



The Legal Architecture of Explainability-by-Design

Operationalizing the GDPR's Right to an Explanation for Automated Decision-Making
Under the Data Protection by Design and by Default Principle.

Author: Giorgia Loredan

Supervisor: Professor Dr. Guido Westkamp

Table of Contents

| | |
|--|-----------|
| 1. INTRODUCTION | 1 |
| 2. EXPLAINABILITY: A ROAD OUT OF THE BLACK BOX | 3 |
| 2.1. BACKGROUND ON DEEP LEARNING AND THE ‘BLACK BOX’ PROBLEM | 3 |
| 2.2. XAI: WHAT, WHY, AND HOW? | 4 |
| 3. THE RIGHT TO AN EXPLANATION UNDER EU DATA PROTECTION LAW | 6 |
| 3.1. AUTOMATED DECISION-MAKING UNDER THE GDPR | 6 |
| 3.2. THE RIGHT TO AN EXPLANATION FOR AUTOMATED DECISIONS | 9 |
| 3.2.1. <i>STICKING TO THE LETTER OF THE LAW: GDPR PROVISIONS</i> | 9 |
| 3.2.2. <i>JURISPRUDENCE OF THE COURT OF JUSTICE</i> | 10 |
| 3.3. EXPLAINABILITY-BY-DESIGN: REGULATORY GUIDANCE | 12 |
| 4. MODELLING GDPR-COMPLIANT XAI ACCORDING TO ARTICLE 25 | 13 |
| 4.1. BUILDING GDPR-COMPLIANT EXPLANATIONS FOR DEEP LEARNING MODELS | 14 |
| 4.1.1. <i>BEFORE THE DECISION: GLOBAL EXPLANATIONS</i> | 14 |
| 4.1.2. <i>AFTER THE DECISION: LOCAL EXPLANATIONS</i> | 16 |
| 5. LIMITATIONS OF EXPLAINABILITY UNDER DATA PROTECTION LAW | 19 |
| 5.1. LEGAL RESTRICTIONS | 19 |
| 5.2. PRACTICAL RESTRICTIONS | 20 |
| 6. CONCLUSION | 22 |
| BIBLIOGRAPHY | 24 |

1. Introduction

Since being recognized as a ‘technological revolution’, artificial intelligence has become a key competitive edge for companies in today’s market.¹ The use of AI by organizations jumped 23% between 2023 and 2024, running parallel to record-high investments being poured into the private AI sector.² At the same time, paradoxically, almost all industry leaders express doubt as to whether they are fully prepared to implement the technology responsibly.³ A large part of their concern relates to the veil of opacity under which many machine learning (‘ML’) systems operate, especially those based on deep learning.⁴ The mass scale adoption of automated decision-making for commercial purposes raises a number of legal and regulatory challenges when the decisions cannot be understood or justified. In the European Union, automated decision-making is actively governed under the legal framework of the General Data Protection Regulation (‘GDPR’), which lays down obligations for developers, deployers, and providers of AI-based products. A fundamental – yet controversial – requirement prescribed by EU data protection law is explainability, or more precisely, the right to an explanation.

To date, significant research efforts have been dedicated to exploring the legal status of explainability under the GDPR,⁵ as well as under both the GDPR and the European Union’s Artificial Intelligence Act,⁶ with the main point of contention being whether or not the right to an explanation

¹ E. Papagiannidis et al., ‘Responsible artificial intelligence governance: A review and research framework’ (2025) 34(2) *The Journal of Strategic Information Systems*, p. 1, available at: <https://doi.org/10.1016/j.jsis.2024.101885>.

² N. Maslej et al., ‘Artificial Intelligence Index Report 2025’ (2025) AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2025, p. 3, available at: https://hai.stanford.edu/assets/files/hai_ai_index_report_2025.pdf.

³ C. Giovine and R. Roberts, ‘Building AI trust: The key role of explainability’ (2024) *McKinsey & Company*, available at: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/building-ai-trust-the-key-role-of-explainability>.

⁴ G. Pavlidis, ‘Unlocking the black box: analysing the EU artificial intelligence act’s framework for explainability in AI’ (2024) 16(1) *Law, Innovation and Technology*, p. 294, available at: <https://doi.org/10.1080/17579961.2024.2313795>.

⁵ R. Hamon et al., ‘Bridging the Gap Between AI and Explainability in the GDPR: Towards Trustworthiness-by-Design in Automated Decision-Making’ (2022) 17(1) *IEEE Computational Intelligence Magazine*, pp. 72-85, doi: 10.1109/MCI.2021.3129960; P. Dewitte, ‘AI Meets the GDPR: Navigating the Impact of Data Protection on AI Systems’ (2025) in N. A. Smuha (ed.) *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence*. Cambridge: Cambridge University Press (Cambridge Law Handbooks), pp. 133–157, available at: <https://doi.org/10.1017/9781009367783.010>; F. Sovrano et al., ‘Making Things Explainable vs Explaining: Requirements and Challenges Under the GDPR’ (2021) In: V. Rodríguez-Doncel et al. (eds) *AI Approaches to the Complexity of Legal Systems XI-XII. AICOL 2020, AICOL 2018 XAILA 2020. 13048 Lecture Notes in Computer Science*, pp. 169-182, Springer, Cham, available at: https://doi.org/10.1007/978-3-030-89811-3_12.

⁶ L. Metikoš and J. Ausloos, ‘The Right to an Explanation in Practice: Insights from Case Law for the GDPR and the AI Act’ (2024) *Law, Innovation and Technology*, pp. 1-36, available at: <https://doi.org/10.1080/17579961.2025.2469349>; M. Nisevic et al., ‘Explainable AI: Can the AI Act and the GDPR go out for a date?’ (2024) International Joint Conference on Neural Networks (IJCNN), Yokohama, Japan, 2024, pp. 1-8, doi: 10.1109/IJCNN60899.2024.10649994; B. A. Juliussen, ‘The Right to an Explanation Under the GDPR and the AI Act’ (2025) *MultiMedia Modeling: 31st International Conference on Multimedia Modeling*, MMM 2025, Nara, Japan, January 8–10, 2025, Proceedings, Part IV, pp. 184-197, available at: https://doi.org/10.1007/978-981-96-2071-5_14.

legitimately and meaningfully exists under EU law.⁷ Yet, an unexpected resolution to this debate emerged in the 2025 *Dun and Bradstreet Austria* (C-203/22) judgment by the Court of Justice of the European Union (hereafter ‘CJEU’), which arguably confirmed the right to an explanation for automated decision-making (‘ADM’) processes.⁸ In view of the Court’s recent decision, and of the potential effect a precedent such as this may set for the entire technology industry going forward, the present paper sets out, as a first point, to answer the following question: **To what extent do the provisions of the GDPR together with the CJEU’s judgment in *Dun & Bradstreet Austria* provide for a legally binding and enforceable right to an explanation of automated decision-making?**

Subsequently, this paper turns to the Data Protection by Design and by Default (hereafter ‘DPDD’) principle, promulgated under Article 25 of the GDPR. Under this provision, data protection is transformed from a collection of ideals into concrete legal obligations for controllers.⁹ That being said, an extensive review of the available literature reveals no previous studies have examined the right to an explanation from the perspective of Article 25. Albeit significantly overlooked in the explainability discussion, this provision warrants due attention from researchers, policymakers, and industry stakeholders for multiple reasons. Firstly, it functions as a core pillar of the GDPR, as it resolves the gap between law in theory and law in practice.¹⁰ Secondly, the DPDD obligation has been arguably recognized as a requirement of European human rights law,¹¹ elevating its importance beyond regulatory compliance to a matter of fundamental rights protection. Finally, as argued in Section 4.1 of this paper, EU regulatory guidelines identify an intrinsic and definitive link between explainability and the DPDD principle, thus justifying further analysis of its role in ensuring data protection.

In light of all foregoing facts, the second research question will be: **How can the right to an explanation be modelled in accordance with the requirements of Article 25?** By answering both research questions, the present paper seeks to not only fill a gap in the current literature but also provide crucial insights as to how developers and providers of AI may operationalize GDPR-compliant explanations in line with the DPDD principle. Going further, the final chapter of this dissertation will take a critical look at the legal and practical limitations of a right to an explanation, accompanied by recommendations intended to support industry actors navigate said limits in real-world applications.

⁷ See also: B. Fresz et al., ‘How Should AI Decisions Be Explained? Requirements for Explanations from the Perspective of European Law’ (2024) 7(1) Proceedings of the 2024 AAAI/ACM Conference on AI, Ethics, and Society, AAAI Press, pp. 438–450, available at: <https://doi.org/10.1609/aies.v7i1.31648>.

⁸ S. Rossetti, ‘The Court of Justice of the European Union confirms the existence of the right to explanation of automated decision-making’ (2025) *European Law Blog*, available at: <https://doi.org/10.21428/9885764c.flcc5989>; E. Valgaeren and T. Duquin, ‘CJEU confirms “right of explanation” in battle between trade secrets and algorithmic transparency’ (2025) *Stibbe*, available at: <https://www.stibbe.com/publications-and-insights/cjeu-confirms-right-of-explanation-in-battle-between-trade-secrets-and->

⁹ A controller is the natural or legal person determining the means and purposes of processing personal data, as per Article 4(7) GDPR.

¹⁰ C. Michelakaki and S. B. Vale, ‘Unlocking Data Protection By Design & By Default: Lessons from the Enforcement of Article 25 GDPR’ (2023) *Future of Privacy Forum*, p. 2.

¹¹ *Ibid.*

2. Explainability: A Road Out of The Black Box

2.1. Background on Deep Learning and the ‘Black Box’ Problem

From web searches to recommendations on e-commerce websites, ML powered technology has penetrated nearly all aspects of modern personal and professional life.¹² Fueled by the feasibility of new programming frameworks, broader access to necessary computing power, and the availability of Big Data, ML technology has seen growing levels of application over recent years.¹³ Deep learning (‘DL’), a particularly popular subset of ML algorithms that run on artificial neural networks, faces expedited integration into organizations as a key decision-making tool.¹⁴ DL-augmented decision-making has rapidly diffused across diverse industries including insurance, energy, retail, healthcare, banking, and transportation.¹⁵ That is not to mention the global,¹⁶ and regional,¹⁷ agendas geared at incentivizing business-driven adoption of DL tools.¹⁸

For many commercial applications, DL algorithms outperform their simpler, or ‘shallower’, ML counterparts due to improved learning capabilities and advanced processing techniques.¹⁹ The structural complexity of DL may be succinctly defined by three interrelated features: (i) high dimensionality, (ii) nonlinearity, and (iii) an innumerable number of parameters.²⁰ DL generates multiple layers of non-linear clusters to encode and transform information as representations in a distributed fashion.²¹ Other similarly learned changes occur in the training phase, during which the parameters and ‘weights’ of the model are adjusted as a means of refining performance and accuracy.²²

The accumulation of small, iterative edits during training, combined with the nonlinear dispersion of meaning, renders the reverse-engineering of a specific DL decision into an impractical task. Even when information can be retrieved regarding the decision-making process, the model’s self-

¹² Y. LeCun et al., ‘Deep learning’ (2015) 521 *Nature*, p. 436, available at: <https://doi.org/10.1038/nature14539>.

¹³ C. Janiesch et al., ‘Machine learning and deep learning’ (2021) 31 *Electronic Markets*, p. 685, available at: <https://doi.org/10.1007/s12525-021-00475-2>.

¹⁴ Y. R. Shrestha et al., ‘Augmenting organizational decision-making with deep learning algorithms: Principles, promises, and challenges’ (2021) 123 *Journal of Business Research*, p. 589, available at: <https://doi.org/10.1016/j.jbusres.2020.09.068>.

¹⁵ Ibid.

¹⁶ B. Ammanath and K. Firth-Butterfield, ‘How deep learning can improve productivity and boost business’ (2022) *World Economic Forum*, available at: <https://www.weforum.org/stories/2022/01/deep-learning-business-productivity-revenue/>.

¹⁷ See: Eurostat, ‘Usage of AI technologies increasing in EU enterprises’ (2025), available at: <https://ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20250123-3>.

¹⁸ A. B. Rashid and MD A. K. Kausik, ‘AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications’ (2024) 7 *Hybrid Advances*, Sections 4.8-4.9, available at: <https://doi.org/10.1016/j.hybadv.2024.100277>.

¹⁹ M. R. Karim et al., ‘Interpreting Black-box Machine Learning Models for High Dimensional Datasets’ (2023) 2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA), Thessaloniki, Greece, Abstract, doi: 10.1109/DSAA60987.2023.10302562.

²⁰ See: V. Hassija et al., ‘Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence’ (2024) 16 *Cognitive Computation*, pp. 46-47, available at: <https://doi.org/10.1007/s12559-023-10179-8>.

²¹ LeCun et al., *supra* note 13, p. 436.

²² J. Lu, ‘Gradient Descent, Stochastic Optimization, and Other Tales’ (2024), Preface, available at: <https://doi.org/10.48550/arXiv.2205.00832>.

learned representations may not be designed to correspond to human-understandable attributes or language.²³ Additionally, in the case of reinforcement-based DL systems, the prediction-making routes are continuously updated based on user feedback, making it virtually impossible to replicate, foresee, or translate their behavior.²⁴ Hence, by default, the decision-making process of most DL-based products takes an opaque, unintelligible, and untraceable form.

In recognition of this, the literature distinguishes between ‘white box’ models, interpretable to experts in the applicable domain, and ‘black box’ models, not directly interpretable and extremely demanding for experts to comprehend.²⁵ This taxonomy has been welcomed by policymakers and regulators when considering the ‘ethical’ deployment of black box systems at scale.²⁶ Their position is rooted in a serious concern regarding the inability of AI providers, deployers, or users to understand the outcomes of opaque systems. This sentiment mirrors the opinion of the European Data Protection Supervisor (‘EDPS’), which has found AI systems that obscure the underlying logic behind decisions to be “unacceptable”.²⁷ Ultimately, for both end-users and industries to adopt a given product, it should be perceived as reliable, or trustworthy,²⁸ which will not be possible when a system is characterized by unpredictability, opacity, strangeness, and absence of justification.²⁹

2.2. XAI: What, Why, and How?

One key approach to addressing the ‘black box problem’ revolves around transparency, by which complex AI systems are made understandable, and thus trustworthy. Transparency can shed light on both the internal structure of the model, such as its parameters, as well as the development process, including relevant design choices, usage of data, and testing procedures.³⁰ However, technical transparency alone does not suffice as a guarantee of interpretability. Although users may see behind the veil of opacity, it does not mean they will know what to look for, or how to effectively find meaning in all the information provided. In the same way, transparency also fails to recognize that different stakeholders interacting with complex AI models have varying interests, backgrounds, and competences. Users, sector-specific experts, system engineers, corporate officers, and regulatory entities will each require varying degrees of detail, or contextually tailored answers regarding a specific

²³ Pavlidis, *supra* note 4, p. 295.

²⁴ D. Kamarinou et al., ‘Machine Learning with Personal Data’ (2016) 247 *Queen Mary School of Law Legal Studies Research Paper*, p. 107, available at: <https://ssrn.com/abstract=2865811>.

²⁵ See: O. Loyola-González, ‘Black-Box vs. White-Box: Understanding Their Advantages and Weaknesses From a Practical Point of View’ (2019) 7 *IEEE Access*, pp. 154096-154113, doi: 10.1109/ACCESS.2019.2949286.

²⁶ The Royal Society, ‘Explainable AI: the basics’ (2019) ISBN: 978-1-78252-433-5, p. 8.

²⁷ EDPS, ‘TechDispatch: Explainable Artificial Intelligence’ (2023) p. 2, doi: 10.2804/802043.

²⁸ C. Véliz et al., ‘We might be afraid of black-box algorithms’ (2021) 47 *Journal of Medical Ethics*, 339–340, available at: <https://doi.org/10.1136/medethics-2021-107462>.

²⁹ B. Brožek et al., ‘The black box problem revisited. Real and imaginary challenges for automated legal decision making’ (2024) 32 *Artificial Intelligence and Law*, p. 436, available at: <https://doi.org/10.1007/s10506-023-09356-9>.

³⁰ Hamon et al., *supra* note 6, p. 75.

output, making general transparency unapplicable and unviable for guaranteeing algorithmic accountability.³¹

As a result, explainability becomes a necessity. While transparency unveils the inner technical corpus of DL systems, explainability presents the cognitive steps taken by the machine by providing clear and coherent reasoning for automated decisions.³² Moreover, for legal and regulatory reasons, knowing the *why* behind decisions is just as relevant as the *what*. As neatly summarized by Georgios Pavlidis, the “principle of explainability is neither a simple software problem nor just an intellectually challenging quest for knowledge; it is a prerequisite for accountability, fairness, public trust, and effective regulation and supervision”.³³ In that sense, explainability is arguably a more important, and more difficult, requirement than transparency with respect to enabling AI accountability, trustworthiness, and compliance with data protection law.

It is relevant to note that, in terms of content, explanations can take different forms depending on their audience. Generally, explanations may be classified as either “global” or “local”, each technique serving a unique purpose. Global explanations touch upon the overall operation of the model, its development and training, and its approved use environments.³⁴ On the other hand, local explanations justify individual algorithmic decisions in light of the given circumstances, taking into account, for example, particular features which influenced the respective outcome.³⁵ Collectively, these approaches guarantee the interpretability of automated decisions, allowing users to critically assess both the overall function of the model as well as the reasoning behind a specific result.

Since becoming an extensively researched field, Explainable AI (‘XAI’) continues to garner enormous academic, political, and commercial interest.³⁶ Several studies have demonstrated that XAI effectively enhances trust of end-users in automated decision-making.³⁷ These findings have been consequently translated into practice across multiple sectors, as XAI becomes an increasingly common

³¹ See: S. Grimmelikhuijsen, ‘Explaining Why the Computer Says No: Algorithmic Transparency Affects the Perceived Trustworthiness of Automated Decision-Making’ (2023) 83(2) *Public Administration Review*, pp. 241–262, available at: <https://doi.org/10.1111/puar.13483>.

³² EDPS, *supra* note 33, p. 5.

³³ Pavlidis, *supra* note 4, p. 296.

³⁴ W. Maxwell and B. Dumas, ‘Meaningful XAI Based on User-centric Design Methodology: Combining Legal and Human-computer Interaction (HCI) Approaches to Achieve Meaningful Algorithmic Explainability’ (2023) *CERRE*, p. 14, available at: <http://dx.doi.org/10.2139/ssrn.4520754>.

³⁵ *Ibid.*

³⁶ F. Herrera, ‘Reflections and attentiveness on eXplainable Artificial Intelligence (XAI). The journey ahead from criticisms to human–AI collaboration’ (2025) 121 *Information Fusion*, p. 1, available at: <https://doi.org/10.1016/j.inffus.2025.103133>.

³⁷ M. Xu and Y. Wang, ‘Explainability increases trust resilience in intelligent agents’ (2024) *British Journal of Psychology*, available at: <https://doi.org/10.1111/bjop.12740>; M. Zolanvari et al., ‘TRUST XAI: model-agnostic explanations for AI with a case study on IIoT security’ (2023) 10(4) *IEEE Internet of Things Journal*, pp. 2967–2978, doi: 10.1109/JIOT.2021.3122019; R. Tiwari, ‘Explainable AI (XAI) and its Applications in Building Trust and Understanding in AI Decision Making’ (2023) 7(1) *IJSREM*, doi: 10.55041/IJSREM17592; M. Perlmutter et al., ‘Impact of example-based XAI for neural networks on trust, understanding, and performance’ (2024) 188 *International Journal of Human-Computer Studies*, available at: <https://doi.org/10.1016/j.ijhcs.2024.103277>.

application in both Industry 4.0 and 5.0.³⁸ In the same way, XAI adoption positively correlates to ethical use of customer data in line with information governance frameworks, including – notably – the GDPR.³⁹ Still, explainability is no stranger in the European data protection sphere – in fact, far from one. The following section will therefore lay out the GDPR’s regime on ADM systems, before examining the right to an explanation and its corpus under EU data protection law.

3. The Right to An Explanation Under EU Data Protection Law

3.1. Automated Decision-Making Under the GDPR

A major driver for apprehension of black box models lies in the fact that opacity can have direct adverse effects on individuals when it hides fallacies in AI systems, including the existence of bias, inaccuracies, or hallucinations. Algorithms designed and trained on poor data, for example, may produce decisions that are harmful or discriminatory to individuals, ultimately infringing on their fundamental rights. Yet, black box models will hide such mistakes, meaning that harmful decision may be adopted without a proper checks and balances procedure in place.

For these reasons, Article 22 of the GDPR prohibits any decision “based solely on automated processing” of personal data when it produces either legal or similarly significant effects on the data subject concerned. Here, ‘based solely’ suggests an absence of meaningful human involvement in the decision-making process.⁴⁰ The second threshold of this prohibition, the ‘effects’, focuses on the impact of the automated decision, or in other words, the extent to which the decision affects the rights and freedoms of the data subject. As for ‘legal effects’, that may entail any result which affects an individual’s legal rights, or their legal status.⁴¹ Examples of this kind of effect, as outlined by the Article 29 Working Party, include the cancellation of a contract, the entitlement to or refusal of a social benefit, such as a housing benefit, and the denial of citizenship or admission to a country.⁴²

Data subjects may be further protected by the GDPR’s prohibition on ADM when the significance of the decision in question qualifies as *equivalent* or *similar* to that of a decision with legal effect. Automated decisions will therefore fall under the scope of Article 22 when they significantly affect the circumstances, behavior, or choices of the data subject, impact them in a prolonged or

³⁸ K. Nikiforidis et al., ‘Enhancing transparency and trust in AI-powered manufacturing: A survey of explainable AI (XAI) applications in smart manufacturing in the era of industry 4.0/5.0’ (2025) 11 *ICT Express*, p. 145, available at: <https://doi.org/10.1016/j.ict.2024.12.001>.

³⁹ O. O. Olateju et al., ‘Exploring the Concept of Explainable AI and Developing Information Governance Standards for Enhancing Trust and Transparency in Handling Customer Data’ (2024) 26(7) *JERR*, p. 257, available at: <https://doi.org/10.9734/jerr/2024/v26i71206>.

⁴⁰ Article 29 Data Protection Working Party (hereafter ‘Article 29 WP’), ‘Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679’ (2017), p. 20, available at: <https://ec.europa.eu/newsroom/article29/items/612053/en>.

⁴¹ *Ibid*, p. 21.

⁴² *Ibid*.

permanent way, or result in the exclusion, or discrimination, of that individual.⁴³ That may, for instance, look like the automatic refusal of an online credit application, or e-recruiting practices without meaningful human involvement.⁴⁴

Profiling, which equally falls within the remit of Article 22, may or may not be considered to have legal or similarly significant effects depending on the circumstances at hand. In some cases, profiling activities have been considered significantly impactful even when they did not directly strip any rights away from data subjects. For example, systems determining the likelihood of a job applicant being hired, even when the decisions are used to target applicants with employment counselling, have been ruled to have a significant effect.⁴⁵ Likewise, the profiling of individuals in a medical context may be prohibited when the decision excludes them from receiving additional medical attention.⁴⁶

As for online advertising, the consensus traditionally maintained that advertising based on profiling will usually not have significant effects on individuals.⁴⁷ However, increased integration of Big Data for commercial applications of DL-based ADM (hereafter ‘DLADM’), such as customer profiling, poses serious risks to the rights and interests of data subjects. Individuals that are tracked across different websites, devices, and services become more vulnerable to intrusive surveillance and manipulative design practices,⁴⁸ which ultimately degrades their data protection and fundamental human rights.⁴⁹ Additionally, DLADM processes often generate latent features, or inferences, that deduce personal information from inputs.⁵⁰ Notably, the collection and use of *inferred* personal data for profiling is also prohibited when it results in significant effects for data subjects.⁵¹ That includes situations where, for example, an automated decision based on inferred personal traits leads to differential pricing that excludes someone from certain goods or services.

It is important to mention that DLADM systems, especially those powered by Big Data, will often be in further conflict with the GDPR’s data protection principles.⁵² While this paper will not be examining the tension between DL technology and Article 5 GDPR, a brief but crucial note should be made of the fact that data subjects and regulators both have a vested interest in monitoring the operation of ADM activities also with respect to the data protection principles. The development and deployment

⁴³ Ibid.

⁴⁴ GDPR, Recital 71.

⁴⁵ Supreme Administrative Court Austria, *Verwaltungsgerichtshof* [2023].

⁴⁶ Finnish Data Protection Authority, *Tietosuojavaltuutetun toimisto* [2020].

⁴⁷ Article 29 WP, *supra* note 47, p. 22.

⁴⁸ Ibid.

⁴⁹ See: S. Zuboff, ‘The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power’ (2019) New York: PublicAffairs.

⁵⁰ See: V. Michael et al., ‘Algorithms that remember: model inversion attacks and data protection law’ (2018) 376(2133) *Philosophical Transactions of the Royal Society A*, available at: <https://doi.org/10.1098/rsta.2018.0083>.

⁵¹ Article 29 WP, *supra* note 47, p. 8.

⁵² G. Sartor, ‘The impact of the General Data Protection Regulation (GDPR) on artificial intelligence’ (2020) European Parliamentary Research Services (EPRS), Panel for the Future of Science and Technology, available at: [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU\(2020\)641530_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf).

of DLADM without due consideration for data protection principles risks to infringe fundamental human rights, including human dignity,⁵³ the right to privacy,⁵⁴ protection of personal data,⁵⁵ and non-discrimination.⁵⁶

In terms of scope, the GDPR's ADM framework applies under a set of conditions. First of all, Article 22 strictly governs ADM processing related to 'personal data', that is, information which relates to an identifiable natural and living person and which can be used to directly or indirectly identify them.⁵⁷ In practice, this definition has a very broad remit,⁵⁸ and can thus extend to cases in which conjoining various items of information allows for identification (known as the phenomenon of 'unique combinations').⁵⁹ Secondly, the GDPR extends to activities that (i) take place within the context of an establishment of a controller or a processor in the EU, regardless of where the processing itself takes place.⁶⁰ 'Establishment', here, will be treated as a flexible concept, covering any real and effective activity of a legal entity, even if minimal.⁶¹ The GDPR also governs ADM processing activities when they relate to (ii) the offering of goods or services to data subjects in the EU, or alternatively (iii) the monitoring of their behavior.⁶² This regime, quite evidently, casts a wide regulatory reach.

Finally, the prohibition on solely automated decisions will not apply when processing qualifies as "necessary" for entry into, or performance of, a subject-controller contract, or when the processing is otherwise based on explicit consent.⁶³ In both these cases, responsibility falls on the controller to implement "suitable measures to safeguard the data subject's rights and freedoms and legitimate interests".⁶⁴ Article 22 will also not apply when the ADM processing is legally authorized, provided that the law in question provides for the same "suitable measures".⁶⁵ Therefore, even when exempting the prohibition on ADM, measures must nonetheless be in place to adequately account for the rights, freedoms, and legitimate interests of data subjects. These measures must "at least" account for the data subject's "right to obtain human intervention", to "express his or her point of view", and to "challenge the [automated] decision".⁶⁶

Having outlined the GDPR's restrictions on, and obligations for, ADM activities, this paper will turn to the right to an explanation and its definition under EU data protection law.

⁵³ Charter of Fundamental Rights of the European Union (2000/C 364/01), Article 1.

⁵⁴ *Ibid*, Article 7.

⁵⁵ *Ibid*, Article 8.

⁵⁶ *Ibid*, Article 21.

⁵⁷ GDPR, Article 4(1).

⁵⁸ CJEU, *Breyer v Germany* (C-582/14); CJEU, *Nowak v Data Protection Commissioner* (C-434/16); CJEU, *Single Resolution Board v EDPS* (T-557/20).

⁵⁹ D. Kamarinou et al., 'Protection of Personal Data in Clouds and Rights of Individuals' (2021) in: C. Millard (ed.) *Cloud Computing Law*, (2nd edn, OUP 2021), available at: <http://dx.doi.org/10.2139/ssrn.4255833>.

⁶⁰ CJEU, *Google Spain v AEPD* (C-131/12).

⁶¹ CJEU, *Weltimmo* (C-230/14).

⁶² GDPR, Article 3(1)-(2).

⁶³ GDPR, Article 22(2)(a) and (c).

⁶⁴ *Ibid*, Article 22(3).

⁶⁵ *Ibid*, Article 22(2), subparagraph (b).

⁶⁶ *Ibid*, Article 22(3) and Recital 71.

3.2. The Right to an Explanation for Automated Decisions

Albeit treated as a highly contentious topic in the literature over the years, the right to an explanation has evolved into a well-developed requirement, with the most recent contribution stemming from the CJEU. This section will demonstrate that explainability exists as a standalone, binding and enforceable right under the EU data protection law, first by highlighting GDPR provisions, and then considering CJEU jurisprudence.

3.2.1. *Sticking to the Letter of the Law: GDPR Provisions*

Although the ‘right to an explanation’ is not explicitly articulated under the GDPR, the cumulative effect of several provisions nonetheless promotes a *de facto* right. This combination includes Article 22, on automated decision-making, Articles 13 and 14, on the disclosure of information to the data subject, and Article 15, on the right of access. Under this framework, the “existence of automated decision-making” requires controllers to provide “meaningful information” regarding (i) “the logic involved”, (ii) “the significance”, and (iii) “the envisaged consequences of such processing for the data subject”.⁶⁷ Moreover, all information regarding an automated decision must be provided “in a concise, transparent, intelligible and easily accessible form, using clear and plain language”.⁶⁸

By taking a comprehensive view on the legal constellation of Articles 22, 13, 14 and 15, it becomes evident that a distinct ‘right to an explanation’ exists under the GDPR. This position is strengthened on account of the fundamental rights connected to Article 22, such as the right to an effective remedy and non-discrimination. Moreover, each act of EU law must be interpreted in accordance with its respective context, objectives and purpose,⁶⁹ hence requiring any legal interpretation of the GDPR to be grounded on its objective to protect the fundamental rights and freedoms of natural persons, “in particular their right to the protection of personal data”.⁷⁰ Its provisions should also be read in light of the broader technological context within which the act was adopted, namely the rapid, mass scale adoption of black box DL systems at all levels of commercial and organization decision-making.⁷¹ Taking all of these factors into account, it appears more than fitting to opt for the recognition of a right to an explanation.

Conversely, some scholars assume a more restrictive outlook on this issue, arguing that explainability does not exist in the GDPR as such, and if anything takes the form of a ‘right to be informed’.⁷² Similarly, others propose conserving the language of the GDPR, opting to keep the

⁶⁷ GDPR, Articles 13(2)(f), 14(2)(g), 15(1)(h).

⁶⁸ *Ibid*, Article 12(1).

⁶⁹ Fresz et al., *supra* note 8, p. 443.

⁷⁰ GDPR, Article 1(2).

⁷¹ As discussed in Section 2.

⁷² S. Wachter et al., ‘Why a right to explanation of automated decision-making does not exist in the general data protection regulation’ (2017) 7(2) *International Data Privacy Law*, pp. 76–99, available at: <http://dx.doi.org/10.2139/ssrn.2903469>.

requirement of ‘meaningful information about the logic involved’ rather than extrapolating a new right.⁷³ However, these opinions insufficiently engage with the wording under Recital 71 of the GDPR, which specifically stipulates the right “to *obtain an explanation* of the [automated] decision” (emphasis added) as part of the “suitable safeguards” controllers should implement when processing personal data for ADM purposes. Although recitals are not legally binding as such, they nonetheless significantly influence the interpretation of the GDPR.⁷⁴ Recital 71 therefore effectively colors the collective reading of Articles 13, 14, 15 and 22,⁷⁵ in such a way so as to result in an independent right to an explanation.

Explainability also plays an essential role in enabling the exercise of data subject rights, especially those tied to ADM, including human intervention, expression of opinion, and contestability. Explanations enable affected individuals to exercise their rights by challenging, appealing, or simply investigating automated choices, thereby “transforming explanation from passive transparency into active contestability”.⁷⁶ In support of this, the European Data Protection Board (‘EDPB’) maintains that “the data subject will only be able to challenge a decision or express their view if they fully understand how it has been made and on what basis”.⁷⁷ Interestingly, this description by the EDPB appears to stipulate the need for both global explanations, for understanding *how* a decision has been made, as well as local explanations, for identifying *why* a decision was taken and on what grounds.

On this backdrop, the CJEU confirmed the legal and enforceable status of a right to an explanation for ADM under the GDPR framework. The subsequent discussion will therefore turn to CJEU caselaw, in particular the 2023 *SCHUFA Holding* Opinion and the 2025 *Dun & Bradstreet Austria* Decision, to analyze more narrowly how explainability is applied in line with EU data protection law.

3.2.2. Jurisprudence of the Court of Justice

The principle of explainability is not foreign to the CJEU. In fact, the Court has ruled on the legal requirement of local and global explanations in the context of terrorism-detection algorithms.⁷⁸ According to CJEU caselaw, local *ex ante* explainability (before processing) is a fundamental right

⁷³ A. D. Selbst and J. Powles, ‘Meaningful information and the right to explanation’ (2017) 7(4) *International Data Privacy Law*, pp. 233–242, available at: <https://doi.org/10.1093/idpl/ipx022>; J. Dexe et al., ‘Explaining automated decision-making: a multinational study of the GDPR right to meaningful information’ (2022) 47 *The Geneva Papers on Risk and Insurance - Issues Practice*, pp. 669–697, available at: <https://doi.org/10.1057/s41288-022-00271-9>.

⁷⁴ Thompson Reuters, ‘Recital (EU)’ *Practical Law*, available at:

<https://uk.practicallaw.thomsonreuters.com/Glossary/PracticalLaw/113f404e5785211e79bef99c0ee06c731>.

⁷⁵ See: G. Malgieri and G. Comandé, ‘Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation’ (2017) 7(4) *International Data Privacy Law*, pp. 243–265, available at: <https://doi.org/10.1093/idpl/ipx019>.

⁷⁶ A. Engelfriet, ‘An Uninterpretable Right: Legal and Practical Limits of the Right to an Explanation’ (2025), p. 5, available at: <http://dx.doi.org/10.2139/ssrn.5312780>.

⁷⁷ EDPB, ‘Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679’ (2018) available at: https://edpb.europa.eu/our-work-tools/our-documents/guideline/automated-decision-making-and-profiling_en.

⁷⁸ CJEU, *La Quadrature du Net*, Joined cases C-511/18, C-512/18 and C-520/18 ; CJEU, *Ligue des droits humains v. Council of Ministers*, C-817/19.

under the EU Charter when a decision leads to serious consequences for the rights and freedoms of individuals, while local *ex post* contestability (after processing) is protected as a fundamental right when an individual's access to effective remedy would be compromised without it.⁷⁹

Within the context of data protection law, explainability has been specifically addressed by the Court in the context of profiling for the purpose of credit scoring. In his 2023 Opinion on the *OQ v Land Hesse SCHUFA Holding* case, Advocate General Pikamäe found that 'meaningful information about the logic involved' under Article 15 GDPR requires the controller to provide "sufficiently detailed explanations" of both the method used to reach the decision as well as reasons for the specific outcome.⁸⁰ Therefore, according to the AG, controllers should deliver detailed local *ex post* explanations to justify individual decisions, and global explanations to provide "general information" regarding what features the ADM process draws from and their "respective weight on an aggregate level".⁸¹ In fact, for AG Pikamäe, global explainability exists as a standard obligation on controllers, and necessary to enable the contestability of automated decisions.⁸²

The AG was not alone in his ideas, as the CJEU subsequently reached the same conclusion. In its groundbreaking decision for the *Dun & Bradstreet Austria* case, the Court cements the right to an explanation for ADM. In reference to both Recital 71 and the 2023 *SCHUFA* Opinion, the CJEU found that Article 15(1)(h) of the GDPR affords data subjects "a genuine right to an explanation as to the functioning of the mechanism involved in automated decision-making of which that person was the subject and of the result of that decision".⁸³ Hence, according to the Court, the right to an explanation exists as a *genuine*, enforceable legal obligation under the GDPR. Furthermore, explainability is a two-fold requirement, prescribing both a global explanation, on the 'functioning of the mechanism', and a local one, on 'the result' of the ADM. Whether these are to be presented separately or together in the same explanation is not made clear in the judgment, but both explanations nonetheless merit due attention from data controllers.

In addition, the CJEU (again, echoing Pikamäe) finds contestability to be a major goal of the right to an explanation.⁸⁴ However, the Court goes one step further and acknowledges that explainability, as part of the broader right of access, not only safeguards the rights of data subjects under Article 22(3), but also *all* other data subject rights under the GDPR, including the right to be forgotten under Article 17, the right to restriction of processing under Article 18, the right to object as per Article 21, and the right to remedy as well as compensation under Articles 79 and 82, respectively.⁸⁵ By taking

⁷⁹ Maxwell and Dumas, *supra* note 45, p. 19.

⁸⁰ Advocate General Pikamäe, Opinion in *SCHUFA* (C-634/21), paragraph 58.

⁸¹ *Ibid.*

⁸² *Ibid.*

⁸³ CJEU, *Dun & Bradstreet Austria* (C-203/22), paragraph 57.

⁸⁴ *Ibid.*, paragraph 56.

⁸⁵ *Ibid.*, paragraph 54.

this progressive position, the CJEU frames explainability as a fundamental safeguard for the rights of data subjects at the receiving end of any ADM with legal or similarly significant effects.

While the right to an explanation was only recently crystallized under EU data protection law, the principle of explainability holds long-standing roots in this field. Since the GDPR's entry into force, EU regulatory bodies have consistently invoked explainability in relation to the development of GDPR-compliant AI systems. Overtime, this standard became a well-established concept, serving both as a technical measure as well as a guiding tenet of Data Protection by Design and by Default, as will be evidenced in the following section.

3.3. Explainability-by-Design: Regulatory Guidance

Analysis of European soft law demonstrates that explainability has been progressively framed by EU regulatory bodies as a core technical safeguard within the GDPR's DPDD framework. Since the GDPR's entry into force, bodies such as the Commission's High-Level Expert Group on Artificial Intelligence ('AI-HLEG'), the EDPS, the Commission's Joint Research Center, and the European Parliamentary Research Service ('EPRS') have consistently positioned explainability as both a key obligation and a guiding development principle data controllers must implement to ensure compliance with the data protection principles.

The AI-HLEG, for instance, identifies 'explicability' as one of the four fundamental principles for Trustworthy AI,⁸⁶ arguing that data subjects directly or indirectly affected by automated decisions have an explicit right to both global and local explanations, without which they could not contest decisions.⁸⁷ When local explanations cannot be actualized, such as in the case of black box AI, the Expert Group suggests turning to alternative "explicability measures", including traceability, auditability, and transparency, as a means of ensuring the respect for fundamental rights by "the system as a whole" – or *by design and by default*.⁸⁸

Explainability, specifically, exists as a more practical requirement which concerns "the ability to explain both the technical processes of an AI system and the related human decisions" as a means of ensuring GDPR compliance.⁸⁹ Substantiating explainability involves, once again, considering to what extent interpretability was accounted for and embedded into the AI system *by design*.⁹⁰ XAI is thereby classified as a 'technical method' for ensuring trustworthiness-by-design, considered to be "vital" not only for users to comprehend ADM, but also for deployers to measure the reliability of their

⁸⁶ High-Level Expert Group on Artificial Intelligence, 'Ethics Guidelines for Trustworthy AI' (2019), European Commission, pp. 11-13, available at: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

⁸⁷ *Ibid*, p. 13.

⁸⁸ *Ibid*.

⁸⁹ *Ibid*, p. 18.

⁹⁰ *Ibid*, p. 29.

technology.⁹¹ The EDPS reaches an equivalent conclusion, advising controllers to treat XAI as a ‘*technical* measure’ through which they may demonstrate compliance with data protection principles.⁹² This recommendation parallels the text of Recital 71, where explanations are recognized as “suitable safeguards” to satisfy the demands of Article 22(3).

Aside from being considered as an essential technical measure, explainability has additionally been framed as a guiding principle for proactive GDPR-compliance, in accordance with the DPDD requirement. On the basis of Articles 13 and 14 GDPR,⁹³ the EPRS recognizes explainability as a guiding principle for “the development of AI”, and particularly opaque DL-based models.⁹⁴ Likewise, developers and researchers of complex AI are encouraged by the Commission’s Joint Research Center to invest in the “development principle of explainability-by-design”, with the goal of meeting EU data protection standards.⁹⁵ It is worth noting that, as underlined by the EDPS, XAI will not automatically lead to ‘GDPR compliance-by-design’ – instead, it should be understood as a crucial technical consideration and a guiding principle for providers who seek to develop trustworthy products in line with data protection standards.⁹⁶

Taking all of the foregoing regulatory guidance into consideration, the function, objectives, and characteristics representative of explainability coincide fundamentally with the DPDD principle of Article 25 GDPR. Therefore, the following section will address this paper’s second research question, exploring to what extent the right to an explanation may be modelled and effectively operationalized in accordance with the DPDD principle.

4. Modelling GDPR-Compliant XAI According to Article 25

It has been argued that the entire weight of the GDPR rests on the shoulders of Article 25.⁹⁷ That may be true, considering that the DPDD rule turns data protection from an afterthought into a critical obligation present from the very conception of the product, and it does so in large part by requiring controllers to implement appropriate technical and organizational measures (TOMs).⁹⁸ Crucially, Article 25 prescribes a positive obligation to act, thus non-compliance with this provision can result in fines of up to €20 million, or 4% of the total global annual turnover of the controller from the preceding year.⁹⁹

⁹¹ *Ibid*, pp. 21-22.

⁹² EDPS, *supra* note 33, p. 11.

⁹³ Discussed in Section 3.2.1.

⁹⁴ Sartor, *supra* note 58, p. 54.

⁹⁵ Hamon et al., *supra* note 6, p. 84.

⁹⁶ EDPS, *supra* note 33.

⁹⁷ Michelakaki and Vale, *supra* note 11, p. 2.

⁹⁸ For reasons of scope, the ‘data minimization by default’ aspect of Article 25 will be excluded from the following analysis.

⁹⁹ *Ibid*, Article 83.

As established in Section 3.3, explainability has been decisively recognized by EU regulatory bodies as a recommended technical measure and guiding principle for demonstrating compliance with the data protection principles. Sections 3.1 and 3.2 likewise illustrated the role of explainability in safeguarding data subject rights and fundamental human rights. Having regard to the foregoing, it is clear that explainability qualifies as an ‘appropriate’ TOM, in that it is “designed to implement data-protection principles [...] in an effective manner and to integrate the necessary safeguards [...] to meet the requirements of this Regulation and protect the rights of data subjects”.¹⁰⁰

On that note, the following sections will present the best approaches to modelling GDPR-compliant XAI in line with the DPDD principle of Article 25, taking into account the legal constellation of relevant GDPR provisions, CJEU case law, and EU soft law summarized in Section 3. It should be reiterated that the DPDD principle, and the corresponding XAI models proposed in this paper, will apply to all DL systems governed by the GDPR. That includes processing activities that fall within the GDPR’s territorial scope, material scope, and its definition of automated decision-making under Article 22 (refer to Section 3.1).

4.1. Building GDPR-Compliant Explanations for Deep Learning Models

When an automated DL decision falls within the scope of the GDPR, it should be accompanied by an explanation regarding the logic involved, the significance of the decision, and any foreseeable consequences of the respective processing for the data subject concerned. But when should the explanation be provided by the controller? In what format? And how can the information be ‘meaningful’ for the data subject? The following analysis illustrates how the GDPR framework provides a right to two distinct explanations, one before ADM processing and one after, both of which will be modelled accordingly.

4.1.1. Before the Decision: Global Explanations

Based on the provisions of Articles 13 and 14 GDPR, controllers must provide global *ex-ante* explanations, that is, prior to ADM processing. This requirement derives from the data subject’s right to be informed at the moment of data collection, a right which extends to disclosure of the ‘logic involved’ in the ADM process.¹⁰¹ As a result, data subjects should be informed about (i) the existence of DLADM, (ii) the principal characteristics of the system involved, and (iii) the potential outcomes. Alongside these elements, the EPRS has additionally suggested including information regarding the input data that the system takes into consideration, which data items may favor or disfavor the pursued

¹⁰⁰ Ibid, Article 25(1).

¹⁰¹ GDPR, Recital 63.

outcome, the target values or ‘weights’ computed by the system, and the overall purpose(s) the system aims to achieve.¹⁰²

In terms of language, global *ex-ante* explanations should be clear and plain, ensuring that laypeople can understand the information.¹⁰³ At the same time, the simplicity in language should not undermine the precision or the relevance of explanations.¹⁰⁴ To ensure data subject comprehension, controllers should avoid including detailed technical information, for example regarding the algorithm, when the recipient will not be in a position to understand.¹⁰⁵ Secondly, in terms of format, explanations should be made available in a concise, transparent, intelligible, and easily accessible form.¹⁰⁶ One possible approach, as proposed by the EPRS, entails providing modular information, for example in the form of bullet points accompanied by links to access more in-depth descriptions that, for example, cover technical aspects. Another option may involve pairing information, ideally segmented in short paragraphs, with standardized icons in order to give an easily visible, intelligible and digestible overview of the DLADM processing.¹⁰⁷

Some scholars have alternatively suggested the adoption of ‘black box tinkering’, a reserve-engineering technique that allows data subjects to test a system and analyze its responses to detect fallacies or bias.¹⁰⁸ While this approach could democratize knowledge or control over DL algorithms, it would not meet the requirements of a GDPR-compliant *ex-ante* explanation on its own merit, especially when user-facing. Therefore, taking into account all forementioned requirements, global *ex-ante* explanations should resemble the model outlined in Table 1, below, in both content and form.

Table 1: GDPR-Compliant *Ex-Ante* Explanations

| <i>Element</i> | <i>Content</i> | <i>Example</i> |
|-------------------|--|--|
| Timing | Make available a global explanation prior to the ADM system being deployed or when individuals are first subject to the ADM. | “You are receiving this information because our platform uses a profiling system to tailor marketing content according to your preferences.” |
| Purpose and Aim | Explain the role and objective of the DLADM in the overall service or process that it supports. | “This system profiles consumers to deliver personalized marketing content, promotions, and offers.” |
| DL Logic Overview | Provide a simple description of how the DL model works, without too much technical nuance. | “Our deep-learning powered system analyzes customer demographics, browsing history, and past purchases to predict preferences and engagement.” |

¹⁰² Sartor, *supra* note 58, p. 55.

¹⁰³ GDPR, Article 12.

¹⁰⁴ Sartor, *supra* note 58, p. 56.

¹⁰⁵ Advocate General Pikamäe, *supra* note 86, paragraph 76.

¹⁰⁶ *Ibid.*

¹⁰⁷ GDPR, Article 12(7).

¹⁰⁸ See: M. Perel and N. Elkin-Koren, ‘Black Box Tinkering: Beyond Disclosure in Algorithmic Enforcement’ (2017) 69(1) *Florida Law Review*, available at: <https://scholarship.law.ufl.edu/flr/vol69/iss1/5>.

| | | |
|---------------------------|---|--|
| Data Sources | Identify the sources and types of data used by the system, such as the training data. | "We use data from your platform interactions, purchase records, customer surveys, and loyalty program participation." |
| Weights and Target Values | Describe which features influence the decision most and which decision thresholds are in place. | "Recent browsing history and frequency of purchases carry the most weight. Scores above 0.75 trigger targeted ad delivery." |
| Potential Outcomes | Mention the possible decisions and explain what they mean for the individual. | "You may receive personalized discounts, product recommendations, or notifications about relevant sales." |
| Data Subject Rights | Indicate the relevant data subject rights and how these may be exercised by individuals. | "You have a number of rights, including the right to access your personal data, obtain an explanation regarding our profiling, correct inaccuracies, and opt out of personal marketing. Click here to know more about your data protection rights and how to exercise them." |

4.1.2. After the Decision: Local Explanations

The second facet of the right to an explanation entails accounting for potential data subject access requests, or 'explainability requests', following the provision of an automated decision. The most appropriate approach, to guarantee transparency, contestability, and accountability, takes the form of local *ex-post* explanations. In line with the CJEU's *Dun & Bradstreet Austria* judgment, *ex-post* explanations should describe both the procedure as well as principles applied by the model "in such a way that the data subject can understand which of his or her data have been used in what way".¹⁰⁹ Thus, data subjects should understand which specific features, both system-based and derived from the input data, were identified by the algorithm as having contributed most to the ultimate result. Moreover, local *ex-post* explanations should be easily accessible as well as complete.¹¹⁰ In that way, data subjects can effectively assess whether there is "an objectively verifiable consistency and causal link" between the logic involved and the output received,¹¹¹ which in turn enables the exercise of numerous rights awarded to them under Chapter III as well as Articles 79 and 82 of the GDPR.

Content-wise, a fundamental requirement when it comes to local explainability revolves around ensuring the *meaningfulness* of information. Meaningful explanations must, in a concise, reasonably comprehensible, and intelligible manner, communicate the purpose of the DL system and the principles involved in the respective procedure, as well as delineate which factors led to the automated decision. Just as for global explanations, *ex-post* descriptions with overly technical or mathematical information

¹⁰⁹ CJEU, *Dun and Bradstreet Austria*, paragraph 61.

¹¹⁰ Advocate General Richard de la Tour, Opinion in *Dun & Bradstreet Austria* (C-203/22), paragraph 75.

¹¹¹ *Ibid*, paragraph 71.

would not meet the requirements of meaningfulness.¹¹² It is also crucial that they underline which rights data subjects may exercise with respect to the decision, and how.

The first step in choosing appropriate explanation methods involves considering the contemporary progress in XAI technology available on the market, or as phrased in Article 25(1), the ‘state of the art’. Currently, as of August 2025, the state of the art consists of an array of options, a majority of which may be classified as ‘feature-attribution’ XAI.¹¹³ Such methods produce a value to represent how relevant a specific feature is for the ADM outcome, meaning the higher the value assigned to a feature, the higher the relevance of that data item for the automated decision concerned. A similar, and potentially simpler, approach resides with contrastive explanations.¹¹⁴ These justify a specific decision by contrasting it with other possible, more favorable, outcomes as a way of underlining what features influenced the decision-making process, i.e., illustrating the causal link of the ADM processing. Counterfactual scenarios may be employed to enhance contrastive explanations, by presenting hypotheticals of specific variable changes that would have resulted in a different decision.

It is relevant to note that, from a data protection perspective, contrastive explanations may prove to be the favored XAI method. In *Dun & Bradstreet Austria*, the CJEU identifies contrastive explanations as a particularly “appropriate” approach of satisfying the right to an explanation.¹¹⁵ Interestingly, the Court’s judgment is supported by empirical findings. Research by the EDPS shows users tend to prefer contrastive explanations over other forms of local explanations,¹¹⁶ and focus studies with data protection experts demonstrate an identical trend.¹¹⁷ Nevertheless, determining which form of explanation is the most “appropriate” under Article 25 will depend on the cost of implementation, the nature, scope, context and purposes of processing, as well as the risks of varying likelihood and severity for data subjects’ rights and freedoms.¹¹⁸ All of these factors should therefore be taken into consideration on a case-by-case basis by controllers, an assessment which begins from the very beginning of the development process.

It is important to recognize, as highlighted by the EDPB, the Article 29 Working Party, and the CJEU, that the complexity of DL operations does not relieve controllers of their duty to provide

¹¹² CJEU, *Dun and Bradstreet Austria*, paragraph 59.

¹¹³ Notable examples include Shapley Additive Explanations (SHAP), Local Interpretable Model-Agnostic Explanations (LIME), Deep Learning Important Features (DeepLIFT), and Integrated Gradients.

¹¹⁴ Key examples of contrastive state-of-the-art methods include Diverse Counterfactual Explanations (DICE) and Local Rule-Based Explanations (LORE).

¹¹⁵ CJEU, Press Release No 22/25 (2025), available at:

<https://curia.europa.eu/jcms/upload/docs/application/pdf/2025-02/cp250022en.pdf>.

¹¹⁶ EDPS, *supra* note 33, p. 14.

¹¹⁷ L. State et al., ‘The explanation dialogues: an expert focus study to understand requirements towards explanations within the GDPR’ (2025) *Artificial Intelligence and Law*, p. 34, available at:

<https://doi.org/10.1007/s10506-024-09430-w>.

¹¹⁸ GDPR, Article 25(1).

explanations.¹¹⁹ Appealing to ‘technical impossibility’ would also directly conflict with Article 25(1), as it demands the implementation of DPDD at both the moment that controllers determine the means of processing as well as the moment of processing itself. With all of this in mind, Table 2 (below) combines all key requirements for GDPR-compliant local *ex-post* explanations under Article 25:

Table 2: GDPR-Compliant *Ex-Post* Explanations

| <i>Element</i> | <i>Content</i> | <i>Example</i> |
|---|--|---|
| Timing | Local explanations should be given upon a data subject request for access or explanation following the provision of an automated decision. | “You are receiving this explanation because you requested further information on your profile” / “you asked why you were shown this particular promotion”. |
| Purpose and Aim | Explain the role and objective of the DLADM in the overall service or process that it supports. | “This system profiles consumers to deliver personalized marketing content, promotions, and offers.” |
| DL Logic Overview | Provide a simple description of how the DL model works, without too much technical nuance. | “Our deep-learning powered system analyzes customer demographics, browsing history, and past purchases to predict preferences and engagement.” |
| Automated Decision | Clearly state the decision taken and its implications for the data subject. | “You were profiled into the ‘active lifestyle’ segment. You are therefore targeted with offers for fitness clothing.” |
| Relevant Features (Feature Attribution) | Identify which specific data features, whether system-based or derived from input data, most contributed to the ADM processing outcome. | “Why you got this result: our system picked up on your repeated visits to the fitness section, your last purchase of running shoes, and your loyalty program engagement.” |
| Contrastive Insights (Counterfactuals) | Explain the decision by contrasting it with a more favorable outcome and, if appropriate, provide ‘what-if’ scenarios to illustrate the causal link. | “You received a 10% discount on sportswear based on your recent browsing and purchase activity. Had your profile fit in another segment such as home enthusiast you would have received a 10% promotion for kitchenware instead.” |
| Data Subject Rights | Indicate the relevant data subject rights and how these may be exercised by individuals. | “You have a number of rights, including the right to correct your data or opt out of profiling for marketing purposes. Find out more about your data protection rights and how to exercise them here: [link].” |
| Point of contact | Provide details on who the data subject can contact for further questions, requests, or complaints. | “For any further questions on this decision or on your rights, please contact our Data Protection Offer at dpo@company.com or call +33 123 123 123.” |

¹¹⁹ Article 29 WP, *supra* note 47, p. 25; CJEU, *Dun and Bradstreet Austria*, paragraph 61; EDPB, ‘Report of the work undertaken by the ChatGPT Taskforce’ (2024), paragraph 7, available at: https://www.edpb.europa.eu/system/files/2024-05/edpb_20240523_report_chatgpt_taskforce_en.pdf.

As a side note, both global *ex-ante* and local *ex-post* explanations would benefit from including an open ‘user evaluation’ tab in order to test the comprehensibility, accessibility, and overall meaningfulness of the explanations provided. That way, each data subject has the option to express his or her opinion, and controllers monitor the effectiveness of the XAI component.¹²⁰

5. Limitations of Explainability Under Data Protection Law

5.1. Legal Restrictions

The right to an explanation may, at times, conflict with the rights and legitimate interests of others, including controllers. For that reason, the GDPR’s framework delineates some restrictions to the explainability requirement, starting with the content of the explanation. The necessary degree of detail in an explanation may depend on whether the automated decision is likely to have adverse effects on data subjects. The AI-HLEG put forth a similar idea, stating that the degree to which explicability is necessary in the first place is “highly dependent on the context and the severity of the consequences if [the automated decision] is erroneous or otherwise inaccurate”.¹²¹ Likewise, under Article 25, controllers are allotted substantial discretion as to the determination of ‘appropriateness’ of TOMs, which should be assessed with respect to any identified risk(s).¹²² Following this logic, the quality and quantity of content in explanations may be limited by controllers in light of the potential impact of automated decisions (or lack thereof).

The provision of explanations may also be ‘delayed’ when the request in question relates to a highly complex ADM process, in which case controllers will be given extra time to respond.¹²³ Additionally, if they can prove that a particular explanation request is manifestly unfounded or excessive, controllers may refuse to comply with that request, or charge a reasonable fee to cover costs of compliance.¹²⁴ The broadest ground for legal restrictions to the right to an explanation exists under Article 23 GDPR, which stipulates that Union or Member State law can restrict the rights under Articles 13, 14, 15 and 22 when necessary and proportionate for safeguarding particular goals such as national security, or the rights and freedoms of others.

With regard to the rights of others, controllers may leave out or render illegible some requested information when full disclosure would compromise, for example, the data protection or intellectual

¹²⁰ Nevertheless, the evaluation of XAI is in itself a widely recognized challenge of explainability. For more, see: S. Brdnic et al., ‘A Comprehensive User-Centric Method for Evaluating and Comparing XAI Explanations’ (2025) *International Journal of Human–Computer Interaction*, pp. 1-20, available at: <https://doi.org/10.1080/10447318.2025.2470278>.

¹²¹ AI-HLEG, *supra* note 90, p. 13.

¹²² L. Jasmontaite et al., ‘Data Protection by Design and by Default: Framing Guiding Principles into Legal Obligations in the GDPR’ (2018) 4(2) *European Data Protection Law Review*, p. 6, available at: <https://doi.org/10.21552/edpl/2018/2/7>.

¹²³ GDPR Article 12(3).

¹²⁴ *Ibid*, Article 12(5).

property rights of third parties.¹²⁵ In the context of AI explainability, the most commonly invoked restriction relates to the controller's interest in the protection of trade secrets, for example over the algorithm or the training data.¹²⁶ If a controller takes the view that requested information on the 'logic involved' contains trade secrets, they rely on a competent supervisory authority or court to balance the rights and interests at issue and ultimately adjudicate on the data subject's the right to an explanation.¹²⁷ Controllers will, however, most likely never be allowed to outright reject explanation requests, as that would violate the necessity and proportionality requirements.¹²⁸ Here, a crucial question arises as to whether data protection authorities and courts will have sufficient capacity to deal with a spike in explanation requests, a topic which merits due attention from the entire legal domain.

5.2. Practical Restrictions

As with any technical measure, explainability has practical limitations. Firstly, and quite obviously, XAI does not come for free, not in terms of time, nor resources. Maintaining explainability standards may be particularly challenging as it relates to the 'organizational' aspect of Article 25, as it may entail considerable efforts, for example, in terms of explainability audits or collaborative research and development frameworks. In terms of documentation, explainability also adds to the pile of internal workload. In view of the accountability principle, controllers will be expected to document all their design choices from day one, including why particular XAI methods were chosen, clearly mapping each explanation in connection with data subject rights, and doing so potentially as part of Data Protection Impact Assessments.¹²⁹

Nevertheless, said costs need to be balanced against the benefits derived from explainability. For high-risk applications, it might well be worth it to invest in a robust XAI component. It may be equally valuable to involve not only competent people with multidisciplinary backgrounds, but also stakeholders that are directly or indirectly affected by the ADM system in question.¹³⁰ That may necessitate outsourcing experts, practitioners, and specialists from the earliest stages of product development, and on a continuous basis overtime.

An even more pressing question relates to all the AI products already deployed in the market which have not accounted for explainability by-design and by-default, and who perhaps cannot fully implement explainability as such. On one hand, the guidelines and jurisprudence of the EU take a clear stance on this issue, declaring that no excuse may be invoked by controllers to evade the explainability

¹²⁵ Ibid, Article 15(4).

¹²⁶ EU Charter of Fundamental Rights, Articles 16 and 17(2).

¹²⁷ CJEU, *Dun and Bradstreet Austria* (C-203/22), paragraph 74; CJEU, *Norra Stockholm Bygg* (C-268/21), paragraph 58.

¹²⁸ Amsterdam Court of Appeal, *Uber / Drivers* (2023) ECLI:NL:GHAMS:2023:793; ECLI:NL:GHAMS:2023:796.

¹²⁹ GDPR, Article 35.

¹³⁰ O. G. Yalcin, 'Governing Algorithms in the Big Data Era for Balancing New Digital Rights' (2023) Doctor of Philosophy Thesis in Law, Science and Technology, p. 35.

requirement. On the other hand, what about cases where it *really is* technically impossible to provide local explanations, especially for deep learning black box systems? This is an extremely relevant question given that the vast majority of current AI developers and providers have not fully operationalized the explainability-by-design principle.¹³¹ What should be further considered is, when explainability by-design and by-default is not directly possible, would it suffice to implement the AI-HLEG’s alternative ‘explicability measures’ of traceability, auditability, and transparency? No European guidelines, opinions, or court decisions have yet addressed these questions, making the confirmation of the right to an explanation a source of regulatory disruption.

The third limitation relates to the content of local explanations, in particular the fact that these must be adapted to the expertise of the recipient concerned.¹³² The ‘contextualization’, or ‘tailoring’, of explanations may prove challenging in practice considering the many, distinct groups of stakeholders involved, including (i) developers, researchers, and designers, (ii) company owners, (iii) regulators, (iv) domain experts, and (v) end-users, or society at large.¹³³ Several suggestions have been raised in the literature regarding how to best contextualize explanations, including the creation of cloud-based user interfaces,¹³⁴ or interactive digital spaces allowing data subjects to build customized explanations based on their particular background, needs, and interests.¹³⁵ XAI contextualization may be otherwise achieved through the inclusion of different stakeholders during the entire development process, as proposed above. This would nonetheless require significant planning, testing, and updating in response to user feedback. Another possible solution may involve preparing ‘shallower’ global and local explanations for set groups of audiences, and integrating a human component, such as a contact point, for data subjects interested in receiving further information.

Finally, the individual empowerment provided by the right to an explanation may, in practice, function as a subtle transfer of regulatory and interpretive burden onto the data subject. Explainability, in effect, presumes a high degree of digital literacy, legal knowledge, and cognitive, as well as emotional, intelligence from people, who are asked to not only interpret the explanation, but also monitor for errors, understand legal objections, and explore complex redress procedures.¹³⁶ This raises an issue of “transparency fallacy”, that is, the risk that meaningful explanations may not be effectively

¹³¹ Rossetti, *supra* note 9: “We wonder what should be done when the processing is so intricate that it cannot be explained in an understandable manner. This is not a theoretical scenario given the inherent complexity of certain AI systems. Could we reasonably conclude that if the processing is not explainable, it should be simplified or even interrupted? The implications of this assumption may have far-reaching consequences.”

¹³² AI-HLEG, *supra* note 90, p. 18: “Explanation[s] should be timely and adapted to the expertise of the stakeholder concerned”.

¹³³ M. Bello et al., ‘Explainability in Context: A Multilevel Framework Aligning AI Explanations with Stakeholder with LLMs’ (2025) pp. 5-6, available at: <https://arxiv.org/pdf/2506.05887>.

¹³⁴ See: Yalcin, *supra* note 133.

¹³⁵ See: Sovrano et al., *supra* note 6.

¹³⁶ Engelfriet, *supra* note 80, p. 4.

understood by data subjects due to their limited attention, interests, or cognitive capabilities.¹³⁷ Such a problem is exacerbated by the existence of ‘automation bias’, the tendency to regard AI-generated outputs as more reliable than human judgment, resulting in a general scarcity of proper algorithmic scrutiny by users of DLADM models. In fact, research has shown that explanations can increase the chances of humans blindly, and wrongly, following algorithmic recommendations.¹³⁸

As such, over-reliance on and blind trust of ADM technology may limit the true objectives of the right to an explanation, being accountability, fairness, and the protection of fundamental rights. To counteract this, controllers should consider providing a disclaimer stating the potential inaccuracy of the system’s automated decisions, as a means of heightening caution and reducing automation bias.

6. Conclusion

This paper set out to answer two interconnected research questions: (1) To what extent do the provisions of the GDPR together with the CJEU’s *Dun & Bradstreet Austria* judgment in provide for a legally binding and enforceable right to an explanation of automated decision-making, and (2) how can the right to an explanation be modelled in accordance with the requirements of Article 25? As for the first question, this research definitively establishes that the right to an explanation exists as a legally enforceable right under the GDPR, and one that crucially safeguards data subject rights in the context of deep learning automated decision-making. More specifically, through the systematic analysis of Articles 13, 14, 15, and 22, reinforced by Recital 71 and jurisprudence of the CJEU, this dissertation confirms explainability as a binding obligation on controllers, and a dual one at that. Namely, controllers must provide for global explanations on the system’s functioning as well as local explanations to justify individual automated decisions.

Regarding the second question, this research provides a novel perspective on the role of explainability under the Data Protection by Design and by Default framework. The right to an explanation is operationalized as an ‘appropriate technical measure’ under Article 25, and modelled accordingly in two forms: global *ex-ante* and local *ex-post* explainability. These proposed models merge the GDPR’s requirements with state-of-the-art standards, effectively providing a valuable guidance for the implementation of meaningful and user-accessible XAI. The latter part of this dissertation reveals significant challenges ahead across legal, technical, and practical dimensions. The various legal restrictions will shape the treatment of explanation requests, guided by a balancing act between conflicting rights and interests, meanwhile key practical restrictions consist of implementation costs,

¹³⁷ G. Malgieri, ‘“Just” Algorithms: Justification (Beyond Explanation) of Automated Decisions Under the General Data Protection Regulation’ (2021) 1(1) *Law and Business*, Sciendo, pp. 16-28, available at: <https://doi.org/10.2478/law-2021-0003>.

¹³⁸ Maxwell and Dumas, *supra* note 43, p. 31.

the burden of contextualization, the paradoxical fallacies of explainability, and the regulatory uncertainty created by the demands of explainability-by-design.

Future research must therefore examine industry-wide enforcement mechanisms, the capacity of supervisory authorities to handle increases in explanation requests, and whether alternative ‘explicability’ measures may substitute explainability-by-design when technical implementation proves impossible, especially for DL systems. While the CJEU does not shy away from a more progressive regime on ADM, will – or can – the demands of explainability translate into practice? Will our AI-dependent ecosystem bend to accommodate trustworthiness-by-design, or rather surrender data protection to technological inevitability?

Bibliography

Advocate General Pikamäe, Opinion of Advocate General in Case C-634/21, *OQ v Land Hesse* (joined party: SCHUFA Holding AG), delivered on 16 March 2023. ECLI:EU:C:2023:220.

Advocate General Richard de la Tour, Opinion of Advocate General in Case C-203/22, *CK v Dun & Bradstreet Austria GmbH*, delivered on 12 September 2024, ECLI:EU:C:2024:745.

Ali, Sajid, Tamer Abuhmed, Shaker El-Sappagh, Khan Muhammada, Jose M. Alonso-Moral, Roberto Confalonieri, Riccardo Guidotti, Javier Del Ser, Natalia Díaz-Rodríguez, Francisco Herrera, 'Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence' (2023) 99 *Information Fusion*, available at: <https://doi.org/10.1016/j.inffus.2023.101805>.

Alnuaimi, Amer F.A.H. and Tasnim H.K. Albaldawi, 'An overview of machine learning classification techniques' (2024) 97 *BIO Web of Conferences*, available at: <https://doi.org/10.1051/bioconf/20249700133>.

Ammanath, Beena and Kay Firth-Butterfield, 'How deep learning can improve productivity and boost business' (2022) *World Economic Forum*, available at: <https://www.weforum.org/stories/2022/01/deep-learning-business-productivity-revenue/>.

Amsterdam Court of Appeal, *Uber / Drivers* [2023] ECLI:NL:GHAMS:2023:793.

Amsterdam Court of Appeal, *Uber / Drivers* [2023] ECLI:NL:GHAMS:2023:796.

Anjos, Lucas, 'Can Law Ever Be Code? Beyond Google's Algorithmic Black Box and Towards a Right to Explanation' (2021) PhD Thesis, Degree in Sciences juridiques, available at: <https://repositorio.ufmg.br/handle/1843/38861>.

Article 29 Data Protection Working Party, 'Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679' (2017), available at: <https://ec.europa.eu/newsroom/article29/items/612053/en>.

Asghari, Hadi, Nadine Birner, Aljoscha Burchardt, Daniela Dicks, Judith Faßbender, Nils Feldhus, Freya Hewett, Vincent Hofmann, Matthias C. Kettemann, Wolfgang Schulz, Judith Simon, Jakob Stolberg-Larsen, Theresa Züger, 'What to explain when explaining is difficult? An interdisciplinary primer on XAI and meaningful information in automated decision-making' (2021) *Alexander von Humboldt Institute for Internet and Society*, available at: <https://doi.org/10.5281/zenodo.6375784>.

Binoletto, Giorgia, 'EDPB Guidelines 4/2019 on Data Protection by Design and by Default' (2020) 6 *Eur Data Prot L Rev*, pp. 574-579.

Bello, Marilyn, Rafael Bello, Maria-Matilde García, Ann Nowé, Iván Sevillano-García, Francisco Herrera, 'Explainability in Context: A Multilevel Framework Aligning AI Explanations with Stakeholder with LLMs' (2025), available at: <https://arxiv.org/pdf/2506.05887>.

Brdnik, Saša, Sašo Karakatič, Boštjan Šumak, 'A Comprehensive User-Centric Method for Evaluating and Comparing XAI Explanations' (2025) *International Journal of Human-Computer Interaction*, pp. 1-20, available at: <https://doi.org/10.1080/10447318.2025.2470278>.

Brożek, Bartosz, Michał Furman, Marek Jakubiec, Bartłomiej Kucharzyk 'The black box problem revisited. Real and imaginary challenges for automated legal decision making' (2024) 32 *Artificial Intelligence and Law*, available at: <https://doi.org/10.1007/s10506-023-09356-9>.

Charter of Fundamental Rights of the European Union (2000/C 364/01).

Court of Justice of the European Union, *Breyer v Germany* (C-582/14) decided on 19 October 2016, ECLI:EU:C:2016:779.

Court of Justice of the European Union, *Nowak v Data Protection Commissioner* (C-434/16) decided on 20 December 2017, ECLI:EU:C:2017:994.

Court of Justice of the European Union, *Single Resolution Board v EDPS* (T-557/20), decided 26 April 2023, ECLI:EU:T:2023:219.

Court of Justice of the European Union, *Google Spain v AEPD* (C-131/12), decided 13 May 2014, ECLI:EU:C:2014:317.

Court of Justice of the European Union, *Weltimmo* (C-230/14), decided 1 October 2015, ECLI:EU:C:2015:639.

Court of Justice of the European Union, *La Quadrature du Net*, Joined cases C-511/18, C-512/18 and C-520/18, decided 6 October 2020, ECLI:EU:C:2020:791.

Court of Justice of the European Union, *Ligue des droits humains v. Council of Ministers*, C-817/19, decided 21 June 2022, ECLI:EU:C:2022:491.

Court of Justice of the European Union, *Norra Stockholm Bygg* (C-268/21) decided 2 March 2023, ECLI:EU:C:2023:145.

Court of Justice of the European Union, *Dun and Bradstreet Austria* (C-203/22), decided 27 February 2025, ECLI:EU:C:2024:745.

Court of Justice of the European Union, ‘Judgment of the Court in Case C-203/22 (Dun & Bradstreet Austria) – Automated credit assessment: the data subject is entitled to an explanation to understand and challenge the automated decision’ (CP-25 0022) Press Release No 22/25 of 27 February 2025, available at: <https://curia.europa.eu/jcms/upload/docs/application/pdf/2025-02/cp250022en.pdf>.

Dewitte, Pierre, ‘AI Meets the GDPR: Navigating the Impact of Data Protection on AI Systems’ (2025) in N. A. Smuha (ed.) *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence*. Cambridge: Cambridge University Press (Cambridge Law Handbooks), pp. 133–157, available at: <https://doi.org/10.1017/9781009367783.010>.

Dexe, Jacob, Ulrik Franke, Kasia Söderlund, Niels van Berkel, Rikke Hagensby Jensen, Nea Lepinkäinen, Juho Vaiste, ‘Explaining automated decision-making: a multinational study of the GDPR right to meaningful information’ (2022) 47 *The Geneva Papers on Risk and Insurance - Issues Practice*, pp. 669–697, available at: <https://doi.org/10.1057/s41288-022-00271-9>.

Druce, Jeff, Michael Harradon, James Tittle, ‘Explainable Artificial Intelligence (XAI) for Increasing User Trust in Deep Reinforcement Learning Driven Autonomous Systems’ (2021) 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada, available at: <https://doi.org/10.48550/arXiv.2106.03775>.

Engelfriet, Arnoud, ‘An Uninterpretable Right: Legal and Practical Limits of the Right to an Explanation’ (2025), available at: <http://dx.doi.org/10.2139/ssrn.5312780>.

European Data Protection Board, ‘Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679’ (2018) available at: https://edpb.europa.eu/our-work-tools/our-documents/guideline/automated-decision-making-and-profiling_en.

European Data Protection Board, ‘Report of the work undertaken by the ChatGPT Taskforce’ (23 May 2024), available at: https://www.edpb.europa.eu/system/files/2024-05/edpb_20240523_report_chatgpt_taskforce_en.pdf.

European Data Protection Supervisor, ‘TechDispatch: Explainable Artificial Intelligence’ (2023), doi: 10.2804/802043.

Eurostat, ‘Usage of AI technologies increasing in EU enterprises’ (2025), available at: <https://ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20250123-3>.

Finnish Data Protection Authority, *Tietosuojavaltuutetun toimisto* [2020] 6482/186/2020.

Fresz, Benjamin, Elena Dubovitskaya, Danilo Brajovic, Marco F. Huber, Christian Horz, ‘How Should AI Decisions Be Explained? Requirements for Explanations from the Perspective of European Law’ (2024) 7(1) Proceedings of the 2024 AAAI/ACM Conference on AI, Ethics, and Society, AAAI Press, pp. 438–450, available at: <https://doi.org/10.1609/aies.v7i1.31648>.

Georgios Pavlidis, ‘Unlocking the black box: analysing the EU artificial intelligence act’s framework for explainability in AI’ (2024) 16(1) *Law, Innovation and Technology*, available at: <https://doi.org/10.1080/17579961.2024.2313795>.

Giovine, Carlo and Roger Roberts, ‘Building AI trust: The key role of explainability’ (2024) *McKinsey & Company*, available at: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/building-ai-trust-the-key-role-of-explainability>.

Gomez-Uribe, Carlos A. and Neil Hunt, ‘The Netflix recommender system: Algorithms, business value, and innovation’ (2015) 6(4) *ACM Transactions on Management Information Systems*, available at: <http://dx.doi.org/10.1145/2843948>.

Grimmelikhuijsen, Stephan, ‘Explaining Why the Computer Says No: Algorithmic Transparency Affects the Perceived Trustworthiness of Automated Decision-Making’ (2023) 83(2) *Public Administration Review*, pp. 241–262, available at: <https://doi.org/10.1111/puar.13483>.

Hamon, Ronan, Henrik Junklewitz, Ignacio Sanchez, Gianclaudio Malgieri, Paul De Hert, ‘Bridging the Gap Between AI and Explainability in the GDPR: Towards Trustworthiness-by-Design in Automated Decision-Making’ (2022) 17(1) *IEEE Computational Intelligence Magazine*, pp. 72-85, doi: 10.1109/MCI.2021.3129960.

Hassija, Vikas, Vinay Chamola, Atmesh Mahapatra, Abhinandan Singal, Divyansh Goel, Kaizhu Huang, Simone Scardapane, Indro Spinelli, Mufti Mahmud & Amir Hussain, ‘Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence’ (2024) 16 *Cognitive Computation*, available at: <https://doi.org/10.1007/s12559-023-10179-8>.

Herrera, Francisco, ‘Reflections and attentiveness on eXplainable Artificial Intelligence (XAI). The journey ahead from criticisms to human–AI collaboration’ (2025) 121 *Information Fusion*, available at: <https://doi.org/10.1016/j.inffus.2025.103133>.

High-Level Expert Group on Artificial Intelligence, ‘Ethics Guidelines for Trustworthy AI’ (2019), European Commission, available at: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

Janiesch, Christian, Patrick Zschech, Kai Heinrich, ‘Machine learning and deep learning’ (2021) 31 *Electronic Markets*, available at: <https://doi.org/10.1007/s12525-021-00475-2>.

Jasmontaite, Lina, Irene Kamara, Gabriela Zanfir-Fortuna, Stefano Leucci, ‘Data Protection by Design and by Default: Framing Guiding Principles into Legal Obligations in the GDPR’ (2018) 4(2) *European Data Protection Law Review*, available at: <https://doi.org/10.21552/edpl/2018/2/7>.

Juliussen, Bjørn Aslak, ‘The Right to an Explanation Under the GDPR and the AI Act’ (2025) MultiMedia Modeling: 31st International Conference on Multimedia Modeling, MMM 2025, Nara, Japan, January 8–10, 2025, Proceedings, Part IV, pp. 184-197, available at: https://doi.org/10.1007/978-981-96-2071-5_14.

Kamarinou, Dimitra, Christopher Millard, Felicity Turton, ‘Protection of Personal Data in Clouds and Rights of Individuals’ (2021) in: C. Millard (ed.) *Cloud Computing Law*, (2nd edn, OUP 2021), available at: <http://dx.doi.org/10.2139/ssrn.4255833>.

Kamarinou, Dimitra, Christopher Millard, Jatinder Singh, ‘Machine Learning with Personal Data’ (2016) 247 *Queen Mary School of Law Legal Studies Research Paper*, available at: <https://ssrn.com/abstract=2865811>.

LeCun, Yann, Yoshua Bengio, Geoffrey Hinton., ‘Deep learning’ (2015) 521 *Nature*, available at: <https://doi.org/10.1038/nature14539>.

Li, Yuxi, ‘Deep Reinforcement Learning’ (2018), available at: <https://arxiv.org/abs/1810.06339>.

Loyola-González, Octavio, ‘Black-Box vs. White-Box: Understanding Their Advantages and Weaknesses From a Practical Point of View’ (2019) 7 *IEEE Access*, doi: 10.1109/ACCESS.2019.2949286.

Lu, Jun, ‘Gradient Descent, Stochastic Optimization, and Other Tales’ (2024), available at: <https://doi.org/10.48550/arXiv.2205.00832>.

Malgieri, Gianclaudio, ‘“Just” Algorithms: Justification (Beyond Explanation) of Automated Decisions Under the General Data Protection Regulation’ (2021) 1(1) *Law and Business*, Sciendo, pp. 16-28, available at: <https://doi.org/10.2478/law-2021-0003>.

Malgieri, Gianclaudio and Giovanni Comandé, ‘Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation’ (2017) 7(4) *International Data Privacy Law*, pp. 243–265, available at: <https://doi.org/10.1093/idpl/ix019>.

Maslej, Nestor, Loredana Fattorini, Raymond Perrault, Yolanda Gil, Vanessa Parli, Njenga Kariuki, Emily Capstick, Anka Reuel, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Juan Carlos Nieves, Yoav Shoham, Russell Wald, Tobi Walsh, Armin Hamrah, Lapo Santarlasci, Julia Betts Lotufo, Alexandra Rome, Andrew Shi, Sukrut Oak, ‘The AI Index 2025 Annual Report’ (2025) AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, available at: https://hai.stanford.edu/assets/files/hai_ai_index_report_2025.pdf.

Maxwell, Winston, and Bruno Dumas, ‘Meaningful XAI Based on User-centric Design Methodology: Combining Legal and Human-computer Interaction (HCI) Approaches to Achieve Meaningful Algorithmic Explainability’ (2023) *CERRE*, available at: <http://dx.doi.org/10.2139/ssrn.4520754>.

Metikoš, Ljubiša and Jef Ausloos, ‘The Right to an Explanation in Practice: Insights from Case Law for the GDPR and the AI Act’ (2024) *Law, Innovation and Technology*, pp. 1-36, available at: <https://doi.org/10.1080/17579961.2025.2469349>.

Michelakaki, Christina and Sebastião Barros Vale, ‘Unlocking Data Protection By Design & By Default: Lessons from the Enforcement of Article 25 GDPR’ (2023) *Future of Privacy Forum*.

Nisevic, Maja, Arno Cuypers, Jan De Bruyne, ‘Explainable AI: Can the AI Act and the GDPR go out for a date?’ (2024) International Joint Conference on Neural Networks (IJCNN), Yokohama, Japan, 2024, pp. 1-8, doi: 10.1109/IJCNN60899.2024.10649994.

Nikiforidis, Konstantinos, Alkiviadis Kyrtoglou, Thanasis Vafeiadis, Thanasis Kotsiopoulos, Alexandros Nizamis, Dimosthenis Ioannidis, Konstantinos Votis, Dimitrios Tzovaras, Panagiotis Sarigiannidis, ‘Enhancing transparency and trust in AI-powered manufacturing: A survey of explainable AI (XAI) applications in smart manufacturing in the era of industry 4.0/5.0’ (2025) 11 *ICT Express*, available at: <https://doi.org/10.1016/j.icte.2024.12.001>.

Olateju, Omobolaji Olufunmilayo, Samuel Ufom Okon, Oluwaseun Oladeji Olaniyi, Amaka Debie Samuel-Okon, and Christopher Uzoma Asonze, ‘Exploring the Concept of Explainable AI and Developing Information Governance Standards for Enhancing Trust and Transparency in Handling Customer Data’ (2024) 26(7) *JERR*, available at: <https://doi.org/10.9734/jerr/2024/v26i71206>.

Papagiannidis, Emmanouil, Patrick Mikalef, Kieran Conboy, ‘Responsible artificial intelligence governance: A review and research framework’ (2025) 34(2) *The Journal of Strategic Information Systems*, available at: <https://doi.org/10.1016/j.jsis.2024.101885>.

Perel, Maayan and Niva Elkin-Koren, ‘Black Box Tinkering: Beyond Disclosure in Algorithmic Enforcement’ (2017) 69(1) *Florida Law Review*, available at: <https://scholarship.law.ufl.edu/flr/vol69/iss1/5>.

Perlmutter, Maya, Ryan Gifford, Samantha Krening, ‘Impact of example-based XAI for neural networks on trust, understanding, and performance’ (2024) 188 *International Journal of Human-Computer Studies*, available at: <https://doi.org/10.1016/j.ijhcs.2024.103277>.

Rashid, Adib Bin, MD Ashfakul Karim Kausik, ‘AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications’ (2024) 7 *Hybrid Advances*, available at: <https://doi.org/10.1016/j.hybadv.2024.100277>.

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

Rossetti, Stefano, ‘The Court of Justice of the European Union confirms the existence of the right to explanation of automated decision-making’ (2025) *European Law Blog*, available at: <https://doi.org/10.21428/9885764c.flcc5989>.

Sartor, Giovanni, ‘The impact of the General Data Protection Regulation (GDPR) on artificial intelligence’ (2020) European Parliamentary Research Services (EPRS), Panel for the Future of Science and Technology, available at: [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU\(2020\)641530_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf).

Selbst, Andrew D. and Julia Powles, ‘Meaningful information and the right to explanation’ (2017) 7(4) *International Data Privacy Law*, pp. 233–242, available at: <https://doi.org/10.1093/idpl/ix022>.

Shrestha, Yash Raj, Vaibhav Krishna, Georg von Krogh, ‘Augmenting organizational decision-making with deep learning algorithms: Principles, promises, and challenges’ (2021) 123 *Journal of Business Research*, available at: <https://doi.org/10.1016/j.jbusres.2020.09.068>.

Sion, Laurens, Pierre Dewitte, Dimitri Van Landuyt, Kim Wuyts, Ivo Emanuilov, Peggy Valcke, ‘An Architectural View for Data Protection by Design’ (2019) 2019 IEEE International Conference on Software Architecture (ICSA), Hamburg, Germany, pp. 11-20, doi: 10.1109/ICSA.2019.00010.

Sovrano, Francesco, Fabio Vitali, Monica Palmirani, ‘Making Things Explainable vs Explaining: Requirements and Challenges Under the GDPR’ (2021) In: V. Rodríguez-Doncel et al. (eds) *AI Approaches to the Complexity of Legal Systems XI-XII. AICOL 2020, AICOL 2018 XAILA 2020. 13048 Lecture Notes in Computer Science*, pp. 169-182, Springer, Cham, available at: https://doi.org/10.1007/978-3-030-89811-3_12.

Sovrano, Francesco, Fabio Vitali, Monica Palmirani, ‘Modelling GDPR-Compliant Explanations for Trustworthy AI’ (2021), available at: <https://doi.org/10.48550/arXiv.2109.04165>.

State, Laura, Alejandra Bringas Colmenarejo, Andrea Beretta, Salvatore Ruggieri, Franco Turini, Stephanie Law, ‘The explanation dialogues: an expert focus study to understand requirements towards explanations within the GDPR’ (2025) *Artificial Intelligence and Law*, available at: <https://doi.org/10.1007/s10506-024-09430-w>.

Supreme Administrative Court Austria, *Verwaltungsgerichtshof* [2023] ECLI:AT:VWGH:2023:RO2021040010.J09.

The Royal Society, ‘Explainable AI: the basics’ (2019) ISBN: 978-1-78252-433-5, available at: <https://royalsociety.org/-/media/policy/projects/explainable-ai/AI-and-interpretability-policy-briefing.pdf>.

Thompson Reuters, ‘Recital (EU)’ *Practical Law*, available at: <https://uk.practicallaw.thomsonreuters.com/Glossary/PracticalLaw/I13f404e5785211e79bef99c0ee06c731>.

Tiwari, Rudra, ‘Explainable AI (XAI) and its Applications in Building Trust and Understanding in AI Decision Making’ (2023) 7(1) *IJSREM*, doi: 10.55041/IJSREM17592.

Valgaeren, Erik and Thibau Duquin, ‘CJEU confirms “right of explanation” in battle between trade secrets and algorithmic transparency’ (2025) *Stibbe*, available at: <https://www.stibbe.com/publications-and-insights/cjeu-confirms-right-of-explanation-in-battle-between-trade-secrets-and>.

Veale, Michael, Reuben Binns, Lilian Edwards, ‘Algorithms that remember: model inversion attacks and data protection law’ (2018) 376(2133) *Philosophical Transactions of the Royal Society A*, available at: <https://doi.org/10.1098/rsta.2018.0083>.

Véliz, Carissa, Carina Prunkl, Milo Phillips-Brown, Theodore M Lechterman, ‘We might be afraid of black-box algorithms’ (2021) 47 *Journal of Medical Ethics*, pp. 339–340, available at: <https://doi.org/10.1136/medethics-2021-107462>.

Wachter, Sandra, Brent Mittelstadt, Luciano Floridi, ‘Why a right to explanation of automated decision-making does not exist in the general data protection regulation’ (2017) 7(2) *International Data Privacy Law*, pp. 76–99, available at: <http://dx.doi.org/10.2139/ssrn.2903469>.

Wulf, Alexander J. and Ognyan Seizov, ‘“Please understand we cannot provide further information”: evaluating content and transparency of GDPR-mandated AI disclosures’ (2024) 39 *AI & Society*, available at: <https://doi.org/10.1007/s00146-022-01424-z>.

Xu, Min, Yiwen Wang, ‘Explainability increases trust resilience in intelligent agents’ (2024) *British Journal of Psychology*, available at: <https://doi.org/10.1111/bjop.12740>.

Yalcin, Orhan Gazi, ‘Governing Algorithms in the Big Data Era for Balancing New Digital Rights’ (2023) Doctor of Philosophy Thesis in Law, Science and Technology.

Zahra, Atf and Peter R. Lewis, "Is Trust Correlated With Explainability in AI? A Meta-Analysis," in *IEEE Transactions on Technology and Society*, doi: 10.1109/TTS.2025.3558448.

Zolanvari, Maede, Zebo Yang, Khaled Khan, Raj Jain, Nader Meskin, 'TRUST XAI: model-agnostic explanations for AI with a case study on IIoT security' (2023) 10(4) *IEEE Internet of Things Journal*, pp. 2967–2978, doi: 10.1109/JIOT.2021.3122019.

Zuboff, Shoshana, 'The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power' (2019) New York: PublicAffairs.