

From Model Duties to Downstream Fundamental Rights: Contextual Accountability for GPAI under the EU AI Act

Hannan Zhang¹, Xiangyi Liu²

¹ University of Sheffield

² University of Edinburgh

h Zhang242@sheffield.ac.uk, X.Liu-231@sms.ed.ac.uk

Abstract

Under the EU Artificial Intelligence Act (AI Act), general-purpose AI (GPAI) raises a value chain problem for fundamental rights protection. GPAI-specific duties attach primarily to model providers, including requirements to produce and share certain model-level information, while downstream safeguards depend largely on the classification of the AI system and the context in which it is used. Yet rights impacts often become concrete only after a model is integrated into downstream systems and used in institutional settings. This paper examines when and why model-level GPAI obligations may fail to support downstream fundamental rights accountability. It develops a Contextual Accountability Gap Framework that distinguishes model-intrinsic, configuration-mediated, and use-context-dependent risks, and assesses whether upstream information remains specific, actionable, and contestable across the AI value chain. Through illustrative rights-sensitive scenarios, the paper shows how contextual accountability gaps emerge from the uneven translation of model-level information into system design, deployment oversight, and context-specific safeguards. It concludes by identifying three governance mechanisms that can better connect GPAI model duties to downstream rights protection: context-aware documentation, downstream translation records, and contextual fundamental rights review.

1 Introduction

General-purpose AI (GPAI) is increasingly being integrated into systems that support communication, classification, recommendation, summarization, and decision-making across institutional contexts. A GPAI model may be developed upstream, made available through an API or other access mechanism, integrated into multiple downstream systems, and deployed for different purposes. The same model may support low-stakes drafting, recruitment assistance, public-service chatbots, educational feedback, legal or administrative summarization, and other forms of organizational support. This makes GPAI a

particularly revealing case for AI governance: the rights significance of a model is often not fully determined at the point of development, but is shaped by downstream configuration and institutional reliance (Bommasani et al. 2021; Weidinger et al. 2022; Cobbe, Veale, and Singh 2023).

This creates a value chain problem for fundamental rights protection. The actor with the most information about the model may not know the full context of use, while the actor closest to affected persons may depend on information produced upstream. Model providers, downstream system providers, and deployers occupy different positions in the AI value chain, with different forms of knowledge, control, and responsibility (Cobbe, Veale, and Singh 2023; Williams, Schuett, and Anderljung 2026). For GPAI-enabled systems, rights protection therefore depends not only on whether risks are identified at the model layer, but also on whether model-level information can travel downstream and become usable for system design, deployment oversight, and contestability.

The EU AI Act addresses AI governance through a risk-based framework that aims to promote human-centric and trustworthy AI while protecting health, safety, and fundamental rights (European Parliament and Council of the European Union 2024). The Act does not place responsibility for downstream fundamental rights protection solely on GPAI model providers. Its architecture distinguishes between AI systems and GPAI models, and allocates obligations across model providers, downstream system providers, and deployers.

However, this layered structure raises a connective question. GPAI-specific duties are primarily attached to the model layer: GPAI model providers must prepare technical documentation and make certain information and documentation available to downstream providers. Yet many rights-relevant impacts become concrete only after the model is integrated into a downstream system and used in context. This paper therefore asks: *When and why do model-*

level GPAI obligations fail to support downstream fundamental rights accountability in rights-sensitive institutional contexts?

The paper argues that GPAI model-level duties are important for fundamental rights protection, but they do not become downstream safeguards by themselves. Documentation and information-sharing matter because downstream actors depend on upstream knowledge about model capabilities and limitations. But that information must be translated into domain-specific testing, system design, and institutional safeguards before it can support downstream accountability.

The paper makes three contributions. First, it clarifies how the AI Act's distinction between GPAI models and AI systems structures rights-protective duties across the value chain. Second, it develops a Contextual Accountability Gap Framework that distinguishes model-intrinsic, configuration-mediated, and use-context-dependent risks, and evaluates upstream information through the criteria of specificity, actionability, and contestability. Third, it shows why AI Act implementation should focus not only on whether GPAI providers disclose information, but on whether that information remains usable across the value chain.

The remainder of the paper proceeds as follows. Section 2 sets out the AI Act's value chain structure for GPAI. Section 3 develops the Contextual Accountability Gap Framework. Section 4 applies the framework to illustrative scenarios. Section 5 discusses implementation implications. Section 6 concludes.

2 The AI Act's Value Chain Architecture for GPAI

The EU AI Act creates a layered governance structure for AI. For present purposes, the key actors are GPAI model providers, AI system providers, and deployers, operating across the model, system, and deployment layers of the AI value chain. This structure is especially important for GPAI because GPAI models are not tied to one fixed downstream use, but can be integrated into a variety of AI systems and deployed across different institutional contexts.

This section sets out the parts of the AI Act's architecture that are necessary for the argument. It explains how the Act distributes obligations across three layers: the GPAI model layer, the AI system layer, and the deployment layer. This provides the regulatory background for the Contextual Accountability Gap Framework developed later in the paper.

2.1 GPAI Models, AI Systems, and Downstream Integration

The AI Act distinguishes between GPAI models and AI systems. An AI system is defined by reference to a machine-

based system that, for explicit or implicit objectives, infers from input how to generate outputs such as predictions, content, recommendations, or decisions that may influence physical or virtual environments (Article 3(1)). By contrast, a GPAI model is a model that displays significant generality, is capable of competently performing a wide range of distinct tasks, and can be integrated into a variety of downstream systems or applications (Article 3(63)).

This distinction is central to the Act's value chain structure. A GPAI model provides general-purpose capabilities, while downstream AI systems may integrate those capabilities into applications with more specific functions and intended purposes. Deployment then concerns the use of such systems by deployers in particular contexts of use.

This distinction also structures responsibility across the value chain. GPAI model providers are regulated as upstream actors, while deployers are regulated in relation to AI system use in organizational and institutional settings. Obligations therefore attach at different stages.

2.2 Model-Level GPAI Duties

The AI Act imposes several obligations on providers of GPAI models, including technical documentation, information-sharing with downstream providers, copyright-related policies, and training content summaries (Article 53). For this paper, the most relevant duties are documentation and information-sharing. Subject to Article 53(2), providers must prepare and keep up-to-date technical documentation covering the model's training and testing process and evaluation results (Article 53(1)(a); Annex XI). They must also make available information and documentation to providers who intend to integrate the GPAI model into their AI systems, so that those downstream providers can understand the model's capabilities and limitations and comply with their own obligations under the Act (Article 53(1)(b); Annex XII).

The Act imposes additional obligations on providers of GPAI models with systemic risk. These include performing model evaluations, assessing and mitigating systemic risks, tracking and reporting serious incidents, and ensuring an adequate level of cybersecurity protection for the model and its physical infrastructure (Article 55(1)(a)-(d)). Together, Articles 53 and 55 locate a set of obligations at the model layer. They require GPAI model providers to generate and share information about the model, and in the case of models with systemic risk, to conduct additional evaluations and risk mitigation measures. Downstream AI systems integrating GPAI models, and their use, are then addressed through the Act's separate system-level and deployment-level provisions.

2.3 System-Level Duties for AI Systems

The AI Act separately regulates AI systems according to their risk category and intended purpose. Some AI practices are prohibited outright (Article 5). Beyond prohibited practices, the Act's most detailed system-level requirements apply to high-risk AI systems.

A system may be high-risk where it is intended to be used as a safety component of a product, or is itself a product, covered by specific Union harmonization legislation and subject to third-party conformity assessment (Article 6(1); Annex I). It may also be high-risk where it falls within one of the areas listed in Annex III, subject to the conditions and exceptions set out in the Act (Article 6(2)-(3); Annex III). These areas include employment, education, access to essential private and public services, and law enforcement (Annex III).

The Act sets out the following requirements for high-risk AI systems. Providers must establish a risk management system (Article 9), meet data governance requirements where relevant (Article 10), prepare and update technical documentation (Article 11; Annex IV), enable automatic logging (Article 12), ensure sufficient transparency and provide instructions for use (Article 13), enable human oversight (Article 14), and ensure appropriate levels of accuracy, robustness, and cybersecurity (Article 15). These obligations are more application-specific than the GPAI model obligations.

Article 25(4) also requires providers of high-risk AI systems and relevant third-party model suppliers to specify by written agreement the information, capabilities, technical access, and assistance needed for compliance (European Parliament and Council of the European Union 2026).

For AI systems that are not classified as high-risk, the Act generally imposes fewer system-level obligations. The most relevant obligations for this paper are the transparency duties that apply to certain systems, including AI systems intended to interact directly with natural persons or that generate synthetic content (Article 50). These duties are narrower than the high-risk requirements and primarily concern disclosure.

2.4 Deployment-Level Duties and Fundamental Rights Impact Assessment

The AI Act also imposes obligations on deployers, especially deployers of high-risk AI systems. Deployers must take appropriate technical and organizational measures to use such systems in accordance with the instructions for use (Article 26(1)). They must assign human oversight to natural persons with the necessary competence, training, authority, and support (Article 26(2)). They must also monitor system operation based on the instructions for use and notify relevant actors and authorities of risks or serious incidents (Article 26(5)).

The deployment layer is especially important for the AI Act's connection to fundamental rights. Specified deployers must conduct a fundamental rights impact assessment (FRIA) before deploying certain high-risk AI systems (Article 27). The assessment must describe the deployer's processes in which the system will be used, the period and frequency of use, the categories of natural persons and groups likely to be affected, the specific risks of harm likely to affect them, the human oversight measures to be implemented, and measures to be taken if those risks materialize (Article 27(1)).

Article 27 is important because it shifts attention from the AI system in the abstract to the system's use in context. It asks how the system fits into the deployer's processes, who is likely to be affected, and what oversight and mitigation measures are available. In this respect, it is different from both GPAI model documentation and high-risk system technical documentation. It is concerned with the situated use of the system and its potential effects on natural persons and groups.

Article 50 also includes transparency duties relevant to deployment. Deployers of emotion recognition systems or biometric categorization systems must inform exposed natural persons of the operation of the system (Article 50(3)). Deployers of AI systems generating or manipulating content may also be subject to disclosure obligations in specific cases (Article 50(4)).

These deployment-level provisions show that the AI Act also recognizes that the use of an AI system in a particular context may affect natural persons and fundamental rights. In such cases, deployer obligations, human oversight, transparency duties, and FRIAs are designed to address the relationship between the system and the affected persons in context.

2.5 The Value Chain Structure

The AI Act therefore distributes obligations across different layers of the AI value chain. This layered structure reflects the different knowledge and control held by different actors. GPAI model providers may know most about model development and evaluation. Downstream AI system providers may know most about how the model is integrated into a concrete application. Deployers may know most about the institutional context, affected persons, and actual use. The Act's architecture allocates obligations across these different roles.

This layered structure provides the regulatory background for the analysis that follows. The AI Act does not locate GPAI accountability at a single point in the value chain. Instead, it distributes model-level duties, system-level safeguards, and deployment-level responsibilities across different actors and legal mechanisms. The next section builds on this structure to develop the Contextual

Accountability Gap Framework, which assesses when model-level GPAI duties can support downstream fundamental rights protection and when the connection breaks down.

3 The Contextual Accountability Gap Framework

The preceding analysis shows that the AI Act's distinction between GPAI models and AI systems is not merely a technical classificatory choice. It structures how responsibility, information, and rights-protective duties are distributed across the AI value chain.

This section develops the Contextual Accountability Gap Framework for assessing when GPAI model-level obligations can meaningfully support downstream fundamental rights protection, and when they leave a gap between plausible rights impacts and legally available mechanisms of accountability.

3.1 Identifying Rights-Sensitive Downstream Contexts

The first step is to ask whether the GPAI-enabled system operates in a context where fundamental rights interests are plausibly affected. This inquiry is broader than determining whether a system falls within the AI Act's high-risk classification (Veale and Zuiderveen Borgesius 2021). The high-risk regime remains the Act's most developed mechanism for connecting AI governance to fundamental rights protection, but high-risk classification does not exhaust the universe of rights-sensitive AI uses (Veale and Zuiderveen Borgesius 2021; Mantelero 2022). Some GPAI-enabled systems may affect access to opportunities, public services, information, or institutional decision-making without clearly falling within the legal categories that trigger the Act's most intensive obligations (Selbst 2021; Cobbe, Veale, and Singh 2023).

Rights-sensitive contexts include, but are not limited to, employment, education, healthcare, financial access, migration and asylum, policing and security, social benefits, public administration, housing, content moderation, political communication, and access to essential services (Mantelero 2018; Eubanks 2018). These contexts are rights-sensitive not simply because they involve important interests, but because they often combine unequal power relations, informational asymmetries, dependency on institutional decisions, and difficulties for affected persons to detect or challenge AI-mediated influence (Eubanks 2018; Selbst et al. 2019; Kroll 2021). Existing literature on AI and human rights has repeatedly emphasized that algorithmic systems may implicate equality, privacy, due process, freedom of expression, access to services, and effective remedies (Mantelero 2018; Barocas, Hardt, and Narayanan

2023), while accountability is particularly difficult where affected persons cannot understand or contest the role played by automated systems (Kroll 2021; Nannini 2024).

This first step therefore shifts attention from the model in isolation to the setting in which the model is used (Selbst et al. 2019; Mantelero 2022). A GPAI model used to draft generic marketing copy raises different concerns from the same model used to summarize welfare case files, generate risk indicators for students, support recruitment screening, or mediate access to public services. The relevant question is not whether the model is inherently rights-violating, but whether its downstream use occurs in a domain where errors, biases, omissions, or misleading outputs can affect the enjoyment or exercise of fundamental rights (Mantelero 2018; Raji et al. 2020; Selbst 2021).

This contextual inquiry also helps avoid overextension. Not every GPAI-enabled application requires heightened rights analysis. The framework is triggered by plausible rights sensitivity, not by the mere presence of GPAI. Its function is to identify cases in which the downstream context makes the legal adequacy of model-level duties uncertain (Selbst 2021; Mantelero 2022).

3.2 Classifying GPAI-Related Rights Risks

Once a rights-sensitive context has been identified, the second step is to classify the source of the relevant GPAI-related rights risk. This paper distinguishes between three types of risk: model-intrinsic risks, configuration-mediated risks, and use-context-dependent risks. The distinction is analytical. In practice, these categories may overlap. However, the typology clarifies why some risks are more amenable to model-level governance than others (Bommasani et al. 2021; Weidinger et al. 2022; Solaiman et al. 2025).

Model-intrinsic risks arise primarily from the characteristics of the GPAI model itself. These may include biased or stereotyped outputs, hallucinations, privacy leakage, harmful content generation, inadequate robustness, or limitations associated with training data and model architecture (Weidinger et al. 2022; Solaiman et al. 2025). Model-level obligations are most directly relevant to this category. Technical documentation, model evaluation, information about capabilities and limitations, and training content summaries may help downstream providers understand baseline risks before integration (Mitchell et al. 2019; Gebru et al. 2021). In this respect, the AI Act's GPAI obligations support the upstream documentation and governance of such risks (European Commission 2025b).

Configuration-mediated risks arise when downstream actors fine-tune or prompt a GPAI model, connect it to other components, or embed it within a specific AI system (Bommasani et al. 2021; Cobbe, Veale, and Singh 2023). Examples may include retrieval-augmented generation

systems used in legal or administrative contexts, GPAI-based tools that summarize applications or case files, or interfaces that shape how users interpret model outputs. In these cases, the risk does not stem solely from the upstream model or from the downstream system. It emerges from their interaction (Selbst et al. 2019; Cobbe, Veale, and Singh 2023). A model's known limitations may be amplified, reduced, or transformed by system design choices, including data pipelines, prompts, thresholds, and user interfaces (Bommasani et al. 2021; Raji et al. 2020; Selbst 2021).

Use-context-dependent risks arise primarily from the institutional and social context in which the GPAI-enabled system is deployed (Selbst et al. 2019; Selbst 2021). The same technical output may have very different rights implications depending on whether it is used for brainstorming, customer support, student assessment, public benefits triage, or migration advice. These risks are least amenable to model-level governance alone because they depend on institutional purpose, affected populations, decision consequences, procedural safeguards, and the availability of contestation (Mantelero 2018; Kroll 2021; Mantelero 2022). In such cases, upstream documentation may identify general limitations but cannot by itself determine whether the downstream use is fair, proportionate, contestable, or compatible with fundamental rights (Selbst 2021; Nannini 2024).

This typology supports the central claim of the paper: the more a GPAI-related rights impact depends on configuration and use context, the less likely model-level obligations alone are to provide adequate rights protection (Selbst et al. 2019; Cobbe, Veale, and Singh 2023). Model-level duties are necessary because downstream actors depend on upstream information (Brown 2023). However, they are incomplete because fundamental rights impacts often take concrete form when the model is embedded in a system and used within a particular institutional setting (Selbst 2021; Mantelero 2022).

3.3 Assessing the Usability of Upstream Information

The third step evaluates whether upstream information is usable for downstream accountability. The AI Act relies heavily on information flows across the value chain (European Commission 2025a). GPAI model providers must prepare and make information and documentation available to downstream providers, enabling them to comply with their own obligations (European Parliament and Council of the European Union 2024). More generally, scholarship and policy analysis on AI supply chains emphasize that accountability depends on how responsibilities and information are allocated among actors with different levels of knowledge, control, and resources (Cobbe, Veale, and Singh 2023; Brown 2023).

However, the existence of documentation does not guarantee accountability (Ananny and Crawford 2018; Kroll 2021). Information must be usable by the actors who need to act on it. This framework evaluates usability according to three criteria: specificity, actionability, and contestability (Mitchell et al. 2019; Gebru et al. 2021; Nannini 2024).

First, upstream information must have sufficient specificity. Specificity concerns the level of detail and contextual relevance of the information provided. Generic descriptions of model capabilities and limitations may be useful for general transparency, but they may not enable downstream actors to identify which limitations are relevant to a particular rights-sensitive integration or to the fundamental rights interests at stake (Mitchell et al. 2019; Gebru et al. 2021). Specificity requires information about known limitations, foreseeable high-caution uses, relevant performance variations, data-related constraints, unsafe or unsuitable use cases, and conditions under which the model may generate unreliable or discriminatory outputs with potential implications for equality, access, procedural fairness, or contestability (Weidinger et al. 2022; Solaiman et al. 2025; European Commission 2025b). Specificity does not require GPAI providers to predict every downstream use. It does require them to provide information that is meaningful for reasonably foreseeable categories of use, particularly where the model is marketed, documented, or technically suited for integration into rights-sensitive systems (Cobbe, Veale, and Singh 2023; Brown 2023).

Second, information must be actionable. Actionability concerns whether downstream actors can translate the information into concrete governance measures. Even specific information may be practically weak if it does not indicate what downstream providers or deployers should do in response (Raji et al. 2020; Selbst 2021). For present purposes, a limitation disclosed in a model card, technical report, or provider documentation is rights-protective only if it can inform system design, risk assessment, domain-specific testing, user instructions, logging arrangements, human oversight, or deployment limits (Mitchell et al. 2019; Raji et al. 2020). For example, information about hallucination risk may support additional verification mechanisms; information about uneven performance across languages may support domain-specific testing; information about inappropriate use cases may support contractual restrictions or deployment limits. Actionability therefore turns documentation from a disclosure artifact into an input for downstream risk governance (Selbst 2021; Nannini 2024).

Third, information must support contestability. Fundamental rights protection is not exhausted by risk prevention (Mantelero 2018; Kroll 2021). Affected persons, regulators, courts, and oversight bodies may need to understand how a GPAI-enabled system contributed to a decision, recommendation, classification, or denial of

access (Edwards and Veale 2017; Kroll 2021). Contestability requires information that can help identify whether a problem arose from the upstream model, downstream configuration, institutional misuse, insufficient human oversight, or a combination of these factors (Selbst 2021; Cobbe, Veale, and Singh 2023). Transparency and explainability scholarship has increasingly emphasized that disclosure is most meaningful when it supports oversight, challenge, and accountability, rather than merely producing formal compliance artifacts (Ananny and Crawford 2018; Wachter, Mittelstadt, and Floridi 2017; Nannini 2024).

These three criteria recast transparency as a means of supporting accountability (Ananny and Crawford 2018; Kroll 2021). The question is not simply whether GPAI providers have made information available, but whether the disclosed information can travel across the value chain and remain useful in the downstream contexts where rights impacts materialize (Cobbe, Veale, and Singh 2023; Nannini 2024).

3.4 Diagnosing the Contextual Accountability Gap

The final step is to determine whether a contextual accountability gap exists. Such a gap arises when plausible fundamental rights impacts materialize in downstream, context-dependent uses, while the relevant legal duties are located upstream, remain generic, or are insufficiently connected to the conditions of use (Selbst et al. 2019; Selbst 2021; Cobbe, Veale, and Singh 2023).

The framework identifies six indicators of a contextual accountability gap: (1) the GPAI-enabled system is used in a rights-sensitive context; (2) the system does not clearly trigger the high-risk regime; (3) the relevant rights risk depends significantly on downstream configuration, institutional deployment, or user interaction; (4) upstream GPAI documentation is insufficiently specific or actionable for the downstream use; (5) affected persons, regulators, or oversight bodies cannot meaningfully contest the role of GPAI in the relevant decision or interaction; and (6) no equivalent contextual review, such as a fundamental rights-oriented assessment, is conducted by the downstream actor.

These indicators do not function as a formal legal test. They provide an analytical tool for identifying when the AI Act's value chain structure may fail to connect model-level obligations to downstream rights protection (Cobbe, Veale, and Singh 2023). The resulting gap arises from a misalignment between the level at which legal duties attach and the level at which rights impacts become concrete and observable.

This diagnosis is particularly important for GPAI-enabled systems outside the high-risk regime. Article 27's FRIA obligation applies to certain deployers of high-risk AI systems, and the obligation is designed to connect system use with potential effects on affected persons (European

Parliament and Council of the European Union 2024). But where GPAI-enabled systems do not fall within that regime, a similar contextual assessment may not be legally triggered, even though the system may influence access to information, opportunities, public services, or institutional decisions (Mantelero 2018; Selbst 2021; Mantelero 2022).

The framework therefore explains a structural tension in the AI Act. The GPAI regime addresses upstream model governance, while fundamental rights protection often requires downstream contextual accountability. The obligations for high-risk AI systems partially resolve this tension where they apply, but the tension persists where GPAI-enabled systems are rights-sensitive without being legally classified as high-risk (Veale and Zuiderveen Borgesius 2021; European Parliament and Council of the European Union 2024, 2026).

4 Applying the Framework: Illustrative Scenarios

This section applies the framework to GPAI-enabled recruitment support and public-service access chatbots. It examines whether model-level information remains sufficiently specific and practical to support downstream testing, oversight, explanation, contestation, and remedy. This question matters because rights-relevant effects often arise through sociotechnical pathways that are not captured by model properties or legal classification alone (Selbst et al. 2019; Mantelero 2022).

4.1 GPAI-Enabled Hiring Support

Consider a GPAI-enabled tool used by an employer or recruitment platform to assist with hiring. The system does not automatically reject candidates. Instead, it summarizes CVs, ranks applicants according to job descriptions, generates interview questions, and flags perceived inconsistencies in applications. The downstream provider integrates a GPAI model through an API, adds a retrieval layer containing job criteria and employer preferences, and designs a dashboard through which recruiters review candidates. The deployer describes the system as decision support rather than automated hiring.

This scenario primarily illustrates a configuration-mediated risk, although it may also involve model-intrinsic risks. At the model level, the GPAI provider may disclose general information about output reliability, hallucination, bias, language performance, evaluation conditions, or unsuitable uses (Weidinger et al. 2022; Solaiman et al. 2025; European Commission 2025b). If the model qualifies as a GPAI model with systemic risk, Article 55 also requires additional model evaluation, systemic risk assessment and mitigation, incident reporting, and cybersecurity measures. These obligations may improve the upstream provider's

understanding of model behaviour and failure modes. However, they remain primarily model-level obligations. They do not by themselves determine how the model will be used in a specific recruitment workflow.

The rights risk changes when the model is configured downstream. A general limitation disclosed by the model provider becomes consequential through prompts, ranking criteria, retrieval data, interface design, and recruiter workflow design. A model may be documented as capable of summarization, but the downstream system may use summaries as the first or primary representation of a candidate. A model may be documented as having uneven performance across languages or writing styles, but the downstream system may not test whether this affects candidates with non-standard employment histories, migrant backgrounds, disability-related career gaps, or non-native language use. The risk therefore does not arise solely from the model. It emerges from the interaction between model properties and downstream system design (Selbst et al. 2019; Raji et al. 2020; Cobbe, Veale, and Singh 2023).

Where the system is clearly used for recruitment or candidate evaluation, the AI Act's high-risk regime may narrow the accountability gap. Employment and workers' management are included in Annex III's high-risk areas, and systems intended for recruitment or selection, including filtering applications or evaluating candidates, may trigger system-level obligations under the Act. In such cases, risk management, technical documentation, logging, transparency to deployers, human oversight, and deployer obligations can provide a stronger downstream governance structure. The high-risk regime can therefore function as a translation mechanism: it can require the downstream provider to turn upstream model information into system-level risk controls and require the deployer to use the system under specified conditions.

The framework then asks whether upstream information is specific, actionable, and contestable in this context. Specificity requires more than a general warning that the model may hallucinate or produce biased outputs. It requires information that helps downstream actors understand which limitations are relevant to recruitment-related tasks, such as summarizing career histories, comparing applicants, generating interview questions, or interpreting non-standard credentials (Mitchell et al. 2019; Gebru et al. 2021). Actionability requires that the downstream provider can translate those limitations into concrete measures, such as domain-specific validation, restrictions on automated ranking, recruiter instructions, logging of generated summaries, and testing across relevant applicant groups (Raji et al. 2020; Selbst 2021). Contestability requires information that enables candidates or oversight bodies to determine whether a GPAI-enabled system shaped a summary, ranking, or recommendation and to challenge

inaccurate or misleading representations (Kroll 2021; Nannini 2024).

As discussed above, the gap is narrower where the high-risk regime clearly applies, but it is not eliminated. First, the protection depends on whether upstream information is sufficiently specific to the hiring context and whether downstream providers actually translate it into system-level testing and safeguards. Second, candidate-facing accountability remains uncertain if applicants cannot know that a GPAI-enabled system shaped a summary, ranking, or interview prompt, or cannot challenge an inaccurate or misleading representation of their application.

The contextual accountability gap therefore appears not because the AI Act contains no system-level safeguards for hiring. It appears where model-level information fails to connect with system-level and candidate-facing accountability. In this scenario, the framework highlights the need to examine the chain connecting model documentation, system configuration, deployer reliance, and candidate contestability.

4.2 Public-Service Access Chatbots

A second scenario involves a public authority deploying a GPAI-enabled chatbot to help individuals navigate welfare benefits, housing support, healthcare entitlements, immigration procedures, or other public services. The chatbot answers questions, summarizes rules, guides users through forms, and suggests next steps. It does not formally determine eligibility or issue a binding decision. It is presented as an informational tool intended to improve access and reduce administrative burden.

At the model layer, GPAI documentation may again provide important inputs. A model provider may disclose that the model can produce inaccurate factual claims, hallucinate legal or policy details, perform unevenly across languages, or fail in domain-specific contexts without grounding and verification. This information is relevant because public-service settings often involve dependence on institutional information, unequal knowledge, and limited exit options (Eubanks 2018). A user who receives incomplete or misleading guidance may miss a deadline, fail to apply, misunderstand eligibility, omit relevant information, or not pursue an appeal.

This scenario primarily illustrates a use-context-dependent risk, although it may also involve model-intrinsic and configuration-mediated elements. The same GPAI limitation has different consequences depending on the institutional setting. A hallucinated answer in a casual consumer chatbot may be inconvenient; a hallucinated answer about welfare eligibility, housing rights, or immigration procedures may affect access to essential services or legal protection. The risk therefore depends on how the chatbot is grounded, whether information is verified,

whether users can escalate to human staff, whether the system is accessible to non-native speakers or disabled users, and whether erroneous guidance can be corrected.

Here, system-level and deployer-level obligations may not always provide the same connective force as in recruitment. Some AI systems used to determine access to essential public services may fall within high-risk categories. If the chatbot is integrated into eligibility assessment, triage, or decision-making and is classified as high-risk, system-level obligations require risk management, logging, oversight, and documentation. And in such cases, the FRIA obligation may require contextual analysis of the system's impact on affected persons. But if the chatbot is framed as general information, communication support, or user guidance, its legal classification may be less clear. The rights-relevant effect may arise before any formal decision: in whether the person understands a right, submits an application, seeks review, or trusts the information provided.

The relevant high-risk obligations may not apply. In that setting, the most directly applicable AI Act obligation may be Article 50(1), which requires the provider to ensure that users of an AI system intended to interact directly with natural persons are informed that they are interacting with an AI system, unless this is obvious from the circumstances. Yet this form of transparency is thinner than the high-risk regime. It tells the user that AI is involved, but it does not by itself require the public authority to verify the legal accuracy of outputs, provide human escalation, assess rights impacts, or create correction and remedy pathways. This is where the contextual accountability gap widens. The GPAI provider's documentation may say that the model should not be relied upon for authoritative legal or administrative advice. Yet that information would still need to be translated by the public authority into verified knowledge bases, restricted response design, escalation protocols, accessibility safeguards, audit logs, and correction pathways if it is to support rights protection. If this translation does not occur, the existence of upstream documentation provides little protection to the affected person. A notice that the user is interacting with AI may satisfy Article 50's interaction transparency requirement, but it does not allow the user to know whether the answer is legally accurate, whether exceptions apply, or how to challenge misleading guidance (Ananny and Crawford 2018).

The chatbot scenario therefore shows a different pattern from recruitment. The problem is that informational or auxiliary uses of GPAI may materially shape the exercise of rights while falling outside the high-risk regime's system-level safeguards. The contextual accountability gap appears when model-level warnings remain upstream, while downstream institutions fail to convert them into public-service accountability. The relevant governance question is whether public authorities using GPAI have enough usable information and sufficient obligations to turn model

limitations into safeguards that support access, explanation, and remedy.

4.3 Comparative Lessons from the Scenarios

The scenarios show that the key issue is the conditional and uneven connection between model-level GPAI duties and downstream rights accountability. The AI Act's model/system/deployment architecture can provide meaningful safeguards, but its rights-protective force depends on whether model-level information remains specific, actionable, and contestable as it moves into downstream systems and institutional practices.

First, where high-risk system duties clearly apply, the contextual accountability gap is narrower but not eliminated. Hiring illustrates this point. If a GPAI-enabled system clearly falls within Annex III, the AI Act provides stronger mechanisms for risk management, documentation, human oversight, instructions for use, and, in certain cases, a FRIA. These mechanisms can help translate model-level information into downstream safeguards. However, their protective effect still depends on the quality of that translation: whether upstream information is specific to the downstream use, whether downstream providers integrate it into system-level testing and design, whether deployers understand and follow the relevant limits, and whether affected persons can contest GPAI-mediated outcomes.

Second, where GPAI use is auxiliary or informational, the gap widens. Public-service chatbots illustrate this dynamic. A GPAI-enabled system may not make a final decision. It may summarize, explain, route, draft, flag, or recommend. These functions can seem secondary, but they may shape whether a person applies for a benefit. In such cases, model-level documentation may disclose limitations, but there may be no clear downstream mechanism that requires those limitations to be translated into rights-facing safeguards.

Third, the central governance problem is translation across layers. GPAI model providers supply model-level information. Downstream providers configure systems, interfaces, prompts, retrieval layers, and logging practices. Deployers embed those systems into organizational routines and decision environments. Affected persons need explanations, challenge mechanisms, and remedies. The framework asks whether this chain works. A contextual accountability gap arises where information is disclosed upstream but fails to become actionable, contestable, or remedial downstream.

These lessons refine the paper's central claim: GPAI model-level duties are an important foundation for downstream governance, but they do not themselves guarantee fundamental rights accountability. What matters is whether model-level information is translated into system-level safeguards, deployment oversight, and mechanisms for explanation, challenge, and remedy.

5 Implementation Implications: From Model Documentation to Contextual Safeguards

This section develops the implementation implications of the Contextual Accountability Gap Framework by focusing on translation across layers. The aim is to identify what implementation should require at each point in the value chain so that model-level information becomes specific, actionable, and contestable in downstream rights-sensitive contexts.

5.1 Context-Aware GPAI Documentation

The first governance implication concerns the content and purpose of GPAI documentation. The AI Act already places documentation and information-sharing duties on GPAI model providers. These duties matter because downstream actors cannot assess or mitigate risks they cannot see. Model cards, datasheets, transparency reports, and broader ecosystem documentation similarly aim to reduce information asymmetries among model developers, downstream providers and deployers, regulators, and affected communities (Mitchell et al. 2019; Gebru et al. 2021). However, the scenarios above show that documentation is rights-protective only if it can be used downstream.

Context-aware documentation would therefore evaluate model information by its downstream usability. It would not require GPAI providers to predict every possible use or to guarantee the safety of all downstream systems. Rather, it would require them to provide information that is useful for reasonably foreseeable rights-sensitive integrations. For example, a general statement that a model may hallucinate is useful but incomplete. In a public-service access chatbot, the relevant question is whether the model can reliably provide grounded information about eligibility, deadlines, and procedural steps.

In recruitment, the relevant question is whether the model performs unevenly across different CV formats, languages, career trajectories, or employment histories associated with protected characteristics or other socially salient circumstances. In education, the relevant question is whether summaries or feedback differ across dialects, writing styles, or student populations. Context-aware documentation should not resolve these domain-specific questions by itself. But it should make clear where downstream testing and safeguards are necessary.

There are limits. GPAI providers cannot know all downstream contexts, and excessive documentation demands may create burdens without improving rights protection. The appropriate standard should therefore be reasonable foreseeability. If a GPAI model is designed, marketed, technically suited, or widely used for summarization, classification, ranking, advice generation,

educational feedback, or public-facing chatbots, then certain high-caution uses are not wholly unforeseeable. In those cases, model providers can reasonably be expected to identify limitations and safeguards relevant to downstream rights-sensitive integration.

5.2 Translation Records for Downstream Providers

The second implication concerns downstream providers that integrate GPAI models into AI systems. The scenarios in Section 4 show that rights risks often change as the model is configured. A general model limitation becomes a recruitment risk when summaries shape candidate evaluation. A hallucination risk becomes an access-to-services risk when a chatbot provides public-service guidance. These risks are not fully model-intrinsic; they are configuration-mediated or use-context-dependent.

For this reason, downstream providers should treat GPAI documentation as an input into system-level governance rather than as evidence that risks arising from downstream integration have already been addressed. A practical way to implement this connection is a translation record. A translation record would explain how upstream model information was used when designing, testing, and deploying the downstream system. It would make explicit the link between GPAI documentation and system-level governance.

For rights-sensitive GPAI-enabled systems, such a record could include: (1) which GPAI model was used and for which system function; (2) which upstream model limitations were considered relevant; (3) how the model was configured and integrated, including prompts, retrieval systems, fine-tuning, and interfaces; (4) what domain-specific validation or testing was performed; (5) what safeguards were adopted in response to known limitations; (6) what information was communicated to deployers; (7) what human oversight arrangements were established; and (8) what information is retained to support contestability by affected persons.

This proposal builds on the insight that documentation should support communication between actors, not merely record internal development choices (Mitchell et al. 2019; Gebru et al. 2021). It also aligns with sociotechnical accounts of algorithmic accountability, which emphasize that harms often arise through institutional integration rather than from technical artifacts alone (Selbst et al. 2019). A translation record would help regulators, deployers, auditors, and in some cases affected persons understand whether the system provider treated model-level information as a meaningful governance input.

The burden should be proportionate. Lower-risk productivity uses should not require extensive translation records. But where GPAI materially contributes to rights-

sensitive assessment, access, classification, allocation, or remedy, some record of translation is justified. The record need not disclose trade secrets or full model internals. Its purpose is to show how known model limitations were addressed in the downstream system and how responsibility is distributed between the GPAI model provider, downstream AI system provider, and deployer.

5.3 FRIA-Like Contextual Review for Rights-Sensitive Deployment

The third implication concerns deployers. Article 27 requires certain deployers of high-risk AI systems to conduct a FRIA before putting the system into use. This is one of the AI Act's most important mechanisms for connecting AI governance to concrete effects on affected persons. It requires attention to the deployer's context, affected groups, foreseeable risks, human oversight, and mitigation measures.

However, as the scenarios show, some GPAI-enabled systems may affect rights through auxiliary, informational, or support functions that do not always clearly trigger dense high-risk obligations. This does not mean that FRIA obligations should be extended wholesale to every GPAI-enabled system. It means that some rights-sensitive deployments require contextual review even when formal classification is uncertain. The appropriate trigger should be contextual and proportionate. A FRIA-like contextual review should be considered where three conditions are present: the deployment occurs in a rights-sensitive context; the GPAI-related risk is configuration-mediated or use-context-dependent; and model-level information is unlikely to become specific, actionable, or contestable without additional institutional review.

Such a review would ask different questions from model documentation. It would ask who is affected, what interests or rights are at stake, how GPAI outputs influence institutional judgment or user behavior, whether affected persons know about the system's role, what human oversight exists, whether alternatives are available, how errors can be challenged, and how harms can be remedied. These questions are deployment-specific and cannot be answered at the model layer alone.

Impact assessment literature supports this approach. Human rights and social impact assessment frameworks emphasize that rights risks must be evaluated in relation to affected persons, institutional context, power asymmetries, and available remedies (Mantelero 2018; Mantelero 2022). Institutional accounts of algorithmic impact assessment similarly show that AI governance must address organizational use, not merely technical design (Selbst 2021). The GPAI context strengthens this point because model-level information often requires institutional translation before it can become rights-protective.

The purpose of FRIA-like review is to ensure that rights-sensitive deployments receive contextual scrutiny when formal classification does not fully capture the pathway of impact. For example, a public-service chatbot may be framed as informational rather than decisional, but it may still shape whether a person applies for a benefit or understands an appeal right. The key question is not only whether the system falls within a formal category, but whether GPAI materially shapes the conditions under which rights are exercised.

A deployer-level contextual review also helps address a knowledge problem. GPAI providers know most about the model; downstream providers know most about the system; deployers know most about the institutional context and affected populations. No single actor has full visibility. A FRIA-like review can help the deployer to connect upstream documentation, downstream system design, local institutional practices, and affected-person remedies. It therefore plays a crucial role in repairing the connective failure identified in this paper.

6 Conclusion

This paper examined a value chain problem in GPAI governance under the EU AI Act: model-level duties are imposed upstream, while many fundamental rights impacts become concrete only after a GPAI model is integrated into downstream systems and used in institutional contexts. The paper argued that GPAI model-level obligations are necessary but incomplete for downstream rights protection. Their effectiveness depends on whether upstream information remains usable across the value chain.

To analyze this problem, the paper developed the Contextual Accountability Gap Framework. The framework distinguishes between model-intrinsic, configuration-mediated, and use-context-dependent risks, and assesses whether upstream information is specific, actionable, and contestable in downstream contexts. Applied to hiring support and public-service chatbots, the framework shows that the key governance problem is not simply disclosure, but translation: whether model-level information can become system-level safeguards, deployment oversight, and mechanisms for challenge or remedy.

The paper's broader contribution is to clarify that fundamental rights protection in GPAI governance depends on the connection between layers of responsibility. Context-aware documentation, downstream translation records, and FRIA-like contextual review can help strengthen fundamental rights accountability.

References

- Ananny, M.; and Crawford, K. 2018. Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability. *New Media & Society* 20(3): 973–989. doi.org/10.1177/1461444816676645.
- Barocas, S.; Hardt, M.; and Narayanan, A. 2023. *Fairness and Machine Learning: Limitations and Opportunities*. Cambridge, MA: The MIT Press.
- Bommasani, R.; Hudson, D. A.; Adeli, E.; Altman, R.; Arora, S.; von Arx, S.; Bernstein, M. S.; Bohg, J.; Bosselut, A.; Brunskill, E.; et al. 2021. On the Opportunities and Risks of Foundation Models. arXiv:2108.07258.
- Brown, I. 2023. Allocating Accountability in AI Supply Chains. Ada Lovelace Institute. <https://www.adalovelaceinstitute.org/resource/ai-supply-chains/>. Accessed: 2026-05-16.
- Cobbe, J.; Veale, M.; and Singh, J. 2023. Understanding Accountability in Algorithmic Supply Chains. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 1186–1197. New York: Association for Computing Machinery. doi.org/10.1145/3593013.3594073.
- Edwards, L.; and Veale, M. 2017. Slave to the Algorithm? Why a ‘Right to an Explanation’ Is Probably Not the Remedy You Are Looking For. *Duke Law & Technology Review* 16(1): 18–84.
- Eubanks, V. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin’s Press.
- European Commission. 2025a. General-Purpose AI Models in the AI Act – Questions & Answers. <https://digital-strategy.ec.europa.eu/en/faqs/general-purpose-ai-models-ai-act-questions-answers>. Accessed: 2026-05-03.
- European Commission. 2025b. Guidelines on the Scope of Obligations for Providers of General-Purpose AI Models under the AI Act. <https://digital-strategy.ec.europa.eu/en/library/guidelines-scope-obligations-providers-general-purpose-ai-models-under-ai-act>. Accessed: 2026-05-03.
- European Parliament and Council of the European Union. 2024. Regulation (EU) 2024/1689 of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act). <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>.
- European Parliament and Council of the European Union. 2026. Regulation (EU) 2026/1744 of 8 July 2026 Amending Regulations (EU) 2024/1689, (EU) 2018/1139 and (EU) 2023/1230 as Regards the Simplification of the Implementation of Harmonised Rules on Artificial Intelligence (Digital Omnibus on AI). <https://eur-lex.europa.eu/eli/reg/2026/1744/oj/eng>.
- Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J. W.; Wallach, H.; Daumé III, H.; and Crawford, K. 2021. Datasheets for Datasets. *Communications of the ACM* 64(12): 86–92. doi.org/10.1145/3458723.
- Kroll, J. A. 2021. Outlining Traceability: A Principle for Operationalizing Accountability in Computing Systems. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 758–771. New York: Association for Computing Machinery. doi.org/10.1145/3442188.3445937.
- Mantelero, A. 2018. AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment. *Computer Law & Security Review* 34(4): 754–772. doi.org/10.1016/j.clsr.2018.05.017.
- Mantelero, A. 2022. *Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI*. The Hague: T.M.C. Asser Press. doi.org/10.1007/978-94-6265-531-7.
- Mitchell, M.; Wu, S.; Zaldivar, A.; Barnes, P.; Vasserman, L.; Hutchinson, B.; Spitzer, E.; Raji, I. D.; and Gebru, T. 2019. Model Cards for Model Reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229. New York: Association for Computing Machinery. doi.org/10.1145/3287560.3287596.
- Nannini, L. 2024. Habemus a Right to an Explanation: So What? – A Framework on Transparency-Explainability Functionality and Tensions in the EU AI Act. In *Proceedings of the Seventh AAAI/ACM Conference on AI, Ethics, and Society*, 1023–1035. Washington, DC: AAAI Press. doi.org/10.1609/aies.v7i1.31700.
- Raji, I. D.; Smart, A.; White, R. N.; Mitchell, M.; Gebru, T.; Hutchinson, B.; Smith-Loud, J.; Theron, D.; and Barnes, P. 2020. Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 33–44. New York: Association for Computing Machinery. doi.org/10.1145/3351095.3372873.
- Selbst, A. D. 2021. An Institutional View of Algorithmic Impact Assessments. *Harvard Journal of Law & Technology* 35(1): 117–191.
- Selbst, A. D.; Boyd, D.; Friedler, S. A.; Venkatasubramanian, S.; and Vertesi, J. 2019. Fairness and Abstraction in Sociotechnical Systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 59–68. New York: Association for Computing Machinery. doi.org/10.1145/3287560.3287598.
- Solaiman, I.; Talat, Z.; Agnew, W.; Ahmad, L.; Baker, D. K.; Blodgett, S. L.; Chen, C.; Daumé III, H.; Dodge, J.; Duan, I.; Evans, E.; Friedrich, F.; Ghosh, A.; Gohar, U.; Hooker, S.; Jernite, Y.; Kalluri, P. R.; Leidinger, A.; Lusoli, A.; Lin, M.; Lin, X.; Luccioni, S.; Mickel, J.; Mitchell, M.; Newman, J.; Ovalle, A.; Png, M.-T.; Singh, S.; Strait, A.; Struppek, L.; Subramonian, A.; and Vassilev, A. 2025. Evaluating the Social Impact of Generative AI Systems. In *The Oxford Handbook of the Foundations and Regulation of Generative AI*, edited by P. Hacker, A. Engel, S. Hammer, and B. Mittelstadt. Oxford: Oxford University Press. doi.org/10.1093/oxfordhdb/9780198940272.013.0025.
- Veale, M.; and Zuiderveen Borgesius, F. 2021. Demystifying the Draft EU Artificial Intelligence Act—Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach. *Computer Law Review International* 22(4): 97–112. doi.org/10.9785/cr-2021-220402.
- 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. *International Data Privacy Law* 7(2): 76–99. doi.org/10.1093/idpl/ix005.

Weidinger, L.; Uesato, J.; Rauh, M.; Griffin, C.; Huang, P.-S.; Mellor, J.; Glaese, A.; Cheng, M.; Balle, B.; Kasirzadeh, A.; Biles, C.; Brown, S.; Kenton, Z.; Hawkins, W.; Stepleton, T.; Birhane, A.; Hendricks, L. A.; Rimell, L.; Isaac, W.; Haas, J.; Legassick, S.; Irving, G.; and Gabriel, I. 2022. Taxonomy of Risks Posed by Language Models. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 214–229. New York: Association for Computing Machinery. doi.org/10.1145/3531146.3533088.

Williams, S.; Schuett, J.; and Anderljung, M. 2026. On Regulating Downstream AI Developers. *European Journal of Risk Regulation* 17(1): 94–122. doi.org/10.1017/err.2025.10020.