford University Press, 2015) at p. 75.*

Controllership is a central concept in data protection law, as it delineates the extent to which certain actors are responsible to uphold specific obligations under the law. It gives rise to the central distinction between *data controllers,* i.e., the ones who "exercise overall control over the purposes and means of the processing of personal data",[24] and *data processors,* who "act on behalf of, and only on the instructions of, the relevant controller".[25] There are scenarios where it is difficult to distinguish controllership from processing, e.g., in the context of smart homes,[26] but in most cases the distinction will be straightforward and will have significant repercussions for data protection law. A data controller will bear the highest level of compliance responsibility with the law, whereas data processors will have fewer obligations, although still direct ones under the law that are likely to result in liability.[27]

In EU member states, data protection law, and more specifically for present purposes the EU GDPR,[28] applies directly and horizontally. In the UK, the GDPR remains part of UK law as the UK GDPR,[29] in conjunction with an amended version of the Data Protection Act 2018. The relationship between the two legal instruments is in principle significant since it may determine such issues as the breadth of derogations from the data protection requirements.[30] For the purposes of the present article, we focus on the GDPR, as applicable in the UK, due to our interest being more in cornerstone principles of the legislation and their reinterpretation in the new light of emerging technological and social realities for the legal profession. The enforcement of data protection rights in the UK might be an open question after the removal of

---

[24] I.C.O., "Controllers and processors" <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/key-definitions/controllers-and-processors/>.

[25] I.C.O., "Controllers and processors" <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/key-definitions/controllers-and-processors/>.

[26] J. Chen, L. Edwards, L. Urquhart, D. McAuley, "Who is responsible for data processing in smart homes? Reconsidering joint controllership and the household exemption" (2020) 10(4) I.D.P.L. 279 at 280.

[27] I.C.O., ''Legitimate Interests'' <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/legitimate-interests/>.

[28] The EU GDPR is the main focus of the present article because it is the applicable legislation to the conduct of justice data sharing for lawtech purposes, as opposed e.g., to the Law Enforcement Directive which applies for such purposes as 'prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties' by competent law enforcement authorities.

[29] I.C.O., ''The UK GDPR'' <https://ico.org.uk/for-organisations/dp-at-the-end-of-the-transition-period/data-protection-and-the-eu-in-detail/the-uk-gdpr/>.

[30] M. Mourby et al, "Governance of academic research data under the GDPR—lessons from the UK" (2019) 9(3) I.D.P.L. 192 at 198.

article 8 of the EU Charter of Fundamental Rights[31] and the government's aspirations to overhaul UK data protection law.[32] Nonetheless, the UK has received "adequacy" decisions both under the GDPR and the EU Law enforcement Directive and shall be expected to adhere to their requirements for the foreseeable future. [33] Hence, this analysis will remain relevant even after the proposed reform of the UK data protection law by the current government.[34] We now turn to the specific empirical context of the article: sharing justice data for lawtech innovation purposes.

III. Sharing justice data for lawtech innovation: scope and background

Justice data holds a prominent role among the different types of public-sector data generated throughout the operation of large-scale bureaucracies for such purposes as social security and tax collection.[35] The UK government has been very willing to utilise such data to achieve "public service modernisation" and improve civil service delivery through insights produced by data-intensive research.[36] Indeed, the justice system has been at the forefront of realising this aspiration, as noted in the introduction, and making justice data available for lawtech entrepreneurs remains one of the main aims and ambitions of government reformers. Yet, as the Law Society highlighted in its "Report on Algorithms in the Criminal Justice System",[37] existing use of data analytics and algorithmic technologies in the justice system without sufficient interrogation and explainability to the public might "threaten human rights and

[31] S. Aidinlis, "The Right to be Forgotten as a Fundamental Right in the UK after Brexit" (2020) 25(2) C.L. 67 at 70.
[32] D. Erdos, "The Draft Data Protection (Fundamental Rights and Freedoms) Amendment Regulations: Arguably Partially Ultra Vires and Liable to Undercut the UK's International Commitments" (2023) University of Cambridge Faculty of Law Research Paper No. 26/2023.
[33] European Commission, "Decision on the adequate protection of personal data by the United Kingdom - General Data Protection Regulation" (28 June 2021); K. McCullagh, "Post-Brexit data protection in the UK-leaving the EU but not EU data protection law behind" in *G. González Fuster et al (eds.), Research Handbook on Privacy and Data Protection Law (Cheltenham: Edward Elgar, 2022) at p. 36.*
[34] Department for Culture, Media, and Sport, "Data: a new direction" <https://www.gov.uk/government/consultations/data-a-new-direction>.
[35] R. Connelly et al, "The role of administrative data in the big data revolution in social science research" (2016) 59 S.S.R. 1 at 3.
[36] J. Manzoni, "Big data in government: the challenges and opportunities" (21 February 2017) <https://www.gov.uk/government/speeches/big-data-in-government-the-challenges-and-opportunities>.
[37] The Law Society, "Algorithm use in the criminal justice system report" https://www.lawsociety.org.uk/topics/research/algorithm-use-in-the-criminal-justice-system-report.

undermine public trust in the justice system".[38] Data protection law requirements aspire to provide transparency, explainability and lawfulness checks on the use of justice data for lawtech innovation.

What do we mean, then, by referring to making "justice data" available for lawtech innovation purposes? In previous work,[39] some of the present authors defined "justice data" as the "wealth of data on legal processes and outcomes", particularly thinking of the novel data that will be generated in the advent of digitalising courts and tribunals in the UK.[40] Yet, since "justice data" is not a legal term, there is a need to provide here a clearer definition of the types of data that would be of interest for the purposes of the present article.

Across the justice system, there is a wealth of data that could potentially interest lawtech innovation entities, e.g. textual, audio or video data of judgments,[41] meta-data about judgments and cases (for example referring to the average length of judgments or duration of active cases), administrative data about the operation of the courts and tribunals held by the Ministry of Justice/HMCTS, or even data held by other private and public entities that would be relevant for the justice system (as in the case of equalities data).[42] Crucially, these different types of data will exist on different places along the identifiability spectrum, e.g. judgments texts may include some directly identifiable data like the names of judges and litigants, whereas records of how long particular types of cases last may not contain such information. Similarly, there might be differences in terms of the sensitivity or non-sensitivity of the data, all questions of

[38] The Law Society, "Algorithm use in the criminal justice system report" https://www.lawsociety.org.uk/topics/research/algorithm-use-in-the-criminal-justice-system-report at 19.
[39] S. Aidinlis, H. Smith, A. Adams-Prassl and J. Adams-Prassl, "Building a Justice Data Infrastructure: Opportunities and Constraints" (September 2020) AI4LAW UKRI-funded project, University of Oxford <https://www.law.ox.ac.uk/sites/files/oxlaw/ukri_justice_data_report_fv_0.pdf> at p.2.
[40] HM Courts & Tribunals Service, "The HMCTS reform programme" (updated 14 May 2021) <https://www.gov.uk/guidance/the-hmcts-reform-programme>.
[41] See the example of Just: Access, "Just: Transcription" <https://www.just-access.org/how-it-works> offering an easy and affordable solution that turns audio files, including witness testimonies, into written transcripts.
[42] For a more comprehensive categorisation of justice data see N. Byrom, "Digital Justice: HMCTS Data Strategy and Delivering Access to Justice" (The Legal Education Foundation 2019) <https://research.thelegaleducationfoundation.org/blog/digital-justice-hmctsdata-strategy-and-deliveringaccess-to-justice> at 23.

profound significance for the legal implications of the activity, particularly when novel statistical techniques and linkages of data are to be employed by private entities.

Considering the plethora of potential interests in justice data, we confine our analysis here to *textual data*, mostly thinking of judgments, but also potentially expanding to pleadings and submissions by litigants on the basis of which judges decide cases and deliver judgment.[43] Considering that a significant number of judgments is already in the public domain and can be mined to train AI models under the terms and conditions of their respective databases,[44] our analysis will refer to judgments that are not already public and will be shared for the first time by courts and other public bodies, as well as case documents such as pleadings and litigants' submissions that are similarly not in the public domain. Whether a judgment or related case document is already in the public domain has crucial legal implications for personal data sharing, as the status of data as already published weakens the reasonable expectations of privacy of data subjects that their data will not be processed by other entities.[45]

This confinement is justified by virtue both of the topicality of widening publication of and access to such data in the UK,[46] and the significant potential of legal analytics based on such data to enhance access to justice by yielding compelling insights about the potential of legal claims and lowering the costs of triaging parties' circumstances and needs.[47] Crucially, confining the scope to textual data of this kind excludes some of the most controversial and ethically dubious applications of analytics in the justice system, e.g. the use of background information or personal data of judges to predict individual judicial behaviour - a type of legal analytics firmly banned in France.[48] By contrast, the emphasis on textual data seeks to develop overall trends and patterns in

[43] P. Leith and C. Fellows, "Enabling Free On-line Access to UK Law Reports: The Copyright Problem" (2009) 18(1) I.J.L.I.T. 72 at 80.
[44] See for example, BAILII, "Reproduction and Copyright" <https://www.bailii.org/bailii/copyright.html>.
[45] J. Bell et al, "Balancing Data Subjects' Rights and Public Interest Research" (2019) 5 E.D.P.L. 43 at 50.
[46] Ministry of Justice and HM Courts & Tribunals Service, "Boost for open justice as court judgments get new home" (16 June 2021) <https://www.gov.uk/government/news/boost-for-open-justice-as-court-judgments-get-new-home>.
[47] YouGov, "Legal needs of Individuals in England and Wales" <https://legalservicesboard.org.uk/wp-content/uploads/2020/01/Legal-Needs-of-Individuals-Technical-Report-Final-January-2020.pdf>.
[48] M. Langford and M. R. Madsen, "France Criminalises Research on Judges" (Verfassungsblog, 22 June 2019) <https://verfassungsblog.de/france-criminalises-research-on-judges/>.

the case law, drawing inferences about the impact of existing legal principles on a prospective claim.[49] In other, legal-philosophical words, our focus in the present article is legal analytics for *positivists,* i.e. analytics that assume the central role of pre-existing posited norms from previous precedents in predicting future decisions, instead of legal analytics for *realists*, i.e. analytics that assume that extra-legal factors have a dominant role in shaping the outcome of judicial decisions.[50]

Having made this clarification, will it always be the case that justice data of interest to lawtech entities will include "personal data" in the legal sense as defined in article 4 of the EU GDPR?[51] Identifiability is a context-specific question: it cannot be assumed that justice data will always be personal,[52] thus triggering the application of data protection law. In some cases, lawtech innovators may confine themselves to e.g., anonymous and aggregate-level data that suffices to estimate the timeframe within which a particular type of dispute is likely to be resolved. In most cases, however, it is fair to anticipate that such personal information as the name of the judge (or judges) and the names of the lawyers representing the parties will be needed to produce estimations of the potential success of a claim. Data controllers are under an obligation to assess on an *ad hoc* basis whether data protection law and, consequently, the analysis offered in the present article apply. If they do not feel confident that the data of interest is not personal, the EU GDPR, as transposed in the UK post Brexit via the Data Protection Act (DPA) 2018, will apply.

Provided that justice data are personal data, another important clarification about the *applicable reach* of data protection laws relates to whether the discussed data sharing activity serves the purpose of "research" under the GDPR. In case research exemptions apply, such data protection provisions as data subject rights will have limited applicability in this context. Due to the exploratory character of lawtech

[49] See the example of Solomonic, "Litigation Intelligence" <https://www.solomonic.co.uk>.
[50] For a deeper understanding of the differences between positivists and realists, see the seminal paper by B. Leiter, "Legal Realism and Legal Positivism Reconsidered" (2001) 111(2) Ethics 278-301.
[51] Article 4 GDPR: "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".
[52] Mourby et al, ''Are 'pseudonymised' data always personal data? Implications of the GDPR for administrative data research in the UK'' (2018) 34(2) C.L.S.R. 222 at 224.

innovation and its foundations in computing and mathematical science, lawtech actors may claim that their services, and thus their data processing actions, fall within the ambit of scientific research,[53] at least until their product is finalised and validated for broader market use. Important legal implications stem from determining this issue, such as the application of the research derogation under article 89 GDPR which may pre-empt the exercise of certain data subject rights as long as organisational safeguards are applied by the data processors.[54] The boundaries between data-intensive research and innovative commercial legal services, however, may be less clear than what one might think in the EU GDPR. Arguably, the Regulation adopts an inclusive definition of "research", comprising not only of academic research but also of commercial research endeavours.[55] In their "Preliminary Opinion on data protection and scientific research",[56] the European Data Protection Supervisor (hereafter EDPS) clarified that scientific research might be conducted both by not-for-profit entities like academic institutions and by for-profit corporate actors. Yet, in the same opinion, the EDPS confined scientific research to research inquiries that apply the scientific method of observing phenomena, formulating, and testing a hypothesis for those phenomena, and concluding as to the validity of the hypothesis.[57] Where is the line to be drawn between scientific research and other types of lawtech innovation?

Lawtech innovation primarily aims at the application of scientifically advanced methods of data analytics to justice data, mainly judgments, with a view to offering a paid service of legal advice to a broad base of consumers,[58] at the fraction of the cost of specialist legal advice provided by human lawyers. In applying such methods, it may very well be the case that scientific statistical methods are employed, certain

[53] See for example collaborations in the lawtech sector such as the ones between the Legal Innovation Lab in Wales and the Solicitors Regulation Authority to develop an open source platform for the co-development of lawtech apps, https://legaltech.wales/en/the-lab.
[54] For a comprehensive analysis of the research exemptions in the GDPR, see Mourby et al, "Governance of academic research data under the GDPR—lessons from the UK" (2019) 9(3) I.D.P.L. 192.
[55] GDPR Recital 159: "(…) For the purposes of this Regulation, the processing of personal data for scientific research purposes should be interpreted in a broad manner including for example technological development and demonstration, fundamental research, applied research and privately funded research (…)". It shall be noted that GDPR recitals are not legally binding but provide insights into the background behind the provisions and influence the interpretation of binding provisions.
[56] European Data Protection Supervisor, "A preliminary opinion on data protection and scientific research" <https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf>.
[57] European Data Protection Supervisor, "A preliminary opinion on data protection and scientific research" <https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf at 10.
[58] R. Dale, "Law and Word Order: NLP in Legal Tech" (2019) 25(1) N.L.E. 211 at 217.

hypotheses are being tested and refined until the optimal algorithmic solution has been constructed.[59] The main purpose of the activity is, however, certainly different from independent, not-for-profit research, where the aim is the advancement of knowledge and the publication of one's findings in scientific venues. By contrast, commercial interests dictate that advances achieved in the course of developing legal analytics be kept confidential as proprietary interests of the lawtech company. This does not mean that lawtech innovation will not satisfy the GDPR definition of "research" but creates complications as to the availability of information to estimate whether the scientific method is indeed applied. As a result, the assessment will be a case-by-case one: it cannot be assumed from the outset that data sharing for lawtech innovation will benefit from the GDPR research exemptions.

Which are the actors that will need to make or contribute to this assessment? Which are the affected stakeholders whose interests shall be considered in doing so? From the side of entities interested in accessing justice data, the answer is relatively clear: commercial entities interested in accessing justice data, ranging from start-ups to boutique law firms and legal publishers active in the market of legal analytics. An example of a lawtech services provider interested in this kind of justice data would be the following:

> Sheba Ltd is a lawtech start-up which develops sophisticated machine learning techniques for analysing UK High Court judgments and extracting from them key insights for the likely success of future claims before the same court. Sheba claims that, by using their product, law firms will be able to augment, support and contextualise their personalised advice to clients with the support of carefully analysed data, aggregated from multiple sources and reviewed by both robust algorithmic technology and the human expertise of qualified lawyers.

[59] S. Kuleshov et al, "Legal Tech: Documents' Validation Method Based on the Associative-Ontological Approach" in *A. Karpov and R. Potapova (eds), Speech and Computer. SPECOM 2020. Lecture Notes in Computer Science (Berlin: Springerx, 2020)* (online first) https://doi.org/10.1007/978-3-030-60276-5_25.

A recent survey, commissioned by the Legal Services Board (LSB), about the needs of users of the justice system in England and Wales brought together organisations like the LSB, the Law Society, and legal researchers associated with academic institutions to design a questionnaire that was circulated to a representative sample of the general public.[60] With regard to the needs of the public, Byrom has called particular attention to "court users' vulnerabilities, including age, mental and physical disabilities, literacy levels, and gender" in designing data collection policies in the justice system.[61] Voluntary and community sector (VCS) stakeholders are similarly impacted by data sharing arrangements in the justice system, considering that they often rely on data held by the Ministry of Justice and HMCTS to evaluate their impact on triggering social change and scrutinising the government.[62]

The picture is less clear when it comes to entities legally responsible for managing the datasets and granting access to them. Broadly speaking, one would associate involvement in the activity with controllership of justice data. The Ministry of Justice, HMCTS, and the judiciary are arguably the central actors to be considered here. Nonetheless, the collection, processing and controllership of justice data are paradigmatically fragmented and problematic in the UK, cautioning every researcher against committing to a categorical identification of actors granting access to justice data comprised of both data controllers and data processors.[63] For example, a reasonable amount of British and Irish case law has been held and published, for quite some time, by the British and Irish Legal Information Institute (hereafter BAILII), a non-profit entity that is "hosted in the UK and Ireland by the Institute of Advanced Legal Studies, London and the Law Faculty, University College Cork".[64] Controllership of data may also be complicated by reference to the intersection between data protection law and other fields of law, e.g. IP law, when it comes to the copyright entitlements of

[60] YouGov, "Legal needs of Individuals in England and Wales" <https://legalservicesboard.org.uk/wp-content/uploads/2020/01/Legal-Needs-of-Individuals-Technical-Report-Final-January-2020.pdf> at 4.
[61] N. Byrom, "Digital Justice: HMCTS Data Strategy and Delivering Access to Justice" (The Legal Education Foundation 2019) <https://research.thelegaleducationfoundation.org/blog/digital-justice-hmctsdata-strategy-and-deliveringaccess-to-justice> at 4.13 and 4.33.
[62] See for example the case of VCS organisations which triggered the creation of the Justice Data Lab in the UK Ministry of Justice in F. Lyon et al, "Opening access to administrative data for evaluating public services: The case of the Justice Data Lab" (2015) 21(2) Evaluation at 232.
[63] M. Bryan, "Early English law reporting" (2009) 4 University of Melbourne Collections.
[64] BAILII, "About BAILII" <https://www.bailii.org/bailii/>.

judges over the judgment as an intellectual product.[65] For present purposes, we confine our interest to data held and managed by public bodies, such as courts, the Ministry of Justice or HM Courts & Tribunals, as we assume that this will refer to the majority of justice data.

In referencing other fields of law, it is also prudent to consider whether other existing exemptions may apply to this particular data processing practice. There are several exemptions within the Data Protection Act 2018 that are relevant to the processing of justice data. Nevertheless, a closer examination of the provisions suggests that they are not applicable to the secondary uses of justice data explored in this article. For example, Paragraph 26, Schedule 1 of the Data Protection Act 2018 permits the processing of special categories of data necessary to publish judgements or other decisions of courts and tribunals. However, the wording seems specifically to limit this ground to the initial publication of the judgement by the Ministry of Justice, rather than any further processing by third parties. Schedule 11 of the Data Protection Act 2018 provides exemptions under Part 4 of the Act, including certain data protection principles and the rights of data subjects. Paragraph 3, Schedule 11 concerns information required to be disclosed by law or in connection with legal proceedings, including where the disclosure of the data is "necessary for the purpose of obtaining legal advice" or "is otherwise necessary for the purposes of establishing, exercising or defending legal rights". Whilst it is undoubtedly the case that many of the start-ups operating in this space seek to help individuals obtain legal advice or assist in the exercise of their legal rights, it is less clear whether their contributions fulfil the requirement of *necessity* set out Paragraph 3, Schedule 11. The I.C.O. has acknowledged in its guidance on complying with data protection laws that a demonstration of necessity is less demanding than the processing being essential.[66] Nevertheless, it remains unclear at this stage of the development whether commercial entities would be able to demonstrate the necessity of their tools in assisting citizens

[65] L. Street and D. Hansen, "Who Owns the Law? How to Restore Public Ownership of Legal Publication" (2019) 26(2) J. Intell. Prop. L. at 205.
[66] I.C.O., "How do we apply legitimate interests in practice?" < https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/legitimate-interests/how-do-we-apply-legitimate-interests-in-practice/#LIA_process>.

in establishing, exercising, and defending their legal rights.[67] The implications for the purposes of this article are that the exemptions listed in the Data Protection Act 2019 appear to not be applicable to the secondary uses of data examined here.

With this in mind, we do not confine the present analysis to particular organisations, but rather proceed from the controllership or processing of textual justice data and develop an argument that applies to various types of actors, from judges and clerks to policymakers and administrative officials. As the following section will explain, one of the key challenges for such data controllers and data processors will be the identification of the appropriate legal basis to share justice data for lawtech innovation purposes. It is the central aim of this article to provide clarity on this critical question.

IV. ‘‘Legitimate’’ commercial interests and the public interest: three constellations

Identifying the appropriate legal basis for the processing of justice data in this context becomes quite challenging due to the uneasy relationship between the “public interest” under article 6(1)(e) GDPR and “legitimate interests” under article 6(1)(f) GDPR, the two most prominent options for a legal basis. The text of these legal bases reads:

> **6(1)(e) Public interest** - processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller.
>
> **6(1)(f) Legitimate interests** - processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.

[67] This does present a potential paradox as commercial entities would require data to demonstrate the necessity of their tools yet can only legally process data when such processing is already demonstrated to be necessary.

Article 6(1) also mentions that "legitimate interests" shall not be invoked by "public authorities in the performance of their tasks".[68] This is reflected in I.C.O.'s guidance,[69] which clarifies that for purposes "outside the scope of (…) tasks *as a public authority* (emphasis added), legitimate interests may be relied upon by public authorities". It is important to clarify that "outside the scope of tasks" does not mean that something is *ultra vires,* as there might be many "unglamorous, but important"[70] administrative powers of public bodies that are lawful, yet not precisely part of their organisational mission as public authorities. For example, a ministerial department may have such powers as the power to form contracts, conveying property and making ex gratia payments.[71] Arguably, sharing justice data for lawtech innovation purposes can be considered as falling within the ambit of administrative powers that are not exactly part of a public authority's tasks, yet resemble the administrative power of conveying property and forming contracts.

Beyond this caveat, the text of the Regulation allows scope for both public and private actors to invoke these two grounds for processing, potentially even in combination to justify different aspects of the same broader data processing activity. For example, a private water company may invoke the "public interest" as a legal basis for receiving personal data of citizens that is necessary to supply an area of the country with water,[72] as such companies, in I.C.O.'s words, carry out "functions of public administration and exercise special legal powers to carry out utility services in the public interest" under the Environmental Information Regulations (EIR) 2004.[73] At the

[68] Article 6(1) GDPR: "Point (f) of the first subparagraph shall not apply to processing carried out by public authorities in the performance of their tasks".
[69] I.C.O., "Legitimate interests" <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/lawful-basis/a-guide-to-lawful-basis/lawful-basis-for-processing/legitimate-interests/#:~:text=If%20you%20are%20a%20public,consider%20legitimate%20interests%20where%20appropriate>.
[70] A. Perry, "The Crown's administrative powers" (2015) 131 L.Q.R. at 652.
[71] *Rederiaktiebolaget Amphitrite v The King* [1921] 3 K.B. 500 at [503] per Rowlatt J.; *Robertson v Minister of Pensions* [1949] 1 K.B. 227 at [231] per Denning J.; *Crown Lands Commissioner v Page* [1960] 2 Q.B. 274 at [287-288] per Lord Evershed M.R..
[72] I.C.O., "Public task" <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/public-task/>.
[73] I.C.O., "Public authorities under the EIR" <https://ico.org.uk/for-organisations/foi-eir-and-access-to-information/freedom-of-information-and-environmental-information-regulations/public-authorities-under-the-eir/#public1>.

same time, they may invoke “legitimate interests” to further process a very small and aggregated amount of that data to improve its marketing strategy in the future.

Nonetheless, regulatory guidance in the UK and the EU has adopted a narrower approach with regard to the entities that can legitimately invoke the public interest under article 6(1)(e) in cases less clear-cut than the provision of a public utility like water or energy. Data protection authorities have often sought to associate the “public interest” under article 6(1)(e) with the duties and responsibilities of public-sector bodies, advising the latter to think carefully before relying on other grounds for processing, more appropriate for private-sector bodies, like “legitimate interests” under article 6(1)(f) GDPR.[74] In its opinion 06/2014, the Article 29 Working Party, having considered the draft GDPR,[75] associated reliance on the “public interest” ground with public-sector data sharing uses, either for public bodies with official authority (e.g. local government bodies) or for private entities which are mandated to disclose data to public bodies with official authority.[76]

This approach in regulatory guidance became particularly evident in the context of using data to combat the COVID-19 pandemic. CNIL, the French Data Protection Authority (DPA), and Garante, the Italian DPA, opined on the powers of public and private sector data controllers to collect and process health data with a view to curbing the spread of the disease. Both authorities advised private sector employers not to gather systematic data about their employees’ health status, as this would be an activity that serves “public health” and should be carried out by competent

[74] See e.g. the guidance by the UK I.C.O., “When can we rely on legitimate interests?” <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/legitimate-interests/when-can-we-rely-on-legitimate-interests/>. The same view has been adopted by the French and Italian Data Protection authorities regarding data processing for public health purposes, see A. Mole et al, “COVID-19 in the workplace: differing guidance from data protection authorities” <https://www.twobirds.com/en/news/articles/2020/global/covid19-in-the-workplace-guidance-from-data-protection-authorities>.

[75] Article 29 Working Party, “Opinion 06/2014 on the "Notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC” at 5. This opinion was then cited with approval (‘provides useful information’) in the Article 29 Working Party “Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679”, which have been subsequently endorsed by the European Data Protection Board, the successor to the Article 29 Working Party.

[76] Article 29 Working Party, “Opinion 06/2014 on the "Notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC” at 21-23.

authorities.[77] At the same time, a similar measure of data collection and processing would be legitimate to employ in the case of e.g. border control officers at an airport. Before the GDPR, the Article 29 Working Party had similarly confined invocation of the "public interest" ground to public bodies with official authority, such as "local government bodies" and "tax authorities", or to private bodies mandated to disclose data to public bodies with official authority.[78] The limitations emerging from this divide for private entities seeking to access publicly-held data, as is also the case with court judgments, with a view to producing insights that serve the public good, i.e. access to justice in our case, are clear.

This divide in regulatory guidance has been criticised in the health data sharing context by the W.H.O.,[79] whereas the UK Government has recently stressed that the private sector played a "crucial role in helping health providers tackle the COVID-19 virus".[80] This does not suggest, of course, that uncritical engagement with the private sector without adequate safeguards is an advisable course of action. As recent examples such as the case of PPE procurement by the UK public sector during COVID-19 illustrate, there needs to be robust oversight of public-private partnerships that ensures that public interest aims are effectively served without abusing executive discretion and denying transparency.[81]

An analysis of the applicable law to share justice data for Lawtech innovation in the UK shall carefully consider the existing regulatory divide when seeking to respond to the compatibility of different lawful grounds for processing for data controllers and processors in this space. Informed by these limitations in the current state of the law, we are advocating here for a conceptualisation of potential avenues for data sharing

[77] A. Mole et al, "COVID-19 in the workplace: differing guidance from data protection authorities" <https://www.twobirds.com/en/news/articles/2020/global/covid19-in-the-workplace-guidance-from-data-protection-authorities>.
[78] Article 29 Working Party, "Opinion 06/2014 on the 'Notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC'" at 21-23.
[79] World Health Organisation, "COVID-19 Strategy Update" (14 April 2020) https://www.who.int/docs/default-source/coronaviruse/covid-strategy-update-14april2020.pdf.
[80] Department for Digital, Culture, Media & Sport, "Data: A new direction" (10 September 2021) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1022315/Data_Reform_Consultation_Document__Accessible_.pdf>.
[81] UK Parliament, "COVID-19: Government procurement and supply of personal protective equipment: Inquiry" <https://committees.parliament.uk/work/731/covid19-government-procurement-and-supply-of-personal-protective-equipment>.

that can effectively promote access to justice without jeopardising the robust protection of data protection rights of the users of the justice system. Since this is an inherently complex and multi-layered space in terms of data flows and relationships between different public and private bodies, our conceptualisation does not seek to be reductive, aiming to provide a one-size-fits-all solution for what different data collection and processing may be needs in an ever-evolving area of practice. With that in mind, our conceptualisation articulates the possibility of *three data sharing constellations* by reference to data controllership and contribution to access to justice in order to establish lawful processing:

A. "Public interest" processing – when legal analytics exclusively serve access to justice purposes, with the public body remaining the sole data controller and the commercial entity being only a data processor who cannot re-purpose the data to pursue other for-profit purposes – public interest under article 6(1)(e) GDPR applies.

B. Hybrid processing – when a public-private data sharing collaboration is established between the public body and the commercial entity, with a view to serving both access to justice and commercial interests; both entities may re-purpose the data as controllers within prescribed safeguards and limitations - both public interest under article 6(1)(e) GDPR and legitimate interests under article 6(1)(f) GDPR apply, with the latter being invoked by the private actor as a separate legal basis at the point at which the private actor acquires controllership.

C. "Legitimate interests" processing – when legal analytics rely on legitimate interests to serve the commercial interests of the private entity to provide its services to its clients, and the public body relies on the interests of the third party to provide relevant data access; both entities may re-purpose the data as controllers within prescribed safeguards and limitations – legitimate interests under article 6(1)(f) GDPR applies.

As becomes clear from the identified constellations, "controllership" of the data and the scope of the "purpose" of data processing become central considerations in determining the lawfulness of sharing justice data for Lawtech innovation. To an extent, these two legal notions are intertwined: the data controller, under articles 4(7)

and (8) GDPR, is the entity which "determines the purposes and means of the processing of personal data", whereas data processors do the activity of processing "on behalf of the controller".[82] Different configurations and arrangements with regard to controllership impact on the conformity of parties to justice data sharing with the legal requirements under lawful grounds for processing. Under our first constellation, i.e., public interest processing, it would be relatively straightforward for public authorities to invoke "public interest" grounds, as the data controllers and for private entities to borrow this legal basis operating solely as data processors. Both sides would, of course, have to agree on the pre-specified means and purpose of processing through a comprehensive data sharing agreement including provisions on safeguards for data subjects' rights and data security organisational measures, as it often happens in the area of public-private data sharing for law enforcement (e.g., private security experts analyse police data to protect public safety and the vital interests of data subjects).[83]

How satisfactory would the first constellation be for participants in justice data sharing for Lawtech innovation in the UK? In the absence of empirical studies on point, several difficulties with confining our analysis to "public interest" processing can nonetheless be identified. Data-driven innovation, in the Lawtech and other sectors, is inherently exploratory and requires numerous iterations before the specific scope and techniques of processing and analysis are fully specified.[84] For example, a recent systematic study which aimed to predict judicial decisions of the European Court of Human through natural language processing (hereafter NLP) techniques pointed out the problem of data access barriers to additional types of data apart from judgments, "especially lodged applications and briefs",[85] that would help this type of research. With that in mind, both public authorities and private entities would be likely to welcome some more flexibility as regards the terms and scope of data processing, especially if the aim is to establish a long-term and sustained data sharing collaboration. This

[82] Articles 4(7) and (8) of the EU GDPR.
[83] N. Purtova, "Between the GDPR and the Police Directive: navigating through the maze of information sharing in public–private partnerships" (2018) 8(1) I.D.P.L. 45 at 52.
[84] For a similar observation in the case of big data analytics within social-scientific research see D. Erdos, "Systematically Handicapped? Social Research in the Data Protection Framework" (2011) 20 I.C.T.L. at 83: pre-determining the scope of processing would be hard to reconcile with the "fluid and norm- challenging nature of a social science research endeavour" supported by big data analytics.
[85] N. Aletras et al, "Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective" (2016) Peer J Computer Science.

flexibility is not present under the first constellation, even if the data sharing agreement allowed more experimentation, as mandatory data protection law would require a clear and exclusive connection between data processing for research and innovation purposes and the public interest. It might be the case, however, that lawtech innovation entities pursued various exploratory stages of processing before arriving at a result that can be shown to be even remotely relevant to the public interest.

In our second constellation, i.e., “hybrid processing”, the boundaries between data controllers and data processors become more porous. Access to justice, as an aim that serves the public interest in the context of the justice system, remains a primary consideration that drives data sharing on behalf of the public authority. It is accepted, however, that the commercial entity might need to re-purpose the means and processing of data in the interest of exploring further capabilities and analytical techniques that can go beyond the state of the art and provide better analytical insights about the justice system. This means that the commercial entity exercises “controllership” over the data, to the extent that their actions do not exclusively serve a pre-determined access to justice policy goal as agreed with the public authority, but also develop legal analytics that are of proprietary interest to the commercial entity itself, for example in using their technology to provide individualised legal advice to consumers or selling the right to using their technology to a law firm.

This constellation allows more flexibility, which might be welcomed by both sides in data processing, whilst also creating legal complications with regard to the second data protection principle, i.e. purpose limitation. Can data originally shared to further the “public interest” be processed for a different purpose, i.e. legitimate commercial interests? As long as the data processing carried out by the commercial entity can be considered “research”, one could argue that the activity would be exempt from the purpose limitation principle due to the interplay between Article 5(1)(b) and Recital 50 GDPR. These two provisions regulate the conformity of further processing for scientific research purposes with the purpose limitation principle:[86]

[86] Article 5(1)(b) and Recital 50 GDPR.

> **Article 5(1)(b):** further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes
>
> **Recital 50:** further processing for (…) scientific (…) research purposes or statistical purposes should be considered to be compatible lawful processing operations

Taken at face value, provided that article 89 safeguards like pseudonymisation are applied, data controllers who have collected their data lawfully would be able further to disclose it for research purposes without limitation as to which entity obtains access or the aims served by the data share.[87] As some of us have pointed out elsewhere, however, article 5(1)(b) and recital 50 GDPR should only be read as addressing purpose limitation rather than as covering all the data protection principles, including lawful basis.[88] Considering the sensitivity of the data in the context of the justice system, particularly in the case of the most vulnerable users of the system,[89] a more conservative interpretation of these provisions would be preferable, requiring a separate treatment of the lawful basis underpinning the sharing justice data for Lawtech innovation in the UK.[90]

What does this mean in practice for public authorities and commercial entities interested in engaging in hybrid processing? In principle, it will be crucial to observe

[87] The UK D.P.A., the I.C.O. seem to endorse this reading: "If your purposes change over time or you have a new purpose which you did not originally anticipate, *you may not need a new lawful basis as long as your new purpose is compatible with the original purpose*". [Emphasis added] See I.C.O., "Lawful basis for processing" <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/>.

[88] J. Bell et al, "Balancing Data Subjects' Rights and Public Interest Research" (2019) 5 E.D.P.L. 43 at 50.

[89] N. Byrom, "Digital Justice: HMCTS Data Strategy and Delivering Access to Justice" (The Legal Education Foundation 2019) <https://research.thelegaleducationfoundation.org/blog/digital-justice-hmctsdata-strategy-and-deliveringaccess-to-justice> at 4.13 and 4.33.

[90] On this matter, also see a relevant opinion by the Article 29 Working Party, "Opinion 03/2013 on purpose limitation" <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf> 33.The Article 29 Working Party rejected the idea of conflating the notion of 'compatibility' under the purpose limitation principle with the notion of 'lawful grounds' for processing under Article 6: "Further, these provisions also confuse two different concepts: the notion of 'compatibility' under Article 5(b) of the proposed Data Protection Regulation and the notion of 'lawful ground' under Article 6. As explained earlier, these two requirements are cumulative".

the proportionate relationship between the "public interest" and the "legitimate" commercial interests that may be served through the further processing of justice data. While the two are not mutually exclusive, substantial and tangible contributions to the public interest (access to justice) will be needed to justify the invocation of the relevant lawful ground for processing by the public authority. The degree to which commercial interests are furthered is also crucial. An internal exploratory analysis that seeks to improve company knowledge and IPR, or individualised legal advice provision at the fraction of the regular cost of specialised legal advice, are arguably more consistent with the public interest basis compared to other types of uses that could rely on justice data (e.g. employee surveillance analytics).[91] Provided that this ad-hoc assessment yields the conclusion that sufficient contributions are made to access to justice, the commercial entity will also need to conform with certain requirements under the "legitimate interests" ground under 6(1)(f) GDPR, which will be analysed under section 4 of this article. Conformity with the legal basis principle is only one part of creating a robust governance framework.

Is there, then, any scope for justice data sharing for Lawtech innovation in the UK without invoking the "public interest" lawful ground for processing and, thus, the potential of the data sharing activity to serve access to justice purposes? In principle, as our third constellation illustrates, there is some scope for both public authorities and commercial entities to rely on legitimate interests under 6(1)(f) GDPR. The provision enables processing that is "necessary for the purposes of the legitimate interests pursued by the controller *or by a third party*" (emphasis added).[92] The CJEU has consistently held that commercial interests in developing a proprietary data-driven technology can seen as legitimate;[93] the same is implied in recital 47 GDPR on legitimate interests, which refers to purely commercial purposes such as "direct

[91] I. Ebert et al, "Big Data in the workplace: Privacy Due Diligence as a human rights-based approach to employee privacy protection" (2021) 8(1) Big Data & Society.
[92] GDPR Article 6(1)(f).
[93] See for example the provision of Internet search engine services in *Google Spain SL and Google Incorporated v Agencia Española de Protección de Datos ('AEPD') and Costeja González* (C-131/12) EU:C:2014:317; [2014] Q.B. 1022; [2014] 3 W.L.R. 659 at [73]: "processing such as that at issue in the main proceedings carried out by the operator of a search engine, that processing is capable of being covered by the ground in Article 7(f)"; or ensuring the interoperability of online media services in *Breyer v Germany* (C-582/14) EU:C:2016:779; [2020] 1 WLUK 618; (2020) 71 E.H.R.R. 17 at [64]: "the objective aiming to ensure the general operability of those services may justify the use of those data after consultation of those websites".

marketing purposes". Lawtech companies could directly invoke this ground, whereas public authorities could claim to share data on the basis of the legitimate interests pursued by such third parties. On the one hand, this constellation is less demanding from both entities participating in data sharing with regard to demonstrating a sufficient contribution to the "public good" of access to justice. This is no longer a requirement and the association with the commercial interests of the Lawtech entity will be far more straightforward to demonstrate.

On the other hand, it is important to note the additional limitations which stem from exculsive reliance on "legitimate interests" compared to the "public interest" as a lawful ground for processing. The clearest difference relates to the application of the legitimate interests balancing test, i.e., the assessment of whether the invoked interests "override" the "rights and freedoms" of data subjects.[94] By contrast, under the "public interest" lawful basis, it would be sufficient that the processing is "necessary" to meet the purpose of serving access to justice.[95] We elaborate on the requirements within the balancing test in the following section of this article. In the light of recent UK government announcements, there is a chance that the balancing requirement will not apply for case where are processed for "business innovation purposes aimed at improving services for customers", but the scope of this exemption needs further clarification. [96] Furthermore, when relying on the "public interest" ground for data processing in the context of research, data controllers have more exemptions available to them e.g. from the data subject right to object to "processing of personal data concerning him or her" as long as the exercise of this right would seriously impair the aim of processing.[97] Similarly, in the case of sensitive personal data sharing under article 9 GDPR, there is arguably more scope to justify data processing for a purpose in the "public interest" compared to "legitimate" commercial interests.[98] The UK Government has also announced that it is planning to introduce amendments to the

[94] GDPR Article 6(1)(f).
[95] GDPR Article 6(1)(e).
[96] Department for Digital, Culture, Media & Sport, "Data: A new direction" (10 September 2021) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1022315/Data_Reform_Consultation_Document__Accessible_.pdf>.
[97] GDPR Article 21(6).
[98] See, for example, articles 9(2)(g) and (j) GDPR: "processing is necessary for reasons of substantial public interest (…)" and "processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) based on Union or Member State law which shall be proportionate to the aim pursued".

application of the GDPR in the UK, clarifying that "further processing for an incompatible purpose may be permitted when it safeguards an important public interest".[99] If this were to be enacted into legislation, it may enable further sharing of justice data for lawtech innovation to a greater extent than under the current regime, as elaborated upon with regard to the first (lawful basis) and second (purpose limitation) data protection principles in this section.

There are also, however, barriers in exclusively relying on the public interest ground. One stems from article 6(3) GDPR, which reads:

> **Article 6(3):** The basis for the processing referred to in (…) (e) of paragraph 1 shall be laid down by: (a) Union law or (b) Member State law to which the controller is subject. The purpose of the processing shall be determined in that legal basis or, as regards the processing referred to in point (e) of paragraph 1, shall be necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller (…)

This provision may create complications with regard to the domestic legislation that determines the purpose of the processing which is necessary for the "task carried out in the public interest" or the "exercise of official authority vested in the controller". Does this power need to be enshrined in statutory legislation or could it also stem from a non-statutory legal source, e.g., a common law power or an implied power?[100] If the data sharing activity is deemed to serve "research" purposes, based on the conditions specified earlier in this article, the Digital Economy Act (DEA) 2017 simplifies the matter by providing an explicit statutory gateway:

> **Section 64(1):** Information held by a public authority in connection with the authority's functions may be disclosed to another person for the purposes of research which is being or is to be carried out.

[99] Department for Digital, Culture, Media & Sport, "Data: A new direction" (10 September 2021) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1022315/Data_Reform_Consultation_Document__Accessible_.pdf>.

[100] For common law powers, see A. Perry, "The Crown's administrative powers" (2015) 131 L.Q.R. at 652; for implied powers see the seminal UK Supreme Court judgments in *Ward v Commissioner of Police of the Metropolis and another* [2005] UKHL 32 | [2006] 1 A.C. 23; *R (New London College Ltd) v Secretary of State for the Home Department* [2013] UKSC 51; [2013] 1 W.L.R. 2358.

If this condition is not satisfied, however, it might be more difficult to rely on the "public interest" ground. It is true that there is no absolute requirement in article 6(3) GDPR that the domestic legislation is statutory so as to rule out reliance on non-statutory legal powers.[101] Nonetheless, UK human rights law mandates that data processing, particularly when it happens on a big scale, needs to satisfy the requirements of foreseeability and specificity of scope not to allow excessive discretion that infringes on the right to private life under article 8 ECHR.[102] While different in theory from the right to data protection under article 8 of the EU Charter of Fundamental Rights, which draws on values other than privacy,[103] in practice the two rights have been interpreted by UK courts as effectively the same. For example, in *Christian Institute,* the UK Supreme Court found that discussing the compatibility of the impugned legislation with article 8 E.C.H.R. exhausts the issue of personal data protection.[104] Experience shows that public bodies in the UK are more likely to err on the side of caution by not promoting public interest research through data sharing rather than engage in data sharing they do not see as clearly empowered by the law.[105]

Considering not only the legal requirements, but also the organisational interests of both parties and the practicalities of justice data sharing, our second constellation, i.e. "hybrid data processing", is very likely to emerge as the preferrable option in practice. With that in mind, it is important to turn our attention to "legitimate interests" as integral within this constellation to justify further data processing by commercial entities. "Legitimate interests" may also become the sole basis that can be relied upon in case regulatory cautiousness or a narrow interpretation of article 6(3) GDPR nudge data controllers against relying on the "public interest" ground. Hence, we now provide a

[101] J. Bell et al, "Lawful disclosure of administrative data for research purposes in the UK" (2019) 2(3) J.D.P.P. at 264.
[102] *S and Marper v United Kingdom* [2008] 12 WLUK 117 | (2009) 48 E.H.R.R. 50.
[103] Such as non-discrimination, the requirement of fair processing, and the role of consent within the legislative framework, see P. De Hert and S. Gutwirth, "Data Protection in the Case law of Strasbourg and Luxembourg: Constitutionalism in Action" in S. Gutwirth et al (eds.) *Reinventing Data Protection?* (Berlin: Springer, 2009) Ch. 1.
[104] *Christian Institute v Lord Advocate* [2016] UKSC 51 at [104] per Lady Hale, Lord Reed and Lord Hodge.
[105] G. Laurie and L. Stevens, "Developing a Public Interest Mandate for the Governance and Use of Administrative Data in the United Kingdom" (2016) 43 J.L.S. 360.

more thorough examination of the application of the legal test under article 6(1)(f) GDPR in the justice data context.

V. The Boundaries of Legitimacy: Criteria and Safeguards

Reliance on legitimate interests for lawtech innovation has been complicated by the recent developments related to OpenAI's ChatGPT. More specifically, Garante, the Italian data protection authority temporarily banned ChatGPT in Italy and launched an investigation into OpenAI for suspected breaches of the GDPR. The Italian DPA alleged that, in respect of collecting data to train the model's algorithms, Open AI collects personal data from users without a lawful basis and without ensuring the processing of accurate data. Garante's decision raises a much broader question: can commercial entities claim a legitimate interest to process personal data for the purposes of developing and improving advanced analytics?

As discussed in Section 3, a reliance on the "legitimate interests" basis to process data requires a careful examination of the specific circumstances. There are also further caveats to its applicability as it can only be used by a public authority when outside the scope of its tasks as a public authority. The I.C.O. specifically refers to the pursuit of commercial interests as an end that may permit a public authority to rely on this basis.[106] However, it remains unclear as to how far a hybrid purpose, of commercial interests and improving access to justice, could preclude the reliance on this basis. We, therefore, propose that the "legitimate interests" ground may be a route to demonstrating the lawfulness of certain secondary uses of justice data discussed in this article.

Whilst the data processing at the heart of this piece is novel, we argue that guidance provided by the I.C.O. on the "legitimate interests" basis for processing and existing case law can be tailored to appropriately govern the processing of justice data in this context. The test requires a data controller to ensure their processing passes three

[106] I.C.O., "Legitimate interests", <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/legitimate-interests/>.

steps to ensure compliance with the law.[107] A data controller must pursue a legitimate interest, the processing must be necessary for that purpose, and the individual's interests must not override the legitimate interest for the data processing to be lawful. The rest of this section will consider these three tests, i.e., the "purpose test", the "necessity test", and the "balancing test" to the data processing activity explored in this article. The application of these tests will be supported using the I.C.O.'s legitimate interests' assessment (hereafter LIA),[108] which expands upon the relevant considerations and factors when considering the tests. The aim is to demonstrate the new partnerships fostered by the digitalisation of the UK's justice system and the innovations in data analytics that can be applied to justice data may still be appropriately governed by existing mechanisms.

1. Demonstrating Legitimacy: Finding a Legitimate Interest

The first step in relying on legitimate interests is to identify a genuine and legitimate reason for processing personal data. Official guidance from the Article 29 Data Protection Working Party (hereafter A29WP) outlines two conditions for the legitimacy of the interest of the data controller: it must represent a "real and present interest" and it must be "sufficiently clearly articulated".[109] The guidance issued by A29WP acknowledged that an "interest" is broader than a "purpose" as the former relates not to the aim of the processing but rather the "broader stake that a controller may have in the processing, or the benefit that the controller derives - or that society might derive - from the processing".[110] This allows the understanding of the "benefit" accrued

[107] I.C.O., "Legitimate interests", <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/legitimate-interests/>.
[108] I.C.O., "How do we apply legitimate interests in practice?" < https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/legitimate-interests/how-do-we-apply-legitimate-interests-in-practice/#LIA_process>.
[109] Article 29 Working Party, "Opinion 06/2014 on the "Notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC" at 24. This opinion was then cited with approval ('provides useful information') in the Article 29 Working Party "Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679", which have been subsequently endorsed by the European Data Protection Board, the successor to the Article 29 Working Party.
[110] Article 29 Working Party, "Opinion 06/2014 on the "Notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC" at 24.

through data processing to encompass the data controllers, data processors, data subjects, third parties, and the broader public. Nevertheless, some specificity of this interest is required to enable the balancing test between the legitimate interests and the rights of the data subject.

This has been interpreted widely by the CJEU in practice, encompassing not only the protection of legal rights such as the property, health, life of family and the individual,[111] or obtaining the information of a person who damaged the controller for legal action purposes,[112] but also purely commercial interests such as the provision of Internet search engine services[113] or ensuring the operability of online media services.[114] Indeed, the CJEU has noted that Member State have a broad margin of discretion in determining what falls within the scope of a legitimate interest.[115] As AG Bobek noted in *Rigas*,[116] the concept is "elastic enough" to encompass a range of considerations beyond the traditional categories of health and family life. The Court has also clarified that Member States are not allowed to categorically dismiss or exclude the possibility of processing certain categories of personal data to be based on legitimate interests.[117] The wide construction of legitimate interests is further supported by GDPR recital 47, which makes explicit reference to fraud prevention and direct marketing purposes, both potentially referring to purely commercial interests of the controller, as falling within the ambit of legitimate interests. This is in line with I.C.O. guidance in the UK. The I.C.O. has acknowledged that "legitimate interests" is a broad term, excluding unethical and unlawful data processing activities but potentially incorporating trivial and controversial uses of data.[118] The I.C.O. guidance also explicitly states that the

[111] *Ryneš v Úřad pro ochranu osobních údajů* (C-212/13) EU:C:2014:2428 [2015] 1 W.L.R. 2607; *TK v Asociatia de Proprietari bloc M5A-ScaraA* (C-708/18) EU:C:2019:1064 [2020] 1 W.L.R. 2286.
[112] *Valsts Policijas Rigas Regiona Parvaldes Kartibas Policijas Parvalde v Rigas Pasvaldibas SIA Rigas Satiksme* (C-13/16) EU:C:2017:336 [2017] 4 W.L.R. 97.
[113] *Google Spain SL and Google Incorporated v Agencia Española de Protección de Datos ('AEPD') and Costeja González* (C-131/12) EU:C:2014:317; [2014] Q.B. 1022; [2014] 3 W.L.R. 659.
[114] *Breyer v Germany* (C-582/14) EU:C:2016:779; [2020] 1 WLUK 618; (2020) 71 E.H.R.R. 17.
[115] *Asociacion Nacional de Establecimientos Financieros de Credito (ASNEF) v Administracion del Estado* (C-468/10) EU:C:2011:777 [2011] E.C.R. I-12181 [2011] 11 WLUK 715.
[116] *Valsts Policijas Rigas Regiona Parvaldes Kartibas Policijas Parvalde v Rigas Pasvaldibas SIA Rigas Satiksme* (C-13/16) EU:C:2017:336 [2017] 4 W.L.R. 97.
[117] *TK v Asociatia de Proprietari bloc M5A-ScaraA* (C-708/18) EU:C:2019:1064 [2020] 1 W.L.R. 2286; *ASNEF-EQUIFAX Servicios de Informacion sobre Solvencia y Credito SL v Asociacion de Usuarios de Servicios Bancarios (AUSBANC)* (C-238/05) EU:C:2006:734 [2006] E.C.R. I-11125 [2006] 11 WLUK 540.
[118] I.C.O., "Legitimate interests", <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/legitimate-interests/>

interests of third parties and commercial interests may both fulfil the requirement of a legitimate interest.[119]

What would be the legitimate interest of lawtech innovation actors for processing personal data for developing legal analytics? Under official guidance,[120] an interest is more likely to be perceived as legitimate if it corresponds with a general public interest or a third party's interest. Considering that legal analytics can provide various benefits for a wide variety of users, including but not limited to personalised legal advice, evaluation of the possibilities for a successful legal claim, and improved education about the law, there are reasons to believe that lawtech actors have a legitimate interest in developing and improving legal analytics as an innovative and widely beneficial AI technology. Furthermore, there is a collective benefit in removing barriers to access to justice and a better understanding of the workings of the justice system, whereas the desirability of encouraging innovation and stimulating economic growth within this sector may fall within the scope of the broad understanding of a legitimate interest that is promoted by regulators.

Hence, the guidance regarding what constitutes a "legitimate interest" does not appear to preclude a reliance on this ground, so long as the commercial entity's activities are not too speculative. However, it is important to note that the less compelling a legitimate interest is, the more likely it will be overridden at the "balancing test" stage, whereby the legitimate interest is balanced against the individual's interests. Therefore, if reliance on the processing undertaken to support the societal interests in improving access to justice are excluded, as tasks within the scope of a public authority acting as a public authority, the legitimate interest here may be slight.

2. Demonstrating Legitimacy: The Necessity Test

The second step in relying on legitimate interests is to show that the processing of personal data is necessary for pursuing the identified legitimate interest. Under official guidance, this means that there should be no "less invasive means available to serve

[119] I.C.O., "Legitimate interests", <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/legitimate-interests/>
[120] Article 29 Working Party, "Opinion 06/2014 on the "Notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC" at 24.

the same end".[121] As the CJEU has stressed, processing would be considered necessary if the same legitimate interest "cannot reasonably be as effectively achieved by other means less restrictive of the fundamental rights and freedoms of data subjects",[122] particularly privacy and data protection. It should be noted that this is not an overly strict interpretation of "necessity", focusing rather on the "reasonable effectiveness" of other means compared to enquiring into whether the purpose could be achieved at all through other means. In fact, AG Bobek argued in *Rigas* that the necessity requirement should not turn the:

> "Realisation of a legitimate interest into a Kafkaesque treasure hunt, strongly resembling an episode of Fort Boyard, in which the participants are sent from one room to another to collect partial clues to eventually work out where they are supposed to go"[123]

Similarly, according to the I.C.O., "necessary" is a bar lower than "essential". Instead, the processing must be "targeted and proportionate" in light of the purpose of the processing.[124]

Is the processing of personal data that are part of the justice system necessary for pursuing the legitimate interest in developing and improving legal analytics? In principle, lawtech innovators need to process personal data from users of the justice system to train their models and improve their performance. Without processing personal data, legal analytics would not be able to function properly or provide high-quality advice and outputs. Crucially, alternative yet piecemeal sources of information do not prevent the assertion of a legitimate interest in processing data from a single comprehensive source. This means that the possibility of commercial entities being able to trawl a vast array of databases and information sources to collect the same data would not prevent a legitimate interest legitimising the processing of this data from a more comprehensive data source as the one potentially offered by the Ministry

[121] Article 29 Working Party, "Opinion 06/2014 on the "Notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC" at 24.
[122] *TK v Asociatia de Proprietari bloc M5A-ScaraA* (C-708/18) EU:C:2019:1064 [2020] 1 W.L.R. 2286.
[123] *TK v Asociatia de Proprietari bloc M5A-ScaraA* (C-708/18) EU:C:2019:1064 [2020] 1 W.L.R. 2286 at [75].
[124] I.C.O., "What Is the 'Legitimate Interests' Basis?" <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/legitimate-interests/what-is-the-legitimate-interests-basis/#when_is_processing>.

of Justice and other public bodies. Indeed, there is a clear legitimate interest in processing such datasets as this would translate into potentially significant savings in terms of time and cost.

Nonetheless, the range of data processing activities that may be undertaken as part of lawtech innovation prevents a definitive answer to the necessity of the processing being given here. Those interested in determining the applicability of the legitimate interests ground for justifying their data processing must instead assess their intended processing on a case-by-case basis. The principles in Article 5 GDPR, particularly the provisions on data minimisation, and a precise understanding of the purpose may provide some guidance and clarity. Special regard should be shown for the sensitivity, relevance, and accuracy of the data to avoid an excessive and indiscriminate processing of personal data that goes beyond what is needed to achieve the legitimate interest. Data security and retention policies should be in the public domain for data subjects to be informed, as well as a clear demonstration of safeguards and mechanisms for ensuring the quality, accuracy, or integrity of the data. That said, there is nothing inherently problematic with reuses of justice data fulfilling the necessity test.

3. *Demonstrating Legitimacy: Balancing Legitimate Interests and Individuals' Interests*

The final test consists of considering the legitimate interest, in light of the interests of individuals to see if the latter should nevertheless override the former. This is why the definition of the legitimate interest may incorporate trivial or controversial purposes as such processing that will not be lawful where it is outweighed by the interests and fundamental rights of others. It must be stressed, however, that a negative impact on some individuals does not automatically override the legitimate interests identified at the first stage. The balancing test enables a data processor to examine the severity of the impact to determine if, on balance, the impact is warranted. This makes it all the more important to identify all the benefits accruing to the full range of actors as a result of the data processing to allow for a detailed examination of the balance between the legitimate interest and others' fundamental rights and interests.

Does the legitimate interest of lawtech innovators in developing and improving legal analytics outweigh or override the rights and interests of other parties? Similarly to the assessment of necessity of processing, we need to evaluate several factors to conclusively determine our answer to this question. An argument in favour of lawtech innovators would be that legal analytics provide significant benefits for users and society at large. Legal analytics and advanced software could enhance user experience, satisfaction, and engagement with the justice system by providing natural and engaging advice on various legal topics. It could also advance scientific knowledge, innovation, and progress by demonstrating the capabilities and potential of AI technology in law. Considering that domestic courts in the EU (e.g., in Lithuania)[125] have drawn on the criteria developed by the European Court of Human Rights in balancing rights, one could refer to the constant line of Strasbourg case law which shows that forms of genuine public interest often trump individual qualified rights such as the rights to privacy and data protection.[126] On the other hand, and considering the "freemium" business model under which most commercial large language models operate, one might argue that the real capabilities of lawtech analytics will largely remain behind paywalls and will not benefit society at large, unless users pay considerable fees for accessing them.

Another crucial consideration relates to whether the justice system data will be drawn from publicly available sources. This was found to be a critical consideration by the CJEU in *Rigas* and *ASNEF*.[127] Kamara and de Hert point out that the reasonable expectations of the data subject are an important consideration within the balancing test: the more foreseeable and acceptable from the side of the data subject the processing operation is, the more likely it is that the balancing exercise will find in favour of the data controller.[128] For example, submissions of litigants to the court may include their private information, as part of their right to private life under article 8 of

[125] N. Bitiukova, "Lithuanian Supreme Administrative Court Undertakes a Legitimate Interests Assessment in a Seminal Case on Journalistic Expression" (2022) 8(1) E.D.P.L. 128 at 133.
[126] *Eweida v United Kingdom* [2013] 1 WLUK 142; *Dahlab v. Switzerland*, dec., No.42393/98, ECtHR (Second Section).
[127] *Valsts Policijas Rigas Regiona Parvaldes Kartibas Policijas Parvalde v Rigas Pasvaldibas SIA Rigas Satiksme* (C-13/16) EU:C:2017:336 [2017] 4 W.L.R. 97; Case C-238/05, *ASNEF-EQUIFAX Servicios de Informacion sobre Solvencia y Credito SL v Asociacion de Usuarios de Servicios Bancarios (AUSBANC)* (C-238/05) EU:C:2006:734 [2006] E.C.R. I-11125 [2006] 11 WLUK 540.
[128] I. Kamara and P. de Hert, "Understanding the Balancing Act Behind the Legitimate Interest of the Controller Ground: A Pragmatic Approach" (2018) 12(4) Brussels Privacy Hub.

the European Convention on Human Rights, which they would not expect to see shared widely. Determining whether the right to privacy would override the lawtech company's legitimate interests would hinge on the context and circumstances of each case, e.g., to what extent open justice principles may dictate that parts of the litigants' submissions are made public in court.[129]

It may not be possible to speak of specific expectations towards the reuse of justice data for the purposes set out in this article but the broader point of the effective and efficient use of public resources, such as justice data, may be of relevance to this point. Whilst the relationship between the HMCTS, Ministry of Justice, and the legal tech start-ups may be a new development, partnerships between public and private entities have been a feature of public service delivery in the UK for decades. This, along with the recent decision to more broadly publish and archive judgement data via the FindCaseLaw initiative, suggests an understanding of individuals' reasonable expectations that may allow for the reuse of justice data on legitimate interest grounds.

That being said, there is still the need to consider such factors as the nature of the data, whereby the more sensitive the data and the more the data pertains to vulnerable data subjects, including children, the more likely it is to have an impact that outweighs the legitimate interest. As noted previously, the data which legal tech start-ups will find most useful are not necessarily those concerned with these attributes of the data subject.[130] Other data points, including those relating to the legal representatives and judiciary involved are more likely to provide insights of relevance. Finally, the balancing test also allows for data controllers and data processors to consider potential safeguards, which can "shift the balance" towards their pursuit of the identified legitimate interest. For example, it is important to be transparent about these uses of data, as a potential safeguard. The I.C.O. guidance highlights the importance of keeping a record of the LIA and its outcome. In the context of this relationship and

[129] S. Ahmed, "Online courts and private and public aspects of open justice: Enhancing access to court or violating the Right to Privacy?" (2023) 20 The Age of Human Rights Journal e7516-e7516.
[130] Although there might be cases where a specific aspect of a litigant's personal information (e.g., their race) might be pertinent to predictions of outcome. The ethics of relying on systemic prejudices to create predictive analytics is, naturally, a separate question and one which merits more attention. A recent effort to achieve fairness aware machine learning and tackle systemic biases in violence datasets against people of colour can be found in I. Pastaltzidis et al, "Data augmentation for fairness-aware machine learning: Preventing algorithmic bias in law enforcement systems" (2022) in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency* (pp. 2302-2314).

data processing activity, we would urge a careful and considered examination as to the possibility of publishing LIAs. The use of new technologies provides opportunities not just in terms of their outputs but also in the ability to adopt new processes that promote transparency.

VI. Conclusions: Reflecting upon the past and present of legal services.

Guided by the developments in the lawtech innovation space, this article explored developers' potential interactions with the existing regulatory framework to analyse the ways in which it is possible to capitalise on these advances whilst respecting individuals' rights and interests. In elucidating the potential opportunities made possible by innovation in this context we have demonstrated the range of benefits that may arise, encompassing commercial interests, the promoting of constitutional rights, and the potential for economic benefits to accrue to the public sector and individuals.

This panoply of benefits prompted the exploration of the different ways such data processing could be rendered lawful, drawing upon regulatory guidance and existing case law. Doing so, we advocated in favour of exploring the possibility of "hybrid processing" as means of rendering secondary uses of justice data in the lawtech innovation space as lawful. This ground provides the requisite flexibility to allow all those involved in the data processing to benefit, so long as the proportionate relationship between the "public interest" and the "legitimate" commercial interests that may be served is upheld. Nevertheless, in recognising the potential limits of this, we also provided a further analysis of the role of the "legitimate interests" grounds to demonstrate compliance with the law. The range of data processing practices performed in this context render an absolute answer to the lawfulness of such processing impossible: but there is clear scope for responsible processing to be rendered lawful.

The implications of this argument go beyond the governance of justice data and raise questions as to the potential future of the regulation of legal services more broadly. The exploration of the lawtech innovation space has drawn attention to emerging actors in the justice context and the potential significantly to contribute to

the aims of the justice system, whilst also requiring consideration of their existence as unregulated commercial entities. Whilst existing regulatory guidance in this context my adequately capture legal analytics, it is unclear as to whether this is applicable to all the aspects of legal services provided by commercial entities.