Algorithmic Documentation and the Limits of Freedom of Information: The EU AI Act's Two-Track Disclosure Architecture

Olesia Chizheva

University of Bologna, Bologna, Italy | lesiachizheva@gmail.com | orcid.org/0009-0008-6192-0544

ABSTRACT

Freedom of information law and EU AI governance law share a common premise: government accountability requires access to government information. In practice, they pull in opposite directions. FOIA regimes are reactive: they give citizens the right to request documents after the fact. The EU AI Act is prospective: it requires technical documentation before a high-risk system is deployed. When a FOIA request targets an AI system's decision logic, these two regimes collide. The resulting disclosure asymmetry is documented across five welfare AI systems in EU member states, each a case where public access mechanisms failed to produce algorithmic documentation or performance information that the AI Act would require to exist within the supervisory compliance architecture. The AI Act creates a two-track disclosure architecture: supervisory access to technical documentation is legally enforceable, while public access to model-level documentation is not. For FOIA purposes, algorithmic documentation remains vulnerable to withholding under fraud-prevention, public-control, and commercial-confidentiality exemption grounds in the access-to-documents disputes examined, while adjacent regulatory-disclosure mechanisms likewise failed to produce public model-level documentation. This architecture privileges regulatory insiders over citizens. The paper derives a minimum disclosure standard from the AI Act's existing obligations, one that closes the gap without requiring legislative amendment in its core elements.

CCS CONCEPTS

• Social and professional topics → Governmental regulations; • Computing methodologies → Artificial intelligence; • Security and privacy → Privacy protections.

KEYWORDS

freedom of information, AI Act, automated decision-making, GDPR, algorithmic transparency, government accountability, FOIA exemptions, public records disclosure

Reference Format:

Olesia Chizheva. 2026. Algorithmic Documentation and the Limits of Freedom of Information: The EU AI Act's Two-Track Disclosure Architecture. In Proceedings of the 1st AI & Open Government Workshop (AIOG @ ICAIL 2026), Singapore, June 8, 2026.

1 Introduction

Freedom of information law rests on a structural assumption: the documents a government agency holds are, in principle, available to the public it serves. That assumption was formed in an administrative environment where a decision was a document. A welfare assessor wrote a determination. The document was the decision.

AI systems break that equation. When a fraud-detection algorithm assigns a risk score that determines whether a control procedure is initiated, the operative output is not a document in any conventional sense. It is a numerical output produced by a model whose weights, training data, and validation history may exist in a form no caseworker has read or could read. The document trail that FOIA was designed to reach may not capture the decision at all.

Five cases from EU member states show what happens when public records mechanisms target that gap: in France, La Quadrature du Net spent several years in CADA and litigation proceedings seeking access to algorithmic materials related to a scoring system that had processed 32.3 million individuals annually since 2010; in Denmark, Amnesty International's information request for Udbetaling Danmark's machine learning model documentation was refused on the ground, confirmed by the Ombudsman in 2022, that disclosure would undermine essential considerations relating to public control; in the Netherlands, the Hague District Court found in 2020 that SyRI's risk model had never been publicly documented; in Ireland, nearly four years of DPC investigation produced a compliance decision but not public disclosure of the matching algorithm; in Estonia, the Unemployment Insurance Fund acknowledged automated decision-making but has published no technical documentation.

Existing scholarship on algorithmic accountability has examined GDPR's explanation rights [9, 16], the AI Act's supervisory architecture [10, 17], and the governance of welfare AI systems [12, 13, 18]. The interaction between FOIA regimes and the AI Act's two-track disclosure architecture, as it plays out in documented enforcement proceedings, has not been systematically examined. Larsson and Heintz's analysis of transparency requirements in AI regulation addresses transparency as a design obligation but not as an information access problem under positive law [19]. Scholarship on open government data focuses on structured datasets rather than model documentation,

and does not address this intersection. That gap is what this paper addresses.

2 Legal Framework: FOIA, the AI Act, and the Disclosure Gap

2.1 FOIA Regimes and Algorithmic Outputs

At EU institutional level, Regulation (EC) 1049/2001 on access to documents requires disclosure of documents held by Union institutions, subject to exceptions including protection of commercial interests and the decision-making process. No provision addresses algorithmic models as a document category. The regulation was drafted in 2001; the drafting history contains no reference to machine learning, training data, or model weights.

Member state FOIA regimes vary, but share a common structural feature: disclosure obligations attach to documents defined by their recorded form, not by their functional significance. In the CNAF and Udbetaling Danmark cases, the decisive point was not that no recorded material existed. It was that access to model-related material could be limited through exemptions protecting fraud prevention, public control, and related administrative interests [1, 2].

The CNAF case illustrates this precisely. The CADA did not hold that scoring variables and source code fall outside the legal category of administrative documents. It held that even where such materials qualify as administrative documents, access may still be defeated by exemptions protecting the effectiveness of fraud prevention and administrative controls [1]. The doctrinal gap is therefore not primarily about classification of algorithmic outputs as documents or non-documents. It is about the breadth of exemption grounds that member state FOIA regimes recognise, grounds wide enough to cover fraud-detection logic, risk model parameters, and validation methodology. Closing this gap requires either legislative redefinition of those exemption grounds or a parallel disclosure regime. The AI Act provides the latter, but routes it exclusively to supervisory authorities.

FOIA regimes were designed to give access to outputs of administrative decisions: the letter, the refusal, the determination. They were not designed to reach the process that generated those outputs, except where that process was itself recorded in a conventional document. Automated systems produce outputs without producing the intermediate record FOIA contemplated. That is the doctrinal gap.

2.2 The AI Act's Disclosure Architecture

The EU AI Act (Regulation (EU) 2024/1689) establishes a parallel disclosure regime for high-risk AI systems. Under Article 11, providers must maintain technical documentation meeting the Annex IV specification before placing a system on the market. Deployers of high-risk systems in public administration must conduct a fundamental rights impact assessment before deployment, per Article 27. High-risk systems must also be registered in the EU database under Article 71 [13].

The primary recipients of technical documentation are supervisory authorities, not citizens. National market surveillance authorities can request and inspect Annex IV documentation under Article 74; what they obtain is subject to confidentiality obligations under Article 78. Article 13 is deployer-facing: it requires providers to supply instructions for use and performance-related information enabling deployers to interpret and use the system appropriately. The affected person receives a narrower notice under Article 26(11): that they are subject to the use of a high-risk AI system. Neither provision gives the public access to Annex IV technical documentation.

The resulting architecture privileges accountability through expert regulators over transparency to citizens. That choice is coherent within its own logic. It becomes a problem when the specialist regulator does not act, which is exactly the pattern the five cases in Section 3 establish.

The EU database under Article 71 does not reproduce Annex IV technical documentation. For public-authority deployers, Annex VIII Section C requires registration of a summary of FRIA findings and, where applicable, a DPIA summary. Section A provider information is publicly accessible. This is useful for notice and traceability purposes, but insufficient for independent assessment of model validation, error distribution, or discriminatory performance across a population.

2.3 The Collision

Consider a citizen whose welfare benefits were suspended in France in 2023. Under GDPR Articles 13(2)(f), 14(2)(g), and 15(1)(h), read together with the safeguards in Article 22(3), she may obtain meaningful information about the logic involved and contest the individual decision. She has no right, under any existing instrument, to the aggregate false positive rate of the model that generated her risk score, or to the validation data that would allow her to assess whether the error is systemic rather than individual. The explanation tells her why she was selected. The aggregate statistics tell her whether the selection system is functioning lawfully at scale. These are different legal objects. FOIA is the principal public-access instrument examined here that could potentially reach the second category, and the CNAF proceedings show its limits.

When a citizen submits a FOIA request for the technical documentation of a high-risk AI system used in a government agency, three legal instruments are potentially engaged: the applicable FOIA regime, Article 22 GDPR, and the AI Act's transparency obligations. Taken together, they do not guarantee access to model-level documentation. FOIA can be blocked by fraud-prevention, public-control, or commercial-confidentiality exemptions. The GDPR's information and access provisions, read together with Article 22(3), provide individual-level safeguards and access to meaningful information about the logic involved; they do not provide access to training data, validation methodology, or system-wide performance metrics [9]. The AI Act routes its technical transparency obligations to supervisory authorities, not to the public.

This is the disclosure asymmetry: market surveillance authorities have legally enforceable access to documentation that citizens cannot compel through any existing legal mechanism.

The asymmetry reflects a structural choice to route technical supervision through specialist regulators. The five cases in Section 3 suggest that choice is not adequate.

3 Case Analysis: Five Public Records Encounters with Welfare AI

The five systems examined here were selected because each generated a documented encounter between public accountability mechanisms and algorithmic documentation. The first three involved direct access-to-documents or FOIA-equivalent attempts to obtain model documentation. The Irish and Estonian examples are adjacent regulatory-disclosure cases: they show that even where supervisory or statutory mechanisms acknowledge automated processing, they do not necessarily produce public model-level documentation. The systems are: the CNAF scoring algorithm in France [3]; the Udbetaling Danmark fraud detection suite in Denmark [4]; the SyRI risk indication system in the Netherlands [5]; the Department of Social Protection SAFE 2 biometric system in Ireland [6]; and the Estonian Unemployment Insurance Fund automated decision system [7].

La Quadrature du Net submitted access requests to CNAF beginning in 2021, initiating CADA proceedings and subsequent litigation. The CADA found in December 2022 that CNAF was entitled to withhold the current algorithmic variables and coefficients, not because the materials fell outside the administrative document category, but because the fraud-prevention exemption overrode access to the current model; prior versions were ordered disclosed [1]. A coalition of fifteen civil society organisations filed suit before the Conseil d'Etat in October 2024, seeking annulment of the algorithm on grounds of discrimination and unlawful surveillance. Under that sustained adversarial pressure, CNAF published the source code of its current algorithm on 15 January 2026, not as a result of any FOIA ruling, but as a concession during ongoing litigation [3]. A 2025 internal CNAF study obtained by the coalition confirmed the algorithm's discriminatory effects; the litigation continues. Disclosure came only after years of coordinated action before the highest administrative court.

In Denmark, Amnesty International's 2024 investigation documented that Udbetaling Danmark refused to provide model documentation for review, supplying only heavily redacted materials. The Ombudsman's 2022 decision accepted UDK's withholding of parts of the model documentation on the basis that disclosure would undermine essential considerations relating to public control [2]. The near-90% no-further-action rate of the Model Abroad algorithm, used by Amnesty International as a proxy for false positives, did not become visible through a routine access pathway available to an affected welfare recipient. It emerged through a two-year investigation combining FOI-based material, parliamentary data requests, and interviews, a resource investment unavailable to an individual citizen [4].

SyRI is the starkest case. A WOB request in June 2019 produced operational data about SyRI's deployment, including use statistics and neighbourhood targeting records, but not the risk model's logic. The Hague District Court's February 2020 judgment noted that the model had never been publicly documented and that the court could not assess the discrimination argument because it was not disclosed [5]. Civil litigation under Article 8 ECHR, not information law, was the only mechanism that forced judicial scrutiny of the system itself. The government's position throughout was that no document responsive to a model-specific information request existed: the algorithmic logic was embedded in software, not recorded in any conventional administrative record.

The Irish and Estonian cases illustrate the same asymmetry through regulatory proceedings. The DPC's decision of 9 June 2025 in IN-21-7-3 found infringements of Articles 5(1)(a), 6(1), 9(1), 5(1)(e), 13(1)(c), 13(2)(a), 35(7)(b), and 35(7)(c) GDPR in DSP's operation of SAFE 2 [6]. Nearly four years of investigation yielded a compliance finding, a reprimand, and fines totalling EUR 550,000, together with an order to cease biometric data processing within nine months unless a lawful basis is identified. Neither the decision nor any other instrument produced public disclosure of the facial matching algorithm's technical documentation; the DPC has no mandate to order such disclosure. Under Annex VIII Section C, a public-authority deployer would be required to register a summary of FRIA findings in the EU database, but this obligation did not yet apply when SAFE 2 was deployed, and a FOIA request for the underlying assessment would face the same exemption grounds that blocked model documentation elsewhere. The DSP appealed in July 2025; proceedings are pending. In Estonia, explicit legislative authorisation of automated decision-making under paragraph 23(4) of the Unemployment Insurance Act makes the system legally transparent at the level of its existence. That is the limit of the transparency. No Annex IV-equivalent documentation has been published, no validation data is available to the public, and the sources reviewed for this paper identify no public DPA assessment of whether the profiling models produce differential outcomes across demographic groups [7].

Across all five systems, public records mechanisms yielded partial disclosure at best, years after deployment, through adversarial proceedings sustained by well-resourced civil society organisations. A citizen subject to an individual decision had no routine path to model-level documentation.

4 The Structural Disclosure Asymmetry

The failures share a cause. In each of the five cases, the information needed to evaluate the system's compliance existed, or should have existed, within the deploying agency or its provider. Legal mechanisms designed to compel disclosure either did not reach that information or were successfully resisted for years.

Had the AI Act's relevant high-risk obligations applied, providers would have been required to maintain Annex IV technical documentation, while public-authority deployers would have been required to complete the Article 27 FRIA and register the Annex VIII Section C information in the EU database. Where GDPR Article 35 was triggered, the relevant controller would also have been required to conduct a DPIA before high-risk

processing. The Annex IV technical documentation would have been accessible to the national market surveillance authority under Article 74. It would not have been publicly accessible as Annex IV documentation under the AI Act, although public-authority deployers would register FRIA and DPIA summaries under Annex VIII Section C. Most high-risk AI obligations under the Act apply from 2 August 2026 per Article 113; for systems deployed before that date, this analysis remains counterfactual unless a system undergoes substantial modification after that date.

Supervisory authorities have legally enforceable access to Annex IV documentation, including training data descriptions, validation methodology, and performance metrics. Citizens have access to Section A provider information in the Article 71 database and, for public-authority deployers, summaries of FRIA and DPIA findings under Annex VIII Section C. This is useful for notice and traceability. It is insufficient for independent assessment of whether a system's validated performance matches its actual deployment record or whether its error rates are distributed uniformly across demographic groups. FOIA requests for the underlying Annex IV documentation face the same exemption grounds that defeated every case in Section 3.

Compliance problems documented above would not have been detectably different under an AI Act regime unless supervisory authorities used their Article 74 powers proactively. In none of the five jurisdictions examined is there evidence that they would have done so absent the civil society pressure that produced the proceedings in the first place. The AI Act shifts the documentary burden from litigation to registration without shifting the incentive structure that made proactive supervision unlikely.

Article 22 GDPR adds a second dimension. Where a welfare AI system falls within Article 22(1), affected individuals receive individual-level safeguards under Article 22(3), together with access to meaningful information about the logic involved under the GDPR's information and access provisions [8]. That covers the individual decision, not the system's aggregate behaviour. Access to training data, validation methodology, or performance metrics across the whole population of decisions falls outside it [9, 10]. Accountability at the systemic level requires a different instrument. FOIA was the principal public-access candidate examined here, and the cases show its limits.

5 Towards a Minimum Disclosure Standard

The disclosure asymmetry is a product of how existing obligations have been routed, not of the AI Act's architecture. The Act already contains components of a minimum public disclosure standard, and operationalising those components avoids the need for legislative amendment in several key areas.

The Annex VIII Section C registration requirement already mandates a summary of FRIA findings for public-authority deployers. The remaining transparency gap concerns the content, granularity, comparability, and evidentiary value of that summary, not the complete absence of a publication channel. EDPB and AI Board guidance could specify minimum content standards: the categories of persons assessed as potentially affected, the identified risks, the mitigating measures adopted, and the demographic breakdown of error rates where differential impact was identified. Making those standards mandatory and uniform would transform a nominal registration entry into a genuine accountability instrument.

Beyond summaries, aggregate performance statistics derivable from Annex IV documentation should be disclosed proactively at regular intervals. Aggregate no-further-action rates by benefit category, and the date of the most recent validation exercise, do not reveal feature weights or detection thresholds. What they provide is the ability to compare documented performance against actual deployment records, and to identify whether error distribution is demographically uniform. Udbetaling Danmark's near-90% no-further-action rate on Model Abroad became publicly known only through two years of adversarial investigation. It should have been on a public register.

The objection that aggregate performance statistics enable gaming of fraud detection systems is legitimate but overstated. Article 35 GDPR already requires a DPIA for high-risk processing that contains materially equivalent information: the necessity and proportionality of the processing, the risks to data subjects, and the measures to address those risks. That assessment is produced for supervisory purposes without any reported compromise of fraud detection capacity across the five systems examined. The commercial secrecy and public-control exemptions protect specific detection logic. They do not justify suppressing evidence of structural discrimination in systems that process tens of millions of people annually.

Some elements of the proposed standard can be advanced without legislative amendment, through database design, guidance templates, and standardised FRIA and DPIA summary formats. Article 13 has indirect relevance here: it ensures that deployers receive performance and limitation information from providers that could inform standardised public summaries. A mandatory expansion of database fields beyond those currently specified in Annex VIII would require a clearer legal basis and may ultimately require legislative amendment rather than implementation guidance alone.

Getting here faces institutional rather than purely legal obstacles. The Commission's current priorities centre on AI Act implementation itself: market surveillance infrastructure, notified body designation, the EU AI Office. Member states with large-scale fraud detection programmes have operational incentives to preserve secrecy around targeting logic and may resist transparency duties perceived as weakening control capacity. Civil society pressure remains fragmented across data protection, welfare rights, and access-to-information communities, in contrast to the coordinated NGO pressure that shaped GDPR's Article 22 protections. The measures proposed here are legally available in significant part. Whether they are politically available depends on whether the enforcement failures documented in Section 3 become visible enough to shift that calculus. That is precisely the function that civil society organisations, investigative journalists, and workshops like this one serve.

The gap between what supervisory authorities can access and what citizens can compel through FOIA is a default setting. Existing instruments can adjust it, at least in part.

6 Conclusion

Five systems, five jurisdictions, five encounters between public records mechanisms and algorithmic documentation. In each case, the instruments designed to ensure government accountability could not reach the operative logic of automated decision-making. The information gap persisted for years, sustained by legal exemptions that were valid and rational from the perspective of the agencies asserting them.

The EU AI Act does not close that gap. It creates a supervisory disclosure channel and a public registration database. The channel routes technical documentation to specialist regulators. The database provides registration-level information and, for public-authority deployers, summaries of FRIA and DPIA findings, useful for notice, insufficient for substantive accountability assessment. FOIA requests for Annex IV documentation face the same exemption grounds as before.

What the Act does create is the infrastructure for a minimum disclosure standard that closes the gap in significant part without legislative amendment. Annex VIII Section C and Article 27 have unused public-facing capacity. Article 13 has indirect relevance because it ensures that deployers receive performance and limitation information that could inform standardised public summaries. Operationalised through standardised FRIA summaries, database templates, and guidance on what performance information public deployers should publish in aggregate form, the Act could make meaningful accountability information available without exposing model weights or detection thresholds. The disclosure asymmetry between supervisory insiders and citizens is a design choice, not an architectural constraint. The five cases examined here were available before the regulation was finalised. The final architecture does not appear to have absorbed their central lesson: supervisory access alone does not secure public accountability.

REFERENCES

[1] Commission d'Acces aux Documents Administratifs. 2022. Avis n. 20226179, Seance du 15 decembre 2022. CADA, Paris. https://www.cada.fr/20226179-0

[2] Folketingets Ombudsmand. 2022. Udtalelse 2022-23: Udbetaling Danmark kunne undtage oplysninger fra aktindsigt. Ombudsmanden, Copenhagen. https://www.ombudsmanden.dk/find-viden/udtalelser/2022/2022-23

[3] La Quadrature du Net. 2026. CNAF's Discriminatory Scoring Algorithm: 10 New Organisations Join the Case before the Conseil d'Etat. https://www.laquadrature.net/en/2026/01/20/cnafs-discriminatory-scoring-algorithm-10-new-organisations-join-the-case-before-the-conseil-detat-in-france/

[4] Amnesty International. 2024. Coded Injustice: Surveillance and Discrimination in Denmark's Automated Welfare State. Amnesty International, London. https://www.amnesty.org/en/latest/news/2024/11/denmark-ai-powered-welfare-system-fuels-mass-surveillance

[5] NJCM et al. v The State of the Netherlands (SyRI). 2020. Rechtbank Den Haag, ECLI:NL:RBDHA:2020:1878, 5 February 2020.

[6] Data Protection Commission. 2025. Decision IN-21-7-3, 9 June 2025: Inquiry concerning the Department of Social Protection (SAFE 2 registration). DPC, Dublin. Appeal filed by DSP, July 2025; proceedings pending.

[7] Zolkin, V., Chochia, A., and Hoffmann, T. 2023. Automated Decision-Making in the EU Member State's Public Administration. Eur. Stud. 10, 2, 178-202. https://doi.org/10.2478/eustu-2023-0017

[8] Article 29 Working Party / EDPB. 2018. Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679 (WP251rev.01). EDPB, Brussels. https://ec.europa.eu/newsroom/article29/items/612053

[9] Binns, R. and Veale, M. 2021. Is That Your Final Decision? Multi-Stage Profiling, Selective Effects, and Article 22 of the GDPR. Int. Data Priv. Law 11, 4, 319-332. https://doi.org/10.31235/osf.io/7mq6z

[10] Lazcoz, G. and De Hert, P. 2023. Humans in the GDPR and AIA Governance of Automated and Algorithmic Systems. Comput. Law Secur. Rev. 50, 105833. https://doi.org/10.1016/j.clsr.2023.105833

[11] Palmiotto, F. 2024. When Is a Decision Automated? A Taxonomy for a Fundamental Rights Analysis. Ger. Law J. 25, 2, 210-236. https://doi.org/10.1017/glj.2023.112

[12] Van Bekkum, M. and Zuiderveen Borgesius, F. 2021. Digital Welfare Fraud Detection and the Dutch SyRI Judgment. Eur. J. Soc. Secur. 23, 3, 272-295. https://doi.org/10.1177/13882627211031257

[13] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 (Artificial Intelligence Act). OJ L 2024/1689, 12 July 2024.

[14] Regulation (EC) No 1049/2001 of the European Parliament and of the Council of 30 May 2001 regarding public access to European Parliament, Council and Commission documents. OJ L 145, 31 May 2001.

[15] Enqvist, L. 2024. Rule-Based versus AI-Driven Benefits Allocation. Inf. Commun. Technol. Law 33, 2, 222-246. https://doi.org/10.1080/13600834.2024.2349835

[16] Wachter, S., Mittelstadt, B., and Floridi, L. 2017. Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. Int. Data Priv. Law 7, 2, 76-99. https://doi.org/10.2139/ssrn.2903469

[17] Hohmann, B. and Kollar, G. 2025. Reflections on the Data Protection Compliance of AI Systems under the EU AI Act. Cogent Soc. Sci. 11, 1. https://doi.org/10.1080/23311886.2025.2560654

[18] Malgieri, G. 2019. Automated Decision-Making in the EU Member States. Comput. Law Secur. Rev. 35, 5, 105327. https://doi.org/10.1016/j.clsr.2019.05.002

[19] Larsson, S. and Heintz, F. 2020. Transparency, Integrity, and Access: Three Core Concepts in an AI Policy Landscape. AI Soc. 35, 4, 1073-1080. https://doi.org/10.1007/s00146-019-00937-2

[20] AlgorithmWatch. 2020. Automating Society Report 2020: Estonia. AlgorithmWatch, Berlin. https://automatingsociety.algorithmwatch.org/report2020/estonia/