



UNIVERSITETET
I OSLO

When is an AI system high-risk?

Operationalising the “no significant risk” carve-outs of the AI Act’s Article 6(3)(a) and (d)

Candidate number: 576

Delivery deadline: 15 May 2026

Word count: 17 432

Table of contents

1	INTRODUCTION.....	1
1.1	Background and motivation	1
1.2	The Article 6(3) proportionality “safety valve”	2
1.3	Research question	3
1.4	Methodology and sources	4
1.5	Thesis structure	6
2	THE AI ACT AND THE LEGAL CONTEXT OF ARTICLE 6(3).....	7
2.1	The AI Act’s risk-based regulatory system.....	7
2.2	High-risk AI systems under the AI Act	7
2.3	The function of Article 6(3) in the AI Act	9
2.3.1	Article 6(3) as a proportionality “safety valve”.....	9
2.3.2	The relationship between Article 6(2) and Article 6(3)	10
2.3.3	The Article 6(3) conditions are centred around “tasks”	10
2.3.4	The relationship between Article 6(3) and Article 14	11
2.3.5	The task of assessing and documenting whether an AI system qualifies for the Article 6(3) carve-out rests with the provider of the system	11
2.3.6	The provider/deployer divergence under Article 6(3).....	13
3	LEGAL INTERPRETATION OF ARTICLE 6(3)(A) AND (D).....	14
3.1	Structure and wording of Article 6(3).....	14
3.1.1	The overarching condition of no significant risk of harm	14
3.1.2	Article 6(3) as a closed list of carve-outs	14
3.1.3	The “profiling of natural persons” carve-out from the carve-outs	16
3.2	“Narrow procedural task” under Article 6(3)(a)	16
3.2.1	Linguistic interpretation of “narrow”, “procedural”, and “task”	16
3.2.2	Systematic interpretation of “narrow procedural task”	20
3.2.3	Teleological interpretation: the purpose of carving out procedural tasks within high-risk activities	21
3.3	“Preparatory task to an assessment” under Article 6(3)(d).....	22
3.3.1	Linguistic interpretation of “preparatory” and “to an assessment”	22
3.3.2	Systematic interpretation of “preparatory task to an assessment”	24
3.3.3	Teleological interpretation: the purpose and limits of excluding preparatory tasks from high-risk classification.....	25
3.4	The distinction between “procedural” tasks and “preparatory” tasks.....	26
3.5	“Profiling of natural persons” under Article 6(3) third subparagraph	27

4	OPERATIONALISING THE CRITERIA: A FRAMEWORK FOR APPLYING ARTICLE 6(3)(A) AND (D)	29
4.1	From interpretation to operationalisation.....	29
4.2	The proposed criteria	29
4.2.1	Criterion 1: The profiling threshold – does the task involve profiling of natural persons?	29
4.2.2	Criterion 2: Task delimitation – is the task specific, defined, and bounded?... 30	
4.2.3	Criterion 3: The nature of the operation – syntactic or semantic?.....	31
4.2.4	Criterion 4: Functional role – does the output bear on the substance of an assessment or decision?	32
4.2.5	Criterion 5: Transparency – is the output transparent and verifiable?	33
4.2.6	Criterion 6: Automation bias resilience – is the output inherently resistant to uncritical reliance?	34
4.3	Application to illustrative cases	35
4.4	On the interrelation of criteria 3 to 6.....	36
5	CONCLUSION	38
5.1	Answer to the research question	38
5.2	Summary of findings and contributions.....	38
5.3	<i>De lege ferenda</i> : improving the legislative framework	39
5.4	Residual uncertainties and directions for future research	41
	TABLE OF REFERENCES	44
	Literature	44
	Legal sources.....	47
	Legislation and treaties	47
	Case law	48
	Institutional documents and official guidance	48
	Websites and online materials.....	50
	APPENDIX: APPLYING THE ARTICLE 6(3)(A) AND (D) CRITERIA TO ILLUSTRATIVE EXAMPLES	52
	Application to illustrative examples.....	52
	Document translation for asylum assessment	52
	Selecting job interview candidates: a multi-stage recruitment workflow.....	53
	Writing a medical journal: a multi-stage clinical documentation workflow	56

1 Introduction

1.1 Background and motivation

With the adoption of the Artificial Intelligence Act (hereafter “AI Act”, “the Act”, or “AIA”) on 13 June 2024, the European Union (EU) became the first jurisdiction to enact a comprehensive horizontal regulatory framework for Artificial Intelligence (AI) systems.¹ It represents the EU’s attempt to strike a balance between the often conflicting objectives of 1) the protection of fundamental rights and public safety, and 2) the promotion of innovation and economic competitiveness, including legal certainty for providers and deployers of AI systems.² This balancing act is central in the provisions governing the classification of high-risk AI systems.

The AI Act adopts a risk-based regulatory approach, under which the intensity of regulatory requirements aims to be proportional to the level of risk that an AI system poses to fundamental rights, health, and safety.³ While few AI uses are outright prohibited, the Act introduces a broad “high-risk” category triggering extensive obligations. While the risk-based architecture of the Act has attracted considerable scholarly attention, most contributions have concentrated on the classification framework at a macro level. Early work has highlighted the practical enforceability challenges that flow from the Act’s layered governance structure, including the potential of dual-use and the ease of disguising the “true” market of an AI system.⁴ Subsequent contributions have examined the conceptual coherence of the Act’s risk tiers and the assumptions underpinning the high-risk category,⁵ the Act’s interplay with existing product-safety regulation,⁶ and the adequacy of its conformity-assessment mechanisms.⁷

Article 6(3) of the Act includes a mechanism to exempt individual AI systems that pose no significant risk from high-risk obligations, even if they operate in high-risk areas. The wording of the mechanism was finalised late in the legislative process,⁸ and has received limited doctrinal scrutiny. This thesis aims to complement the existing literature by examining the AI Act

¹ Regulation (EU) 2024/1689 (AI Act).

² Recital (1) frames the purpose as ensuring “a high level of protection of health, safety, fundamental rights ... while ensuring the free movement of AI-based goods and services cross-border.” On the broader policy tension, see Ebers (2024) 689–697.

³ Recital (26) describes a goal of “proportionate and effective set of binding rules for AI systems”, to be achieved by following “a clearly defined risk-based approach ... tailor[ed to] the type and content of such rules to the intensity and scope of the risks that AI systems can generate”.

⁴ Veale and Borgesius (2021) 100–106.

⁵ Mahler (2022) 248–256.

⁶ Ebers (2024) 695–697.

⁷ Smuha and Yeung (2025) 253–257.

⁸ The first attempt at a high-risk “safety valve” appears in the Council of the European Union (2022) compromise text, with a derogation for systems with “purely accessory” output. This was reworked further in the European

Article 6(3)(a) and (d) carve-outs, asking what it means, in operational terms, for an AI system to perform a “narrow procedural task” or a “preparatory task to an assessment.”⁹

1.2 The Article 6(3) proportionality “safety valve”

Casting the regulatory net widely in sensitive domains ensures that AI systems operating in contexts where fundamental rights are at stake are subject to scrutiny. At the same time, this creates a risk of over-inclusion. Not every AI system deployed within a high-risk use context will, by virtue of its specific function, pose a significant risk of harm. Gradually recognising this, the legislator introduced Article 6(3) as what may be characterised as a proportionality “safety valve”: a mechanism that allows certain AI systems within Annex III domains to not be considered to be high-risk where they do “not pose a significant risk of harm to the health, safety, or fundamental rights of natural persons, including by not materially influencing the outcome of decision-making.”¹⁰ Article 6(3) sets out a closed list of four conditions under which the carve-out may apply, in direct quote:

- (a) the AI system is intended to perform a narrow procedural task;
- (b) the AI system is intended to improve the result of a previously completed human activity;
- (c) the AI system is intended to detect decision-making patterns or deviations from prior decision-making patterns and is not meant to replace or influence the previously completed human assessment, without proper human review; or
- (d) the AI system is intended to perform a preparatory task to an assessment relevant for the purposes of the use cases listed in Annex III,

with the following subparagraph establishing a carve-out from the carve-out in that an AI system in Annex III “shall always be considered to be high-risk where the AI system performs profiling of natural persons.”

Of the four conditions, (b) and (c) share the trait that they describe AI system activity that is independent from and temporally placed after a previously completed human activity or assessment. While (b) and (c) are not free of ambiguity,¹¹ the backwards looking nature and independence from the previous human activity/assessment are concrete signposts for AI providers and deployers assessing if their system qualifies for exemption. On the other hand, the concepts

Parliament (2023) first-reading position, before reaching its final form in Article 6(2a) of the Council of the European Union (2024) proposal based on the trilogue agreement of December 2023.

⁹ See e.g. Couneson (2024) in Pehlivan, Forgó, and Valcke (eds) (2024) for a general commentary on Article 6.

¹⁰ Article 6(3), first subparagraph. See also Recital (53), which elaborates on the rationale for the exception and provides illustrative examples.

¹¹ E.g. at which point does systematic AI result *improvement* under Article 6(3)(b) cross the border into de-facto AI result *replacement* of the earlier human activity result?

of (a) “narrow procedural task” and (d) “preparatory task to an assessment” are open-textured and lack precise legal definition in the operative provisions of the Act. This is the motivation for narrowing the scope of this thesis to Article 6(3) conditions (a) and (d).¹²

This interpretative ambiguity gives rise to a dual risk. On the one hand, risk-seeking actors with an incentive to minimise regulatory costs may strategically exploit the carve-outs by characterising the functions performed by their AI systems in narrow terms, avoiding high-risk classification and compliance obligations. This risk of regulatory arbitrage has been identified as a structural concern in the literature on the AI Act.¹³ Risk-averse actors in heavily regulated sectors may, conversely, avoid the carve-outs altogether, preferring the legal certainty of full high-risk compliance – at the cost of unnecessary compliance burdens, dampened incentives for beneficial AI adoption, and a weakened proportionality principle.¹⁴

This thesis seeks to derive, through systematic legal interpretation, a set of operational criteria for whether an AI system’s task qualifies as “narrow procedural” or “preparatory to an assessment”. The criteria aim to reduce the scope for regulatory arbitrage by risk-seeking actors and to give risk-averse actors the legal certainty needed to invoke the carve-outs where they legitimately apply.¹⁵

1.3 Research question

The open-textured conditions for exemption from high-risk classification require operationalisation if they are to serve their intended function as a proportionate corrective to the breadth of Annex III. This thesis addresses the following research question:

What criteria can be derived from the AI Act and analogies from relevant EU law to operationalise the notions of “narrow procedural task” and “preparatory task to an assessment” in Article 6(3)(a) and (d), and how can these criteria be applied in practice to reduce regulatory arbitrage while enabling legitimate use of the Article 6(3) carve-outs?

The research question contains two analytically distinct but interrelated components. The first is interpretative: it asks what substantive criteria can be derived from the legal sources to give

¹² The Act’s ambiguity is not limited to Article 6(3). See e.g. Smuha and Yeung (2025) 241-242, 247-248 for further discussion on the interpretative challenges arising from the Act’s use of open-textured concepts.

¹³ See Gasiola (2025) for a discussion of structural loopholes, and Veale and Borgesius (2021) 106 on concerns related to risk self-assessment.

¹⁴ Ebers (2024) 689–697 argues that the Act does not follow a truly risk-based approach in all respects and 660–661 that the absence of a risk-benefit analysis may lead to disproportionate regulatory burdens.

¹⁵ The need for interpretative guidance is expressly recognised in Article 6(5), which mandates the Commission to provide guidelines specifying the practical implementation of the Article 6 “no later than 2 February 2026”. At the time of writing, these guidelines have not yet been provided.

operative content to the concepts of “narrow procedural task” under Article 6(3)(a) and “preparatory task to an assessment” under Article 6(3)(d). The second component is applied: it asks how such criteria can be operationalised to guide the practical classification of AI systems, to reduce the scope for regulatory arbitrage by risk-seeking actors while preserving the legal certainty necessary for risk-averse actors to make legitimate use of the carve-outs.

The scope of the inquiry is delimited in several respects. First, the thesis focuses on the conditions in Article 6(3)(a) and (d), rather than on all four alternative conditions listed in Article 6(3). As discussed in the previous section, the conditions of Article 6(3)(b) and (c), concerning AI systems that improve the result of a previously completed human activity and AI systems that detect decision-making patterns, are relatively easy to operationalise through their clear temporal and functional separation from decision processes. Second, the thesis does not undertake an empirical study of how providers and deployers currently intend to apply the Article 6(3) exemptions; the practical examples interspersed throughout the text and detailed in the appendix are illustrative and designed to test the derived criteria, not to make claims about industry practice.¹⁶

By answering this research question, I aim to provide tools that are of value not only for academic discourse, but also for practitioners, including AI providers, deployers, and national market surveillance authorities, who must navigate the Article 6(3) assessments in practice.

1.4 Methodology and sources

For the interpretative part of the research question, the thesis employs legal dogmatic method: the provisions are interpreted through recognised EU-law sources and the ordinary textual, systematic, and teleological techniques. Because all language versions of EU legislation are equally authentic, a full multilingual comparison is relevant, but for reasons of scope, this thesis proceeds from the English text and uses other sources to control its interpretation.¹⁷

The analysis first examines the wording of Article 6(3)(a) and (d), especially “procedural”, “preparatory”, and “task”. It then reads those terms in the context of the Act as a whole, including Annex III, Article 14 on human oversight, Article 3 definitions, and the Act’s risk-based structure. Finally, it considers the purpose of Article 6(3): maintaining fundamental-rights pro-

¹⁶ The extensive use of examples is inspired by the AIA’s own methodology: Recital (53) provides a few brief high-level examples, and Article 6(5) mandates the Commission to provide “a comprehensive list of practical examples of use cases” to illustrate the classification rules.

¹⁷ On the legal dogmatic method in the context of EU law, see Lenaerts and Gutiérrez-Fons (2013). On the ECJ’s methodology for the interpretation of EU secondary legislation, see Case C-283/81 (CILFIT) paragraphs 17-20.

tection while avoiding disproportionate high-risk classification for genuinely low-risk functions. GDPR Article 22 is drawn on by analogy to illuminate the boundary between preparatory processing and decision-influencing automation.¹⁸

The doctrinal analysis is supplemented by the syntactic/semantic distinction from information theory. The distinction is used only as an interpretive aid: it helps clarify whether an AI task operates on the form of information or its evaluative meaning, without displacing the legal-dogmatic analysis.

The thesis is primarily concerned with *lex lata*. Section 5.3 adds *de lege ferenda* observations where the doctrinal analysis reveals structural weaknesses and unfortunate disincentives in the current carve-out framework.

In accordance with settled ECJ case law, the Act's recitals are used as an aid to interpretation of the operative provisions, without being treated as autonomous sources of binding legal obligations.¹⁹ Recital (53) is particularly important because it gives examples of "narrow procedural" and "preparatory" tasks and explains the rationale of the exemption mechanism.

ECJ case law supplies both general interpretative principles and specific analogies. Because the Act has not yet generated relevant jurisprudence, SCHUFA is especially useful: its treatment of GDPR Article 22 shows that a functionally preparatory automated output may still be legally decisive where the subsequent decision strongly draws on it. The Court's approach to derogations also frames Article 6(3): exceptions are interpreted strictly, but not so strictly as to deprive them of *effet utile*.²⁰

Soft law is used only subsidiarily as interpretative aids. Article 6(5) requires Commission guidelines on the practical implementation of Article 6, but their absence at the time of writing leaves this interpretation without an authoritative soft-law benchmark.²¹

¹⁸ Regulation (EU) 2016/679 (GDPR) Article 22(1) provides that "the data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her." Interpretations of "decision" and "profiling" in the context of GDPR Article 22 are relevant for the systematic interpretation of AIA Article 6(3).

¹⁹ See Case C-162/97 (Nilsson) paragraph 54: "the preamble to a Community act has no binding legal force and cannot be relied on either as a ground for derogating from the actual provisions of the act in question."

²⁰ On the requirement to give derogations their *effet utile*, see Case C-201/13 (Deckmyn) paragraph 23. The tension between strict interpretation and *effet utile* is particularly relevant in the context of Article 6(3), which is simultaneously a derogation from the high-risk classification and a mechanism essential to the proportionality of the regulatory framework.

²¹ On the legal status of Commission guidelines in EU law, see Case C-226/11 (Expedia) paragraphs 28–31, establishing that guidelines do not bind national courts but may constitute a relevant point of reference.

1.5 Thesis structure

Chapter 2 situates Article 6(3) within the AI Act's risk-based architecture, the provider/deployer allocation of responsibility, and Article 14 on human oversight. Chapter 3 interprets Article 6(3)(a) and (d). Chapter 4 turns that interpretation into six operational criteria. Chapter 5 answers the research question, summarises the contribution, and identifies points for legislative or supervisory clarification and improvement. The Appendix sets out the criterion-by-criterion application of the framework to three illustrative case studies summarised in Section 4.3.

2 The AI Act and the legal context of Article 6(3)

2.1 The AI Act's risk-based regulatory system

The AI Act establishes a horizontal regulatory framework for AI systems in the EU, applicable across sectors and (almost) irrespective of the technology's underlying architecture.²² The Act's structural lineage is the EU's New Legislative Framework for product safety (NLF).²³ Its central regulatory mechanism is a purportedly risk-based approach under which the intensity of legal obligations imposed on providers and deployers of AI systems is proportional to the level of risk that those systems pose to fundamental rights, health, and safety.²⁴ This approach reflects both the principle of proportionality enshrined in Article 5(4) of the Treaty on European Union (TEU) and the harmonisation objective of the Act's legal basis in TFEU Article 114.²⁵

The Act organises AI systems into five semi-tiered (but not all mutually exclusive) categories of risk and regulatory obligations. Article 5 prohibits unacceptable-risk practices.²⁶ Article 6 defines high-risk systems, the category carrying the most extensive provider and deployer obligations.²⁷ Article 50 assigns transparency obligations to systems with deception risk.²⁸ In a break with the otherwise activity centred risk-based approach, Articles 53 and 55 introduce technology-centred obligations to providers of general-purpose AI models. For systems that fall outside the scope of any of the above categories, the Act introduces no new obligations beyond general Article 4 AI literacy,²⁹ though Article 95 encourages voluntary codes of conduct.³⁰

2.2 High-risk AI systems under the AI Act

Article 6 establishes two distinct pathways through which an AI system may be classified as high-risk. The first pathway, set out in Article 6(1), concerns AI systems that are intended to be used as safety components of products, or that are themselves products, covered by the EU

²² Article 1(1) and Recitals (1)–(8).

²³ See Couneson (2024) 204 for more on the NLF and the concept of “high-risk”.

²⁴ Recital (26). See also Ebers (2024) 689–697 for a critical analysis of the extent to which the AIA follows a truly risk-based approach.

²⁵ TFEU Article 114 provides the legal basis for harmonisation measures aimed at the establishment and functioning of the internal market. Article 5(4) TEU states that “Under the principle of proportionality, the content and form of Union action shall not exceed what is necessary to achieve the objectives of the Treaties”.

²⁶ Article 5(1)(a)–(h).

²⁷ Articles 8–15.

²⁸ Article 50. See also Recitals (132)–(134).

²⁹ Article 4 currently imposes an obligation of providers and deployers to “ensure, to their best extent” sufficient AI literacy, but this obligation is likely to be softened by the digital omnibus on AI. The European Parliament (2026) suggests the obligation should be to “shall take measures to support the improvement of AI literacy”, specifying that “this obligation does not cover any guarantee of a specific level of AI literacy of any individual.”

³⁰ See also Recital (165).

harmonisation legislation listed in Annex I.³¹ This pathway integrates the AI Act into the existing EU product safety legislation by ensuring that AI systems embedded in regulated products such as machinery, medical devices, and motor vehicles are subject to the high-risk requirements where the product itself undergoes third-party assessment.

The second pathway, set out in Article 6(2), is the one of primary relevance for this thesis. It provides that AI systems “referred to in Annex III shall be considered to be high-risk.” Annex III enumerates broad areas of use in which AI systems are presumed to present significant risks to fundamental rights, health, or safety.³² The Annex III classification follows an activity-based logic. Rather than prescribing requirements based on the technical characteristics or capabilities of the AI system itself, the Act identifies areas of use in which AI systems are presumed to pose significant risks by virtue of the context in which they operate. AI systems deployed in domains where fundamental rights are directly at stake – such as recruitment, credit scoring, criminal justice, or migration control – warrant heightened scrutiny regardless of their underlying technical design or surrounding processes. As Recital (52) notes, the Annex III enumeration is placed in an annex rather than in the operative text to enable future amendment via delegated acts as the risk landscape evolves.^{33 34}

Classification as high-risk triggers a comprehensive set of requirements and obligations set out in Articles 8–49. The aggregate compliance burden associated with these obligations makes the boundary between high-risk and not-highrisk classification a matter of considerable economic and strategic significance for providers and deployers.

This activity-based classification approach has significant strengths. It provides a degree of technology-neutrality that allows the framework to adapt to technological change; it aligns with the EU’s established tradition of sector-specific safety regulation; and it allows the same underlying technology to be subject to different regulatory treatment depending on the context of its deployment. An AI model used for entertainment recommendations may face no mandatory obligations, while the same model repurposed for credit scoring falls within Annex III’s high-risk category.

There is, however, a risk of over-inclusion. The Annex III categories are drawn broadly enough to capture AI systems that, while deployed within a high-risk use context, perform functions

³¹ Article 6(1)(a) and (b). Annex I lists the EU harmonisation legislation to which this provision applies.

³² Annex III (1)–(8). Each area specifies particular use cases within the broader category. See Recitals (48)–(61) for the rationale underlying the selection of these areas.

³³ The use of delegated acts for amending Annex III is provided for in Article 7(1) and (2), subject to conditions specified therein. Most notably, Article 7(1)(a) specifies that only use-cases *within* the existing areas in Annex III can be added or modified by delegated acts.

³⁴ See Couneson (2024) 204–208 for more details on the two pathways to high-risk classification.

that do not, in and of themselves, pose a significant risk of harm – an AI system performing automated document sorting in a recruitment workflow falls within Annex III area 4 (employment), but presents a fundamentally different risk profile from one that autonomously screens and ranks candidates. The risk-tier architecture operates primarily as a tool of legislative proportionality rather than as a framework for individualised, empirical risk management.³⁵

Article 6(3) addresses this structural over-inclusion. Its legislative history shows an attempt to add a horizontal corrective to Annex III, so that systems unlikely to cause serious rights or safety risks are not captured merely because they operate in a sensitive domain.³⁶ In a framework that classifies risk primarily by use-area rather than by individualised empirical assessment, Article 6(3) is the principal mechanism for restoring proportionality at the level of the specific AI system. It gives providers a limited route to rebut the Annex III high-risk presumption where the system’s actual task is too remote from the protected interests to justify high-risk treatment.

2.3 The function of Article 6(3) in the AI Act

2.3.1 Article 6(3) as a proportionality “safety valve”

Article 6(3) establishes that an AI system falling within one of the Annex III areas “shall not be considered to be high-risk where it does not pose a significant risk of harm to the health, safety or fundamental rights of natural persons, including by not materially influencing the outcome of decision making”.³⁷

The provision creates a mechanism through which AI systems that fall within the broad sweep of Annex III can, under certain conditions, be exempted from the full range of high-risk obligations. In this sense, Article 6(3) functions as a proportionality “safety valve” that relieves the regulatory pressure on systems that, despite operating in a high-risk domain, do not themselves present a significant risk of harm.³⁸ Without a corrective mechanism, the framework would impose the full weight of high-risk compliance on AI systems performing genuinely low-risk functions, in tension with the proportionality principle articulated in Recital (26) and anchored in TEU Article 5(4). Recital (53) elaborates on this rationale, noting that “there may be specific cases in which AI systems referred to in [Annex III] do not lead to a significant risk of harm to

³⁵ Mahler (2022) discusses this in depth, see esp. 247–248, 265–267. Ebers (2024) 686–689 stresses the hidden underuse costs by the absence of risk-benefit analysis.

³⁶ Council of the European Union (2022) IV.1.5.

³⁷ Article 6(3) first subparagraph.

³⁸ Couneson (2024) 214 explains the Article 6(3) mechanism was referred to as the “filter” during the negotiations. I think the “safety valve” label better captures the regulatory-pressure-relieving purpose of the derogation.

the legal interests protected under those areas because they do not materially influence the decision-making or do not harm those interests substantially”.³⁹

2.3.2 The relationship between Article 6(2) and Article 6(3)

The relationship between Article 6(2) and 6(3) is best understood as a two-step classification process. First Article 6(2) establishes a rebuttable presumption from the legislator: all AI systems falling within the Annex III categories are presumed to be high-risk. Next, Article 6(3) provides the mechanism through which this presumption may be rebutted by the provider who would bear the burden of compliance, subject to the satisfaction of specific conditions, through an individual case-by-case risk assessment. The starting point is one of classification as high-risk; the Article 6(3) carve-out operates as an exception to this default, with the burden of proof placed with the provider. Since the risk designation is tied to the provider-supplied intended purpose, a deployer cannot use Article 6(3) to escape Article 26 and 27 deployer high risk obligations, even if the deployer’s specific intended use of the system is low risk.⁴⁰

2.3.3 The Article 6(3) conditions are centred around “tasks”

The four conditions of Article 6(3) do not refer to the AI *system* as such, but to the *task* it is intended to perform.⁴¹ Condition (a) speaks of a “narrow procedural task”, condition (d) of a “preparatory task to an assessment.” An AI system may have broad technical capabilities, for example a general large language model (LLM)-based AI assistant may be capable of performing tasks ranging from software development to document summarisation to risk assessment, but the classification under Article 6(3) depends not on its general capabilities but on the particular task for which it is “intended.”⁴²

As already mentioned, the intention referenced here is that of the provider, not the deployer. Article 3(12) of the AI Act defines “intended purpose” as “the use for which an AI system is intended by the provider, including the specific context and conditions of use, as specified in the information supplied by the provider in the instructions for use, promotional or sales materials and statements, as well as in the technical documentation.” The definition anchors the assessment to the provider’s declared purpose, not to the system’s inherent technical capacity. The intention must be documented and communicated, and deployers ignoring the documented

³⁹ See Couneson (2024) 214 and references therein for more on the origin of the Article 6 derogation.

⁴⁰ A likely consequence is that otherwise identical AI systems will be offered by providers in both high risk and not-high risk configurations, where the only technical difference is the documented intended purposes.

⁴¹ For a brief high-level treatment of each of the four conditions, see Couneson (2024) 215.

⁴² The importance of the “intended purpose” for the classification assessment is also reflected in Article 6(2), which refers to AI systems “referred to in Annex III”, the descriptions in which consistently use the formula “AI systems intended to be used for...”

“intended purpose” risk themselves being designated a provider of high-risk AI through the mechanism of Article 25(1)(c).⁴³

The task-based approach also has implications for AI systems that perform multiple functions. Where an AI system is configured to perform both a narrow procedural task (such as sorting incoming applications) and a substantive function (such as scoring applicants), the carve-out would apply only if the system, as deployed for its intended purpose, satisfies one of the Article 6(3) conditions. A system whose intended purpose encompasses both procedural and substantive functions cannot claim the benefit of the carve-out for the procedural component alone, as the classification assessment applies to the system as a whole.

2.3.4 The relationship between Article 6(3) and Article 14

If a system is classified as high-risk, Article 14 requires the provider to design it in such a way that it can be effectively overseen by natural persons, including by enabling the user to correctly interpret the system’s output and to decide not to use it or to override it.⁴⁴ Where an AI system is exempted under Article 6(3), these design requirements do not apply.

Recent literature has identified a consistent pattern in which formal human involvement in AI-assisted processes fails to translate into meaningful control, due to automation bias, limited competence or authority, and organisational pressure toward deference.⁴⁵ The Article 14 safeguards, even if imperfect, represent a risk-mitigation minimum for Annex III systems; their removal through the carve-outs is important context for interpreting Article 6(3).

2.3.5 The task of assessing and documenting whether an AI system qualifies for the Article 6(3) carve-out rests with the provider of the system

Article 6(4) requires the provider to document this assessment before placing the system on the market or putting it into service, and to make the documentation available to national competent authorities upon request. In addition, Article 49(2) requires providers who have determined their Annex III system not to be high-risk to register the system in the EU database established under the Act, thereby ensuring that authorities and the public have visibility of such self-as-

⁴³ Most general AI system providers include terms of service wording specifying that the intended use is not AI Act high risk, e.g. the Microsoft Enterprise AI Services Code of Conduct (2025) “Customers must not use the services ... to make decisions or take actions without appropriate human oversight as part of an application that may have a consequential impact on any individual’s legal position, financial position, life opportunities, employment opportunities, or human rights, or may result in physical or psychological harm to an individual.”

⁴⁴ See Panezi (2024) 358–368 for a detailed treatment of Article 14.

⁴⁵ See Crootof, Kaminski, and Price (2023) 437–439, 472–473 on the “MABA-MABA trap”, Enarsson, Enqvist, and Naarttijärvi (2022) on “rubber-stamping” across multiple decision-making contexts, and Lazcoz and de Hert (2023) on the need for human governance beyond individual intervention.

sessed “low-risk-use-in-a-high-risk-sector” systems. The national-level post-market enforcement counterpart to this self-assessment is set out in Article 80.⁴⁶ The evaluation under Article 80 may itself be escalated to the Commission through the Union Safeguard Procedure of Article 81 to ensure consistent application across the EU.⁴⁷

The self-assessment model with a registration safeguard reflects a regulatory design choice that seeks to balance administrative efficiency against the need for transparency and accountability. Risk self-assessments have been the subject of criticism in the literature, with the prediction that the AI Act will face the same challenges as the EU product-safety regime from which it draws inspiration, since self-assessment frameworks produce higher risk of harm to fundamental rights and safety than those involving third-party evaluation.⁴⁸

These concerns were sharpened by the Commission’s Digital Omnibus proposal of 19 November 2025, which proposed to eliminate the Article 49(2) registration obligation for self-assessed Annex III systems while retaining the documentation duty.⁴⁹ The EDPB and EDPS criticised the proposal as one that “would significantly decrease the accountability” and create incentives for providers to claim Article 6(3) exceptions without critical analysis.⁵⁰ The Council subsequently agreed on an amended omnibus proposal with the registration requirement reinstated in a simpler form.⁵¹ The European Parliament adopted its plenary position in March 2026 supporting a simplified registration obligation. Provisional agreement was reached on 7 May 2026, with formal endorsement and legal-linguistic revision pending at the time of writing.⁵²

For risk-seeking providers and deployers, the combination of vague Article 6(3) conditions, risk self-assessment, and (potentially) mostly ad-hoc transparency obligations create an incentive to characterise AI system functions and deployments in terms that bring them within the carve-outs, thereby avoiding the substantial compliance burden of high-risk classification. Risk-averse actors may avoid the carve-out altogether and accept unnecessary compliance costs. Both reactions weaken the proportionality function of Article 6(3). The asymmetry can only be reduced through operational criteria that are concrete enough to guide self-assessment and review.

⁴⁶ For more details on Article 80, see Mahler and Schwemer (2024) 1142–1148.

⁴⁷ For Norway and the other EEA EFTA states the escalation would be to the EFTA Surveillance Authority. For more details on Article 81, see Tosoni and Esayas (2024) 1149–1157.

⁴⁸ Smuha and Yeung (2025) 253–254 and references therein.

⁴⁹ European Commission (2025) (the Digital Omnibus on AI) paragraphs 6, 14, and 32.

⁵⁰ European Data Protection Board and European Data Protection Supervisor (2026), paragraphs 17–21.

⁵¹ Primarily by deleting the Annex VIII section B point 7 requirement to register a summary of the grounds on which the AI system is considered to be not-highrisk.

⁵² European Parliament (2026) amendments 11, 36, 46, 83.

2.3.6 The provider/deployer divergence under Article 6(3)

A provider designs and declares an AI system's intended purpose, and the question of whether that purpose qualifies as "narrow procedural" or "preparatory" is constrained to the deployment scenario(s) described by the provider. As indicated above, this can produce a mismatch between the provider's declared classification and the deployer's actual use. Where an AI system is declared by its provider to be intended for high-risk purposes, a particular deployer may use it only for a narrower procedural or preparatory function that would, in isolation, satisfy Article 6(3). The Act does not allow the deployer to reclassify the system on this basis to escape the deployer's high-risk obligations of Articles 26 and 27.

The opposite mismatch, a system declared low-risk by the provider but deployed for a high-risk task, is captured by Article 25(1)(c), which loops the deployer into provider obligations.⁵³ There is, however, a clear potential for responsibility pulverisation through the use of "constructive ambiguity" in the documented deployment requirements, where an ambiguously worded deployment requirement is initially interpreted strictly to support not-highrisk self-assessment by the provider, and later interpreted permissively by the deployer to justify minimising deployment costs. This has implications for the operative criteria developed in Chapter 4: criteria that turn on the conditions of deployment — transparency, verifiability, and resilience to automation bias — require the provider's declared intended purpose to set clear demands on the deployment environment. The *de lege ferenda* implications of this constraint are taken up in Section 5.3.

⁵³ For more on the Article 25(1)(c) "role-change" mechanism, see Riede and Talhoff (2024) 532–533. See Gallagher (2024) 541–551 for commentary on the Article 26 deployer obligations.

3 Legal interpretation of Article 6(3)(a) and (d)

3.1 Structure and wording of Article 6(3)

Keeping in mind the context set out in Chapter 2, I now move on to derive operative criteria from Article 6(3)(a) and (d), following the methods of interpretation set out in Section 1.4. I first examine the structure and wording of Article 6(3) as a whole.

3.1.1 The overarching condition of no significant risk of harm

Article 6(3) first subparagraph provides that an AI system falling within one of the Annex III areas “shall not be considered to be high-risk where it does not pose a significant risk of harm to the health, safety or fundamental rights of natural persons, including by not materially influencing the outcome of decision making.” This overarching condition — the absence of significant risk, including the absence of material influence on decision-making — establishes a general threshold that is both abstract and subjective.⁵⁴ The “materiality” qualifier is also inherently subjective, and may allow providers to claim that their system's influence on decisions is immaterial even where it significantly shapes the informational basis on which those decisions are made.⁵⁵ If the overarching condition is too indeterminate to constrain the self-assessment effectively, the specified conditions must serve as the primary contributor to legal certainty.

3.1.2 Article 6(3) as a closed list of carve-outs

The second subparagraph sets out four specific conditions under which the first subparagraph “shall apply”. They read as follows, with the conditions that are the focus of this thesis emphasised in bold text:

- (a) the AI system is intended to perform a narrow procedural task;**
- (b) the AI system is intended to improve the result of a previously completed human activity;
- (c) the AI system is intended to detect decision-making patterns or deviations from prior decision-making patterns and is not meant to replace or influence the previously completed human assessment, without proper human review; or
- (d) the AI system is intended to perform a preparatory task to an assessment relevant for the purposes of the use cases listed in Annex III.⁵⁶**

⁵⁴ Crucially, Mahler and Schwemer (2024) 1144 note that the threshold for “significance” in Article 6(3) is ambiguous as between risk level and risk acceptability.

⁵⁵ See Levitina (2024) 136–137.

⁵⁶ Article 6(3)(a)-(d).

The list is connected by the disjunctive “or”, indicating that the conditions are alternative rather than cumulative: satisfying any one of the four is sufficient, provided that the overarching condition of no significant risk of harm is also met.

The question arises whether the four conditions are exhaustive: Can a provider invoke Article 6(3) on a basis not enumerated in the list? Do the four conditions serve as safe harbours, sufficient but not necessary, rather than as an exclusive gateway? The enumerative reading finds some textual support in the structure of the provision itself: the first subparagraph states the “no significant risk” standard as an apparently self-standing rule, and the second subparagraph opens with “the first subparagraph shall apply” where one of the conditions is met, which could be read as triggering a presumption rather than defining the only permissible path.⁵⁷

Textual and systematic considerations support the conclusion that the list is closed: First, the legislative technique of enumerating specific derogation conditions following a general prohibition clause is a well-established mechanism in EU law for delimiting the scope of a derogation.⁵⁸ Had the legislator intended the conditions to be illustrative, one would expect the use of qualifying language such as “in particular” or “for example” – language that is absent here.⁵⁹ Second, Article 6(6) empowers the Commission to adopt delegated acts amending the conditions set out in Article 6(3), “adding new conditions to those laid down in paragraph 3, or modifying them.” The express conferral of a power to add conditions would be superfluous if the existing list were already open-ended; the provision implies that the Article 6(3) conditions can only be expanded through formal legislative amendment.⁶⁰ Third, reading the conditions as a closed list is consistent with the principle of strict interpretation of derogations in EU law, as established in the ECJ’s jurisprudence. This ensures that the scope of the carve-out is legally determinate and that providers cannot invoke unenumerated grounds to escape high-risk classification.⁶¹ The consequence of this reading is that an AI system deployed within an Annex III area that does not satisfy any of the four enumerated conditions remains classified as high-risk, even if the provider considers the system’s specific function to pose no significant risk of harm. The question is not beyond doubt, and future guidance from the Commission or the ECJ may

⁵⁷ The earliest formulation of what would later become the enacted Article 6(3) did not include the enumerated conditions. The wording was “high-risk unless the output of the system is purely accessory in respect of the relevant action or decision to be taken and is not therefore likely to lead to a significant risk.” Council of the European Union (2022) page 32.

⁵⁸ See, by analogy, the structure of GDPR Article 9(2), which sets out an exhaustive list of conditions under which the (otherwise prohibited) processing of special categories of personal data is permitted.

⁵⁹ For an example of such a non-exhaustive illustrative list in EU legislation, see GDPR Article 4(1) listing “identifiers such as a name, an identification number, location data, ...” in the definition of “personal data.”

⁶⁰ Article 6(6). Article 6(8) further specifies that any amendment “shall not decrease the overall level of protection of health, safety and fundamental rights provided for by this Regulation.”

⁶¹ On the principle that exceptions to a general rule in EU law are to be interpreted strictly, see Case C-201/13 (Deckmyn) paragraph 22.

clarify whether the overarching condition has any independent operative function beyond its role as an interpretive constraint on the four conditions.

A second interpretive question is whether satisfying one of the second-subparagraph conditions deems the overarching “no significant risk of harm” threshold met (the automatic reading), or whether the overarching condition is a separate test the provider must also pass (the cumulative reading). The automatic reading reduces “no significant risk” to an interpretive constraint on the conditions. The cumulative reading is more protective, adding a case-by-case assessment on top of the condition-based classification, at the cost of legal certainty.

The automatic reading is, in my view, the better interpretation. Recital (53) describes the carve-out as covering “specific cases in which AI systems ... do not lead to a significant risk of harm ... *because* they do not materially influence the decision-making...” (my emphasis), indicating the significance/materiality threshold is integrated in the specific case conditions. The Article 6(7) power to delete conditions by delegated act “to maintain the protection of health, safety and fundamental rights” suggests the legislator already viewed the conditions as alone determining the boundary of the carve-out. The automatic reading is also more consistent with the Act’s ambition to provide clear, *ex ante* classification rules. The cumulative reading’s advantage is preserving the overarching condition as a safety net for edge cases the legislator did not foresee. The question would benefit from clarification.⁶²

3.1.3 The “profiling of natural persons” carve-out from the carve-outs

Article 6(3) third subparagraph introduces an exception from the exceptions: AI systems “shall always be considered to be high-risk where the AI system performs profiling of natural persons.”⁶³ This contributes to the systematic interpretation of the derogation condition interpretations, and will play an important role in the operationalised criteria developed in Chapter 4.

3.2 “Narrow procedural task” under Article 6(3)(a)

3.2.1 Linguistic interpretation of “narrow”, “procedural”, and “task”

Article 6(3)(a) provides that the carve-out applies where “the AI system is intended to perform a narrow procedural task.” The condition combines three lexical elements – “narrow”, “procedural”, and “task” – each of which contributes to delimiting the scope of the exception.

⁶² Mahler and Schwemer (2024) 1144 “the assessment of significant risk of harm under Article 6(3) appears to leave some space for a concrete risk assessment in the context of assessing the conditions listed above in letters (a) to (d)” can be interpreted as leaning towards the cumulative reading. Conversely, Couneson (2024) 214 “AI systems which fulfil one or more of [Article 6(3) a-d criteria] shall not be considered as high-risk AI systems.”

⁶³ Article 3(52) defines profiling with reference to GDPR Article 4 (4).

3.2.1.1 *“Task”*

The term “task” has already been discussed briefly in Section 2.3.3 above. The Oxford English Dictionary (OED) gives the literal definition as “A piece of work imposed, exacted, or undertaken as a duty or the like; originally, a fixed or specified quantity of labour or work imposed on or exacted from a person; later, the work appointed or assigned to one as a definite duty.” It follows that “task” describes a defined, fixed and concretely specified piece of work, rather than a broader progress or abstract responsibility. “Identify if the applicant has a valid bus driver’s licence” is a task. “Find and hire a bus driver” is arguably not a task, instead representing a broader process (composed of many tasks). And “Bus driver staffing” is clearly an abstract responsibility rather than a task.

In the context of condition (a), “task” denotes a specific, delimited function that the AI system is intended to carry out: not the system’s general capabilities, but its specific operational role in the context of the process in which it is deployed.

3.2.1.2 *“Narrow”*

While a “task” must be specific and defined, there is still variation: Some activities that might be described as tasks — such as the “find and hire a bus driver” example above — are broader, pushing against the upper bound of the definition, while others are more limited in scope. The adjective “narrow” clarifies we are here operating in the latter end of the range.⁶⁴ A system extracting dates from CVs should qualify. A system that extracts dates, calculates work-history gaps, compares the gaps to a predefined significance threshold, and removes applicants on that basis might not. The narrowness criterion serves as a filter against systems that, while potentially classifiable as “procedural” in some respects, operate across a broader functional domain.⁶⁵

3.2.1.3 *“Procedural”*

The adjective “procedural” is the decisive qualifier for condition (a). Literally defined broadly as relating to “...the established or prescribed way of doing something”, in legal theory, the distinction between “procedural” and “substantive” is a fundamental one. Procedural rules and acts concern the manner, method, or form by which a process is conducted, while substantive rules concern the merits, content, or outcome of the matter at hand. The procedural/substantive distinction is a foundational concept in legal theory and is reflected in the structure of EU law

⁶⁴ OED defines “narrow” as “Limited in range or scope; precise; restricted.”

⁶⁵ The responsibility gaps and questions that arise when narrow procedural task AI systems are built using general-purpose AI model APIs are outside the scope of this thesis.

itself, where procedural rules (such as those governing actions for safeguarding EU rights) are distinguished from substantive rules (such as those defining EU rights and obligations).⁶⁶

Transposed to the AI Act context, a “procedural” task is one that relates to the form, order, or method of processing – the mechanics of handling information according to a fixed and defined course of action – rather than to the evaluation of the content or significance of that information. A system that converts unstructured data into a structured format manipulates the form of information without assessing its substance. A system that, to perform its given function, evaluates the quality, relevance, or risk implications of information performs a substantive task.

Recital (53) offers illustrative examples of tasks that may qualify under Article 6(3): “For instance, AI systems intended to perform narrow procedural tasks such as [1] transforming unstructured data into structured data, [2] classifying incoming documents into categories, or [3] detecting duplicates among documents should not be considered to be high-risk.” (numbers in brackets added for reference):

- 1) *Transforming unstructured data into structured data* involves imposing a pre-defined format on raw information – for example, extracting fields from a text document and mapping them to database entries. This is a clear example of a task involving only operations on the form or structure of data rather than on its content or meaning.
- 2) *Classifying documents into categories* is more difficult: On one hand, sorting based on formal characteristics (document type, language, sender) would clearly qualify as “procedural”. Sorting documents into less formal and more judgment-based categories such as “High risk”, “Low risk” would require evaluating the substance of a document’s content. The line between formal and substantive classification is unclear: even classification by language requires the system to engage with the content of a document to a degree. It must process enough of the text to identify the language, which involves something more than pure syntactic manipulation. The difficulty intensifies for intermediate cases. Consider a system that classifies incoming regulatory filings into pre-defined categories such as “employment matter,” “tax matter,” or “environmental matter.” The categories are fixed and externally imposed; the system does not evaluate the merits of any filing. Yet the classification necessarily requires a degree of content analysis – the system must understand what the filing is about, not merely what format it is in. A provider might plausibly argue that such classification is procedural because it operates within a closed and pre-defined taxonomy, while a regulator might counter that any classification requiring engagement with the substance of a document's content is inherently evaluative. The resolution, in my view, lies in the qualifier “narrow”: where

⁶⁶ See e.g. Case C-268/06 (Impact) paragraphs 44-46.

the classification scheme is fixed (predetermined and independent of context), bounded (with clear ringfencing of the extent of the classification's potential impact), and objectively verifiable, where the "correct" classification can be assessed by reference to formal or widely agreed criteria, the task retains its procedural character, even if it involves some degree of content engagement. Where the classification requires the system to make judgments about significance, quality, or risk, however, the task crosses from the procedural into the substantive.

- 3) *Detecting duplicates* involves comparing records for identity or near-identity, which is an operation on formal similarity.

Crucially, the Recital (53) examples are all deterministic or rule-based in character: the "correct" output is, at least in principle, objectively verifiable. That a date has been correctly parsed, a document has been correctly categorised as "English language" rather than "Norwegian language", or that two files contain the same data, can all be assessed by reference to formal criteria without requiring evaluative judgment. This is not to say that the examples are wholly free of evaluative choice, but these judgments happen at the (human-driven) design stage – for instance, where a data structuring schema is designed to ignore or prioritise certain information (e.g. "capture if the applicant has a bus driver's licence, ignore other certifications"). Such design choices, however, are made once and operate uniformly across inputs; they do not require the system to perform an evaluative judgment about the substance of any particular case. If the data structuring design choices themselves embed "black box" substantive evaluation of the kind that Article 6(3) seeks to capture, e.g. building a structured prioritised list of information snippets considered most "relevant" to a future assessment, the task crosses from procedural into substantive territory.

It is useful to conceptualise this distinction in terms drawn from information theory: the difference between syntactic and semantic processing. Syntactic processing operates on the structure or format of data – the arrangement, classification, or transformation of information based on formal properties. Semantic processing operates on the meaning or evaluative content of data – the assessment of what information signifies for a particular purpose, such as whether a candidate is suitable for a position or whether a claim is likely to be fraudulent.⁶⁷ Article 6(3)(a) can be understood as limiting the carve-out to tasks that are syntactic in character: the AI system processes the form of information without generating evaluative conclusions about its content.

⁶⁷ The syntactic/semantic distinction is drawn here for analytical purposes and does not correspond to a formal legal category. It serves to operationalise the procedural/substantive distinction in the specific context of data processing tasks performed by AI systems. See Floridi (2011) for a (comprehensive) philosophical investigation of the differentiation of data, information, and knowledge, esp. 31–42, 104–106, 134–138.

The syntactic/semantic distinction is challenged by how modern AI systems built on large language models (LLMs) process information. An LLM tasked with “transforming unstructured data into structured data” may internally perform extensive semantic processing: e.g. named entity recognition requires contextual understanding of meaning to distinguish a person's name from an organisation's name or a place name. At the technical level, the boundary between LLMs processing form and processing meaning is far less sharp than the legal distinction suggests. However, *de lege lata* the question under Article 6(3)(a) is not how the system technically processes information internally, but what the system's task is characterised as in terms of its input-output specification and its functional role in the decision-making process. A system that takes an unstructured CV as input and produces a structured table of dates, qualifications, and employer names as output performs a task that is procedural at the level of its observable function, regardless of the complexity of the internal computation required to achieve that result. The syntactic/semantic distinction operates at the level of the task's specification and output, not at the level of the model's internal representations.

3.2.2 Systematic interpretation of “narrow procedural task”

Article 14(4) requires that natural persons to whom human oversight is assigned are enabled “to properly monitor the operation of the high-risk AI system” and to “decide, in any particular situation, not to use the high-risk AI system or to otherwise disregard, override or reverse the output of the high-risk AI system.”⁶⁸ A genuinely narrow procedural task does not produce outputs requiring such override mechanisms. The “procedural” character of a task should accordingly be assessed not only by its logical position in the workflow, but also by whether the output risks engaging automation bias in the subsequent human decision-maker. A task whose output is formally intermediate and procedural, but psychologically decisive – because it frames, narrows, or pre-structures the decision-maker's cognitive field – is not procedural in the sense required by the provision: For instance, a case-management system preparing standardised documentation for cases judged to align with a simple “happy path” may, though the task being purely procedural, subconsciously tempt a time-pressed assessor to ignore signs that further investigation is needed, since investigating would forfeit the time-saving filing.⁶⁹

GDPR Article 22, on automated individual decision-making, supplies a parallel systematic reference. Its distinction between automated processing that produces legal or similarly significant effects on a natural person and processing that does not maps onto Article 6(3)(a)'s proce-

⁶⁸ Article 14(4)(a) and (d).

⁶⁹ Parasuraman and Manzey (2010) distinguish such automation complacency from automation bias and link complacency in particular to high workload and competing manual tasks.

dural/substantive divide: language identification or date-sorting facilitates a later decision without itself producing such effects, whereas evaluative judgment of a person's characteristics or situation engages both Article 22 and the high-risk regime.⁷⁰

The ECJ's judgment in Case C-634/21 (SCHUFA) provides an instructive illustration of this boundary in automated processing. The Court considered whether a credit reference agency's calculation of a credit score amounted to a "decision based solely on automated processing" under GDPR Article 22, even though the formal lending decision was the bank's. The Court rejected the agency's argument that the score was merely preparatory, holding that where the probability value plays a "determining role" in the credit decision, describing it as preparatory would create a gap in the protection afforded by Article 22.⁷¹

The SCHUFA reasoning transposes to Article 6(3)(a) by analogy: the "procedural" character of a task should be assessed not by its formal label or its position in a workflow, but by its *functional role* in relation to the decision affecting the natural person. A formally intermediate task may be functionally decisive.⁷² Sorting job applications by date of receipt is procedural; sorting them by "relevance" to the position, even if the provider characterises this as a classification task, and the deployer is not the one making the hiring decisions, is substantive. See also the discussion below on the interpretation of "preparatory task" in Article 6(3)(d).

3.2.3 Teleological interpretation: the purpose of carving out procedural tasks within high-risk activities

The teleological analysis of Article 6(3)(a) must be situated within the AI Act's dual objectives: the protection of fundamental rights and the promotion of innovation through proportionate regulation.⁷³ The purpose of carving out narrow procedural tasks is to avoid imposing the full weight of high-risk compliance on AI systems that, despite operating in a sensitive domain, perform functions that do not engage the interests the regulation seeks to protect.

⁷⁰ See Granmar (2024) 191–195 for an in-depth exploration of the relationship between AIA Article 14 and GDPR Article 22.

⁷¹ Case C-634/21 (SCHUFA), paragraphs 47–62. The Court reasoned that the "the action of the third party to whom the probability value [credit score] is transmitted draws 'strongly' on that value" to decide whether an individual obtains credit and that circumventing Article 22 by treating the score as a merely preparatory act would undermine the provision's protective purpose.

⁷² The SCHUFA analogy is not without weaknesses: GDPR Article 22 and AIA Article 6(3) serve different regulatory functions: the ECJ's reasoning in SCHUFA was developed for a right-based provision rather than for a risk classification threshold. Still, the underlying logic of the SCHUFA decision, that legal characterisation must track functional reality rather than formal labels, is a general principle of EU law that supports the proposed reading, see e.g. Case C-66/85 (Lawrie-Blum) paragraph 17.

⁷³ Recital (1).

Recital (53) articulates this purpose: there are “specific cases in which AI systems referred to in pre-defined areas specified in this Regulation do not lead to a significant risk of harm to the legal interests protected under those areas because they do not materially influence the decision-making or do not harm those interests substantially.” Narrow procedural tasks should accordingly be read as only those whose output does not bear on the legal interests that justified the inclusion of the area in Annex III.

At the same time, the purpose of the carve-out requires that it not be deprived of its *effet utile*. It is easy to construct scenarios in which even narrow procedural tasks could inadvertently cause harm: a duplication-removal algorithm wrongly drops a job applicant who shares a name and date of birth with another, or a date-parsing algorithm miscalculates the applicant’s years of experience because of inconsistent date formats. If the notion of “narrow procedural task” were read so strictly that any theoretical risk arising from such combinations of AI and human error sufficed to defeat it, no real-world AI system could satisfy the condition, the carve-out would exist on paper only, and the proportionality principle would be undermined. The ECJ has consistently held that derogations must be interpreted strictly, but not so as to deprive them of practical effect.⁷⁴ Concretely, this means that the condition must be capable of covering genuinely low-risk procedural AI functions that are deployed within high-risk areas.

The Act’s dual purposes thus impose a two-sided constraint: Article 6(3)(a) must be read strictly enough to prevent substantive decision-support functions from being relabelled as procedural, yet broadly enough to permit genuinely low-risk procedural functions to escape full high-risk compliance.

3.3 “Preparatory task to an assessment” under Article 6(3)(d)

3.3.1 Linguistic interpretation of “preparatory” and “to an assessment”

Article 6(3)(d) provides that the carve-out applies where “the AI system is intended to perform a preparatory task to an assessment relevant for the purposes of the use cases listed in Annex III.” This condition is linguistically more complex than condition (a), and its interpretation requires attention to each of its constituent elements.

The term “task” must be interpreted as in condition (a) – it anchors the assessment to a specific delimited function performed by the AI system in its intended context of use.

The adjective “preparatory” denotes that this is a task “that prepares or serves to prepare for something following; preliminary, introductory.”⁷⁵ In legal parlance, a preparatory act is one

⁷⁴ See Case C-201/13 (Deckmyn) paragraph 23.

⁷⁵ OED preparatory (*adj.*), sense 1.

that precedes and contributes to a final determination without itself constituting that determination. The distinction between preparatory and definitive acts is a staple of EU administrative law.⁷⁶ The distinction was central in the above-mentioned SCHUFA case, in which the Court found that classifying third party credit scoring as (only) a preparatory act to a subsequent credit decision would cause “a lacuna in legal protection”⁷⁷ where the credit decision-maker would be unable to provide information on, and review the merits of, the automated credit scoring process.

The prepositional phrase “to an assessment” establishes that the preparatory task must be directed towards, and in service of, an “assessment.” This implies a temporal and functional sequence: first, the AI system performs a preparatory task; then, separately, an assessment takes place, using the output of the preparatory task. The use of the indefinite article “an” suggests the preparatory task must be tied to a specific assessment act, not a vague or diffuse process.

The term “assessment” itself is not defined in the AI Act, but its ordinary general meaning denotes an act of estimation or evaluation.⁷⁸ In the context of the Annex III areas, “assessment” would refer to estimations and evaluations such as, but not limited to, a natural person’s state of mind, creditworthiness, insurability, suitability for employment, eligibility for social benefits, criminal risk, or medical condition, the state of critical infrastructure: water quality, highway traffic, power supply, the relevance of facts and laws in justice processes. The assessment task output would then in turn be input to the decision-making function that the high-risk regime seeks to regulate. Interestingly, unlike Article 6(3)(c), there is no specification that the assessment be a “human assessment”.⁷⁹ An AI system could perform a preparatory task to an assessment by a (separate and different) high-risk AI system, and the former system would still qualify for the Article 6(3)(d) derogation condition.

The combined effect of these elements is that Article 6(3)(d) covers AI tasks whose outputs serve as inputs to functionally and temporally separate human- or AI-performed estimative or evaluative acts, without themselves constituting or pre-empting those acts.

⁷⁶ See for instance Case 60/81 (*IBM v Commission*), paragraphs 10-12.

⁷⁷ Case C-634/21 (SCHUFA), paragraph 48.

⁷⁸ OED assessment (n.), sense 5.a.

⁷⁹ Article 6(3)(c) “the AI system is intended to detect decision-making patterns or deviations from prior decision-making patterns and is not meant to replace or influence the previously completed human assessment, without proper human review”

3.3.2 Systematic interpretation of “preparatory task to an assessment”

3.3.2.1 *The distinction between the preparation stage and assessment stage*

The systematic context of the AI Act supports a clear functional separation between the preparation stage and the assessment stage. If the two stages are so intertwined that the AI system’s output effectively constitutes the assessment, the condition is not satisfied.

The distinction finds support in the broader architecture of Annex III. The high-risk categories in Annex III identify specific decision-making contexts – such as “AI systems intended to be used to evaluate the creditworthiness of natural persons” or “AI systems intended to be used by public authorities ... to evaluate the eligibility of natural persons for public assistance benefits and services.”⁸⁰ The verb “evaluate” in these descriptions denotes the assessment stage: the act of forming a judgment about a person’s situation. A preparatory task, by contrast, is one that provides the foundation on which the evaluation is performed – the visualised data, the retrieved documents, the translated texts – without itself performing the evaluation.

3.3.2.2 *The role of human oversight and responsibility*

The systematic relationship between Article 6(3)(d) and Article 14, as for condition (a), has two implications.

First, for a task to be genuinely preparatory it must leave the human decision-maker in a position to conduct the assessment independently. If the output pre-structures the assessment – by filtering information, by ranking, or by issuing a recommendation the assessor must affirmatively reject – the system has shifted the cognitive burden, and the carve-out should not apply.⁸¹

Second, the output of a preparatory task should be transparent and verifiable. Whether it is verifiable in fact depends on the assessor’s capabilities: an Urdu-to-Norwegian translation in an asylum case is preparatory only where the assessor has practical access to Urdu proficiency. The same AI system may therefore qualify in one deployment context and not in another, which makes it important that the provider documents the assumed “specific context and conditions

⁸⁰ Annex III, points 5(a) and 5(b).

⁸¹ Crootoof, Kaminski, and Price (2023) 473–487 identify nine distinct roles that humans may play in algorithmic systems, including “correctional” roles (correcting errors and adding situational context), “friction” roles (slowing down automated decisions to prevent harms), and “stand-in” roles (providing a symbolic demonstration that oversight has occurred without substantive review). The label “human in the loop” is compatible with all of these roles, including those that provide no meaningful safeguard. For Article 6(3)(d) to serve its purpose, the relevant question is not merely whether a human performs the subsequent assessment, but whether the system’s design and the nature of its output support a corrective or substantive human role rather than a merely symbolic one.

of use”⁸² and that the deployer adheres to them, otherwise risking attaining provider obligations through the Article 25(1)(c) mechanism.

Similarly, if a system is classified as not-highrisk on the ground that its task is merely preparatory, the Article 14(4)(b) safeguards against “automation bias” do not apply. This is the case where the output is factual and verifiable (such as a list of documents or the above-mentioned translation example), but less so where the output is probabilistic and evaluative (such as a relevance score⁸³ or an executive summary).⁸⁴ In the latter case, the cognitive dynamics of automation bias mean that the AI system effectively contributes to the assessment, even if the assessor technically has access to the underlying data for review when making the assessment.⁸⁵

3.3.3 Teleological interpretation: the purpose and limits of excluding preparatory tasks from high-risk classification

The purpose of Article 6(3)(d) parallels that of condition (a): to keep the high-risk classification from extending to AI systems whose function does not engage the interests the regulation seeks to protect. Its specific rationale is that many Annex III use cases involve multi-stage processes in which AI is deployed at the preparatory stage but not at the stage of evaluative judgment – for example, an AI database search for candidates matching formal recruitment criteria while a human interviewer performs the substantive suitability assessment, or AI-assisted retrieval and translation of case law while the judicial assessment remains the prerogative of the judge.

The purpose of the carve-out is thus to recognise that the high-risk regime is concerned with AI functions that bear on the substance of assessments and decisions affecting fundamental rights, and that functions which merely facilitate the administrative preconditions for such decisions – making information available, organising files, translating documents – do not require the same

⁸² From the “intended use” definition of Article 3(12).

⁸³ In Enarsson, Enqvist, and Naartijärvi (2022) 141–145 the authors highlight the Swedish Social Insurance Agency’s use of automated decision-support systems for sickness benefit assessments, in which AI systems flag cases meeting certain statistical criteria for expedited processing while the formal assessment is performed by a human administrator. An internal audit found that investigative insufficiencies were present in 67 percent of examined cases, and that only approximately one-fourth of cases were both sufficiently investigated and correctly assessed. These findings exemplify that where AI systems perform “preparatory” work that enriches the informational basis on which human assessors rely, the quality of the subsequent human assessment may be compromised – not because the AI system performs the assessment itself, but because it shapes the cognitive conditions under which the assessment takes place, e.g. through a perceived lack of need for additional (human) investigation.

⁸⁴ In Wachter, Mittelstadt, and Floridi (2017) 78–84 the authors distinguish between ex ante transparency (system-level information about how a system functions) and ex post explanation (decision-specific reasoning about why a particular output was produced). Their analysis supports the point that a preparatory output must be sufficiently understandable and verifiable for the assessor to engage critically with it.

⁸⁵ This argument draws on the SCHUFA doctrine: if the AI system’s output plays a “determining role” in the assessment, the nominal involvement of a human does not change its functional character.

intensity of regulation. At the same time, the carve-out must not be allowed to become a vehicle for exempting decision-support systems that, in substance, perform or pre-empt the assessment. In any case the SCHUFA doctrine of GDPR Article 22, discussed in Section 3.2.2 above, provides an upper boundary by analogy: A task should not be considered genuinely preparatory in the context of Article 6(3)(d) if its output plays a “determining role” in the following assessment.⁸⁶

3.4 The distinction between “procedural” tasks and “preparatory” tasks

What is perhaps less clear in the context of the AI Act is the distinction between the “procedural” tasks of Article 6(3)(a) and the “preparatory” tasks of Article 6(3)(d). As discussed above and in Section 3.2 they both share the characteristic of being separate from and not substantially affecting estimative or evaluative tasks, or the decision-making based on such assessments. Why did the legislator deem it necessary to include both Articles 6(3)(a) and (d) as separate conditions? An obvious difference is that Article 6(3)(a) contains no temporal qualification: A procedural task may well be performed *after* the assessment or decision to which it belongs, for instance to document and communicate the decision according to a set procedure. Apart from this, for tasks performed before an assessment or decision, “preparatory” casts a broad net that absorbs “procedural”: All preceding procedural tasks are by nature also preparatory, but not all preparatory tasks are merely procedural.

Recital (53) reinforces this reading by offering examples of tasks that may fall within the preparatory category. The recital mentions AI systems that constitute “smart solutions for file handling, which include various functions from indexing, searching, text and speech processing, or linking data to existing data sources, or AI systems used for translation of initial documents.” A shared feature of these “preparatory” examples compared to their “procedural” counterparts is that they create or significantly transform information to make more accessible and usable for a later assessment, without assigning evaluative weight. A search system creates indexing metadata to better facilitate the search for relevant files and insights. A transcription system converts audio into more easily searchable text. A translation system converts a document into a language the decision-maker can read. This is something more than purely procedural processing, but neither function involves judgments about the significance of the information for the outcome of an assessment.

As with procedural tasks, the preparatory boundary becomes contested where the AI system’s output does more than organise or present information – where it synthesises, ranks, transcribes, translates, or summarises in a way that structures the decision-maker’s cognitive field. A system that generates indexes that allow retrieval of documents matching a search specification and

⁸⁶ The limits of the SCHUFA analogy are discussed in footnote 72.

presents them in chronological or alphabetical order is preparatory – at least if the indexing/retrieval algorithm is formal and transparent. A system that retrieves documents and also highlights the passages deemed most “relevant” to a given assessment – thereby directing the assessor’s attention and framing the analysis – occupies an intermediate zone: Such a relevance classification might easily qualify as an assessment in itself. And a system that generates a summary “risk classification” or “recommendation” based on a synthesis of all the retrieved information has clearly crossed from pure preparation into the assessment stage.

3.5 “Profiling of natural persons” under Article 6(3) third subparagraph

The third subparagraph of Article 6(3) establishes a categorical override to the carve-out mechanism: “Notwithstanding the first subparagraph, an AI system referred to in Annex III shall always be considered to be high-risk where the AI system performs profiling of natural persons.” This carve-out from the carve-outs is absolute, “always” underscores that it leaves no room for case-by-case risk assessment of profiling systems.

Article 3(52) ties the definition of “profiling” to GDPR Article 4(4): “‘profiling’ means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person’s performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements.”

The profiling override creates an important interaction between the AI Act and the GDPR. Because the AI Act imports the GDPR definition of profiling, developments in the interpretation of that definition – including guidance from the European Data Protection Board and future ECJ jurisprudence – will directly affect the scope of the Article 6(3) carve-out. The Article 29 Working Party (now the EDPB) adopted guidelines which interpret the concept of profiling broadly and emphasise that the evaluative element is a decisive criterion: processing that merely handles personal data without evaluating the person does not constitute profiling, while processing that draws inferences about personal characteristics does constitute profiling, even if no material decision follows immediately from the assessment.⁸⁷ This interpretative guidance, while not binding, provides a relevant point of reference for the application of the profiling override under the AI Act.

A full review of the concept of “profiling” in GDPR is outside the scope of this thesis. For present purposes it suffices to note that the override is engaged whenever an AI task uses personal data within the meaning of GDPR Article 4(1) to evaluate personal aspects of a natural person. The list of aspects in Article 4(4) is illustrative and non-exhaustive: any automated

⁸⁷ Article 29 Working Party (2018) pages 6-8.

evaluation of a personal aspect falls within the definition, whether or not the specific aspect is enumerated.

The term “evaluate” indicates such profiling would involve estimates and assessments that would in themselves quickly challenge the procedural/preparatory boundaries of the Article 6(3)(a) and (d) conditions. The override sets an absolute floor for the carve-outs: an AI system that evaluates personal aspects on the basis of personal data is always high-risk in an Annex III context, with far-reaching consequences in sectors whose core function is the assessment of individuals. For example, in creditworthiness evaluation, AI systems typically assess a person’s “economic situation” and “reliability” — both enumerated profiling aspects — and accordingly cannot benefit from the carve-out regardless of how the provider characterises the task. Health and life insurance prediction provide a parallel example. The categorical override forecloses case-by-case auditing of harmful profiling claims, confining the Article 6(3) carve-outs to purely administrative and logistical tasks without human-evaluative elements.

4 Operationalising the criteria: a framework for applying Article 6(3)(a) and (d)

4.1 From interpretation to operationalisation

The following framework is derived from the legal analysis in Chapters 2 and 3 and is offered as a doctrinal contribution to the applied interpretation of Article 6(3)(a) and (d). The criteria are designed to be applied sequentially: each criterion must be satisfied before proceeding to the next. A negative answer at any stage terminates the assessment with the conclusion that the carve-out does not apply. This sequential structure reflects the principle of strict interpretation of derogations: each criterion functions as a necessary condition, and only the cumulative satisfaction of all criteria is sufficient to support the conclusion that the AI system’s task falls within Article 6(3)(a) or (d).⁸⁸

4.2 The proposed criteria

4.2.1 Criterion 1: The profiling threshold – does the task involve profiling of natural persons?

The assessment starts with determining whether the AI system performs profiling of natural persons within the meaning of the GDPR Article 4(4).⁸⁹ As discussed in Section 3.5, the profiling override is absolute: where it applies, the AI system is always high-risk, regardless of any other consideration. The criterion gives a preliminary gate: If the AI system performs profiling, the Article 6(3) carve-outs are categorically unavailable.

Beginning the assessment with this criterion is practical. A significant number of AI systems deployed in Annex III areas — creditworthiness assessment, insurance risk scoring, predictive policing, recidivism prediction, employee performance monitoring — involve profiling tasks by their very nature. For these tasks, any further analysis of the “procedural” or “preparatory” character is superfluous. At the same time, the profiling override applies at the level of the task, not of the domain: Other AI systems within the same workflow — for example, those performing data deduplication, document classification, or speech-to-text transcription — remain eligible for the carve-out where the substantive criteria below are satisfied. The profiling threshold filters out many common assessment-bearing tasks at the outset, concentrating the remaining

⁸⁸ The sequential structure draws on the approach proposed for AI risk classification in Hanif et al (2024), who advocate decision-tree frameworks for operationalising complex regulatory classifications.

⁸⁹ Article 6(3) third subparagraph, read in conjunction with Article 3(52) and GDPR Article 4(4). See also GDPR Article 29 Working Party (2018) pages 6-8 on the broad interpretation of the evaluative element in profiling.

analysis on the auxiliary tasks that do not involve the evaluative assessment of personal aspects (but might still be high risk through other mechanisms, discussed in the following criteria).⁹⁰

The criterion also has a methodological function. As observed in Section 3.5, the EDPB’s interpretation of profiling emphasises the evaluative element: processing that merely handles personal data without evaluating the person does not constitute profiling. This distinction maps directly onto the boundary between syntactic and semantic processing discussed in Section 3.2.1. An AI system that processes personal data for purely formal purposes – verifying the format of a national identity number, detecting duplicate entries in a database of applicants, or converting a handwritten application form into structured digital text – handles personal data but does not evaluate the person. Such systems pass this first criterion and proceed to the substantive assessment under the remaining criteria.⁹¹

4.2.2 Criterion 2: Task delimitation – is the task specific, defined, and bounded?

The second criterion follows from the textual analysis of “task” in Section 3.2.1 and the “narrow” qualifier applicable to condition (a). To qualify under either Article 6(3)(a) or (d), the function performed by the AI system must constitute a specific, delimited, and bounded piece of work within a broader process. This requirement has three sub-dimensions.

First, the task must be identifiable as a discrete function with clear inputs and outputs. It must be possible to describe what the AI system does in concrete operational terms: what data it takes in, what operations it performs, and what output it produces. An AI system whose function is only described in vague or open-ended terms — “assisting with recruitment decisions” or “supporting insurance assessments” — fails this requirement, because the description does not identify a delimited task but rather a broad functional domain.

Second, the task must have definable boundary conditions: a clear beginning and end within the process in which it is embedded. The function should not extend across multiple stages of the decision-making workflow. An AI system that both collects and organises evidence and then prioritises or ranks it for the decision-maker is performing two functionally distinct tasks, one of which may be preparatory while the other may be substantive. The provider cannot bundle these into a single “task” description and claim the benefit of the carve-out for the composite.⁹²

⁹⁰ In Section 5.3 I briefly discuss *de lege ferenda* whether the automatic disqualification of all profiling tasks should be replaced by a more nuanced approach.

⁹¹ EDPB’s guidance in Article 29 Working Party (2018) emphasises that not all automated processing of personal data constitutes profiling; only processing that “involves some form of assessment or judgment about a person” qualifies.

⁹² See discussion in Section 2.3.3.

Third, for Article 6(3)(a) specifically, the task must be “narrow”: it must operate at the limited end of the scope spectrum. An AI system performing a single, well-defined data transformation (extracting dates from documents) satisfies the narrowness requirement. An AI system performing a range of data transformation tasks across a broader workflow, even if each individual task is procedural, may not satisfy the narrowness criterion where its combined function spans a significant portion of the information processing chain. The narrowness qualifier, however, applies only to Article 6(3)(a). Article 6(3)(d) speaks of a “preparatory task” without that qualifier, so a preparatory task may be broader in scope than a narrow procedural task, provided the other criteria are met.⁹³

4.2.3 Criterion 3: The nature of the operation – syntactic or semantic?

The third criterion draws on a central distinction developed in the textual analysis of Article 6(3)(a) and (d): the distinction between syntactic and semantic processing. A task is syntactic where it operates on the form, structure, or format of information without engaging with its evaluative content. A task is semantic where it operates on the meaning, significance, or evaluative implications of information in relation to the decision to be made.⁹⁴ The criterion applies to both Article 6(3)(a) and Article 6(3)(d), but with different stringency: Article 6(3)(a) requires the task to be strictly syntactic, while Article 6(3)(d) tolerates a degree of semantic processing provided it does not generate evaluative conclusions about the subject matter of the subsequent assessment. The two cases are addressed in turn below.

For Article 6(3)(a), this criterion is decisive: only tasks that are syntactic in character qualify as “procedural.” The Recital (53) examples – transforming unstructured data into structured data, classifying documents into formal categories, detecting duplicates – are all syntactic operations whose correctness is, in principle, objectively verifiable by reference to formal criteria. A system that categorises incoming insurance claims by document type (medical certificate, police report, invoice) performs a syntactic operation; a system that categorises the same claims by estimated severity or likely outcome performs a semantic one.

For Article 6(3)(d), the syntactic/semantic distinction operates differently. As discussed in Section 3.3.2, a preparatory task may involve a degree of semantic processing – translation, transcription, indexing – that goes beyond purely formal data manipulation. The Recital (53) examples of preparatory tasks include “text and speech processing” and “translation of initial documents,” which inherently involve engagement with the meaning of language. The decisive question under Article 6(3)(d) is therefore not whether the operation is strictly syntactic, but

⁹³ See the discussion in Section 3.4 on the distinction between “procedural” and “preparatory” tasks.

⁹⁴ See Section 3.2.1.3, drawing on Floridi (2011) for the distinction between syntactic and semantic information processing.

whether it generates evaluative conclusions about the subject matter of the subsequent assessment. A translation makes information accessible in another language; it does not evaluate its significance for the decision. A transcription converts speech to text; it does not assess the credibility or relevance of what was said. An indexing system creates retrievable metadata; it does not determine which information matters.

This yields a more nuanced formulation of Criterion 3 for the two conditions: Under Article 6(3)(a), the task must be strictly syntactic – limited to operations on the form, structure, or format of data. The system must not produce outputs that require the exercise of evaluative judgment to generate, as opposed to the application of formal, rule-based, or deterministic criteria.⁹⁵ Under Article 6(3)(d), the task may involve a degree of evaluative semantic processing, but the output must not contain evaluative conclusions about – or assign evaluative weight to information in relation to – the subject matter of the subsequent assessment. This is closely tied to Criterion 4 below. In practice the Article 6(3)(d) formulation will be the most common application of Criterion 3, since all preceding procedural tasks are also by nature preparatory.

4.2.4 Criterion 4: Functional role – does the output bear on the substance of an assessment or decision?

Even where a task is delimited and operates within the syntactic or permissible semantic range, the Article 6(3) carve-out requires that it does not bear on the substance of the relevant decision or assessment. This criterion operationalises the SCHUFA doctrine discussed in Chapter 3: the legal characterisation of a task depends not on its formal position in a workflow, but on its functional role in relation to the decision affecting the natural person.

The operative question is: Could a reasonable decision-maker’s conclusion be different depending on how the AI system performs its task, compared to another (differently designed) AI system, or a human assistant? If the answer is yes — if choices made through the AI system’s design, inference model, and output presentation shapes, narrows, or pre-structures the informational basis on which a substantive assessment or decision is made in a way that could alter the outcome — then the AI system’s task bears on the substance of the decision and the carve-out does not apply. If the answer is no — if the AI system’s task is such that performing it with

⁹⁵ This formulation reflects the verifiability characteristic identified in the Recital (53) examples: the “correct” output of a procedural task is objectively determinable. See the discussion in Section 3.2.1.

a different system, or having a human perform the task, would not change the substantive information available to the decision-maker (barring outright system malfunctions and human error) — then the task does not bear on the substance of the decision.⁹⁶

This criterion can be further operationalised through a counterfactual test: Would the decision-maker, in a plausible scenario, reach a different substantive conclusion if the AI system was unavailable, and the decision maker had to perform the procedural or preparatory task themselves? For instance, the task of sorting job applications by the applicants’ alma mater does not bear on the substance of a hiring decision: the same applications are before the assessor regardless of their ordering, and no substantive information is added or subtracted. Sorting applications by “relevance” to the job description, by contrast, structures the assessor’s cognitive field in a way that could affect which candidates receive closer attention, bearing directly on the substance of the decision.

As indicated above, what amounts to a “different substantive conclusion” should not be extended to include errors from system malfunctions (or human error) in the otherwise purely procedural or preparatory function itself. The *effet utile* principle discussed in Section 3.2.3 requires that the carve-out not be deprived of its practical effect. Every AI system is capable of error, and a translation error, a transcription mistake, or even a mis-parsed date could in principle lead a decision-maker to a different conclusion. If the counterfactual test were applied so stringently as to capture every possible error in a procedural or preparatory function, the carve-out would be rendered empty. The question must be to the designed functional role of the output, not to the consequences of malfunction.⁹⁷

4.2.5 Criterion 5: Transparency – is the output transparent and verifiable?

The fifth criterion follows from the systematic relationship between Article 6(3) and Article 14 on human oversight, as analysed in Sections 3.2.2 and 3.3.2.2. For a task to be genuinely procedural or preparatory, the output must be transparent and verifiable by the person who performs the subsequent assessment or decision.

Transparency here means that the nature of the operation performed by the AI system must be comprehensible to the person receiving the output. The assessor should understand what the system has done: that it has translated a document, converted speech to text, or extracted dates from a collection of files. Verifiability means that the recipient of the output must be able, in

⁹⁶ This counterfactual test is inspired by the SCHUFA reasoning, in which the Court assessed whether the credit score played a “determining role” in the lending decision by examining whether the third party would reach a different conclusion absent the score. See Case C-634/21 (SCHUFA), paragraph 50.

⁹⁷ See the discussion of *effet utile* in Section 3.2.3: the carve-out must not be interpreted so restrictively that no real-world AI system could satisfy it.

principle, to check the output against the underlying data. A translated document can be checked against the original by a person with the requisite language proficiency. A date extraction can be verified against the source document. A duplicate detection result can be confirmed by comparing the flagged records.

Where the output is opaque — where the system produces a result that the assessor can observe but cannot meaningfully verify — the conditions for the carve-out are not satisfied. The distinction between *ex ante* transparency (system-level information) and *ex post* explanation (decision-specific reasoning) is relevant here.⁹⁸ The Article 6(3) carve-out does not require full algorithmic transparency, but it does require that the output be of a kind that a competent assessor can critically evaluate. A list of extracted data fields is verifiable. A probabilistic relevance score, generated by an opaque model, is not – or at least not without the kind of technical infrastructure that Article 14 is designed to mandate.

This criterion also has a contextual dimension. Whether an output is verifiable depends not only on the nature of the output itself but on the capabilities and resources of the person who receives it. As noted in Section 3.3.2.2, translating an asylum application from Urdu to Norwegian is verifiable only if the assessor has access to proficiency in the original language. A system that is deployed in a context where practical verification by the assessor is impossible is, from a functional perspective, not producing a transparently verifiable output – even if the output is, in abstract, of a kind that could be verified. The verifiability assessment must be made with reference to the actual conditions of deployment, not to an idealised scenario.⁹⁹

4.2.6 Criterion 6: Automation bias resilience – is the output inherently resistant to uncritical reliance?

The final criterion addresses the risk of automation bias; a concern identified in Sections 3.2.2 and 3.3.2 as central to the systematic interpretation of both conditions (a) and (d). AI systems exempted under Article 6(3) are relieved of the very human oversight design requirements that are intended to counteract automation bias. The exemption is justified only where the nature of the AI system’s task and output is such that the risk of uncritical reliance is inherently limited.

This criterion asks whether the output of the AI system is of a kind that, by its nature, invites independent assessment rather than deference. Factual, discrete outputs – a list of extracted dates, a translated document, a set of identified duplicate records – are inherently resistant to automation bias because they present information without suggesting a course of action. They

⁹⁸ Wachter, Mittelstadt, and Floridi (2017) 78–84 explore this distinction in the context of algorithmic decision-making under the GDPR.

⁹⁹ This contextual approach is consistent with the Act’s definition of “intended purpose” in Article 3(12), which anchors the assessment to “the specific context and conditions of use.”

do not tell the decision-maker what to think about the information; they merely make the information available in a usable form. By contrast, outputs that aggregate, evaluate, or assign probabilities — an executive summary emphasising “key findings,” a relevance ranking, a risk score — are inherently susceptible to automation bias because they compress complex information into a cognitively convenient form that the decision-maker is incentivised to adopt rather than scrutinise.¹⁰⁰

The AI Act’s treatment of automation bias has been criticised for requiring awareness of the risk without adequately aligning provider and deployer obligations with the cognitive, design-related and contextual mechanisms through which automation bias arises, despite Article 14’s provider-side human-oversight design duties and Article 26’s deployer-side competence and training obligations.¹⁰¹ This concern reinforces the importance of Criterion 6 for the Article 6(3) assessment: if even the Article 14 and 26 safeguards are not to apply, it follows systematically that the nature of the task and its output must be such that the risk of cognitive reliance is inherently very low. The assessment must consider both the formal characteristics of the output and the cognitive dynamics of the context in which it is received. An output that appears “neutral” in isolation may become decision-determinative when presented to an overworked assessor processing a high volume of cases under time pressure.¹⁰²

4.3 Application to illustrative cases

The framework is tested in three illustrative case studies detailed in the Appendix, each drawn from a distinct Annex III sector. The summaries below highlight what each case contributes to the analysis; the criterion-by-criterion application is set out in full in the Appendix.

The first case examines an AI system used to translate asylum applicants’ written statements into the language of the receiving state, the translation feeding into the eligibility assessment by the case officer. The case illustrates the contextual sensitivity of Criterion 5: the same AI system may qualify for the Article 6(3)(d) carve-out in one deployment context and not in another, depending on whether the case officer has practical access to native-language proficiency to verify the translation. Where verification is illusory, a translation that is in abstract verifiable becomes functionally determinative, and the carve-out does not apply.

¹⁰⁰ See Crotoft, Kaminski, and Price (2023) pages 473-487 on the distinction between “correctional” roles (in which humans genuinely assess and override AI outputs) and “stand-in” roles (in which human involvement is symbolic rather than substantive).

¹⁰¹ Laux and Ruschemeier (2025) pages 1524-1526, especially Section III and III.1.

¹⁰² See Enarsson, Enqvist, and Naarttijärvi (2022) on the Swedish Social Insurance Agency audit findings discussed in Section 3.3.2.2, footnote 83.

The second case examines a multi-stage recruitment workflow – data extraction, rule-based scoring, soft-criteria evaluation, and final ranking. It illustrates the filtering function of the profiling override: the earliest stage (data extraction) qualifies for the carve-out under both Article 6(3)(a) and (d), while each subsequent stage that introduces evaluative engagement with candidates’ personal aspects – even through transparent rule-based criteria – is foreclosed by the third-subparagraph override. The result is a characteristic gradient: AI can support the administrative preconditions of recruitment without engaging the high-risk regime, but the moment it begins evaluating the candidate, the carve-out is categorically unavailable.

The third case examines a clinical documentation workflow comprising audio transcription, speaker classification, statement classification, removal of non-relevant content, and confidence scoring of classifications. It demonstrates, on the one hand, that AI systems can perform highly semantic operations and still qualify for the Article 6(3)(d) carve-out where the operation is undeniably preparatory and presents no significant risk of steering the subsequent assessment. On the other hand, it produces a paradoxical result discussed in Section 5.3 below: adding confidence scores to an otherwise carve-out-eligible classification task takes the system out of the carve-out, because the confidence scoring directs the human assessor’s attention in a manner that bears on the quality and thoroughness of the medical review.

4.4 On the interrelation of criteria 3 to 6

Criteria 3 to 6 are interrelated aspects of a single underlying inquiry: whether the AI system’s task remains functionally separate from the substantive assessment or decision. They are not collapsible into a single test that covers both 6(3)(a) and (d) pathways, however, because the boundary between procedural/preparatory and substantive is itself multi-dimensional.

Criterion 3 asks what kind of operation the system performs: does it process the form of information (syntactic) or its evaluative content (semantic)? A system may be syntactic in its operation – producing the output through formal, rule-based parsing – yet substantive in its functional role, where that output has strong impact on an outcome: e.g. a date-extraction algorithm whose parsed value is solely relied upon to decide whether a statutory time limit has been met. Such a system passes Criterion 3 but fails Criterion 4. Conversely, a system may perform semantic processing, translation, transcription, whose output does not bear on the substance of the subsequent assessment because it merely makes information accessible without assigning evaluative weight. Such a system may fail Criterion 3’s strict syntactic test under Article 6(3)(a) yet still qualify under Article 6(3)(d) if Criterion 4 is satisfied. Finally, a non-preparatory procedural system (e.g. generating documentation after a decision has been made) may fail the Criterion 3 syntactic test while passing Criterion 4.

Criterion 4 asks what functional role the output plays: does it contribute to the substance of the decision? This is a question about the output's relationship to the decision-making chain, not about the nature of the operation that produced it. The SCHUFA-derived counterfactual test operationalises this inquiry. A system that performs a syntactic operation may still fail Criterion 4 if its output, despite being formally procedural, plays a “determining role” in the subsequent assessment. Criterion 4 thus catches cases that Criterion 3 alone would miss.

Criterion 5 asks whether the output is transparent and verifiable, whether the human assessor can check the AI's work. This is analytically distinct from both the nature of the operation (Criterion 3) and the functional role of the output (Criterion 4). A system may perform a semantic but non-evaluative operation (passing Criterion 3 for Article 6(3)(d)) whose output does not bear on the substance of the decision (passing Criterion 4) but which is nevertheless opaque, e.g. a search system that uses a proprietary relevance algorithm that the assessor cannot interrogate to retrieve and translate news articles for anti-money-laundering know-your-customer research. Such a system may fail Criterion 5 even though it passes Criteria 3 and 4.

Criterion 6 adds the cognitive dimension. Even transparent outputs with no formal bearing on the substance of the decision may invite complacent deference if presented as confidence scores, rankings, or colour-coded priorities. The clinical documentation confidence-score example summarised in Section 4.3 and developed in full in the Appendix illustrates this dynamic.

In short, Criterion 3 asks *what the system does*; Criterion 4 asks *what the output means for the decision*; Criterion 5 asks *whether the human can check*; and Criterion 6 asks *whether the human will check*. Each captures a dimension that the others, on their own, would miss.

5 Conclusion

5.1 Answer to the research question

Six operative criteria are derived in this thesis from the AI Act's text, systematic context, analogies from GDPR Article 22 and the ECJ's SCHUFA judgment, and information theory, to operationalise the notions of “narrow procedural task” and “preparatory task to an assessment” in Article 6(3)(a) and (d):

1. whether the task involves profiling of natural persons, which categorically precludes the carve-out,
2. whether the task is specific, bounded, and – for Article 6(3)(a) – narrow,
3. whether the operation is syntactic rather than semantic, or – for Article 6(3)(d) – whether it is preparatory without generating evaluative conclusions about the subject matter of the subsequent assessment,
4. whether the output bears on the substance of the decision, assessed through a counterfactual test derived from the SCHUFA “determining role” doctrine,
5. whether the output is transparent and verifiable by the human assessor in the actual conditions of deployment; and
6. whether the output is inherently resistant to uncritical reliance in the absence of Article 14 human-oversight safeguards.

Applied sequentially and cumulatively, these criteria reduce the risk of regulatory arbitrage by providing a structured methodology that prevents substantive decision-support functions from being re-characterised as “procedural” or “preparatory,” while preserving the carve-outs' *effet utile* for genuinely low-risk administrative and logistical AI functions deployed within Annex III areas.

5.2 Summary of findings and contributions

The answer to the research question rests on the interpretive analysis conducted in Chapters 2 through 4. The Article 6(3) carve-out operates as a proportionality “safety valve” on a blanket high-risk presumption within Annex III; the assessment is task-centred rather than system-centred; and the self-assessment model places a premium on the clarity of the substantive criteria. Textual analysis, drawing on information theory, identified the key distinctions: between “narrow” and broad tasks, between “procedural” (syntactic) and substantive (semantic) processing, and between “preparatory” functions and those bearing on the substance of an assessment. The systematic analysis drew on Article 14 on human oversight, the GDPR Article 22 framework on automated individual decision-making, and the ECJ's SCHUFA judgment, which established

that the legal characterisation of a task depends on its functional role in the decision-making chain rather than on its formal position in a workflow. The teleological analysis confirmed a two-sided interpretive constraint: the carve-outs must be read strictly enough to prevent substantive decision-support functions from being relabelled as procedural or preparatory, yet broadly enough to preserve their *effet utile* for genuinely low-risk administrative functions. From this the operationalised criteria were derived.

The illustrative examples summarised in Section 4.3 and developed in full in the Appendix demonstrate the framework's practical utility across Annex III sectors. Applying the criteria to a clinical documentation workflow illustrates a counterintuitive but analytically interesting result: an AI system may lose its Article 6(3)(d) exemption by providing more information — such as confidence scores — that steers the human quality assurance process, even where the underlying classification task itself remains preparatory. The recruitment workflow illustrated the profiling override's decisive filtering function: the moment an AI system begins evaluating personal characteristics of candidates, even through transparent rule-based criteria, the carve-outs are categorically unavailable.

The criteria are intended to serve as a practical interpretive tool for three principal audiences: Providers, who bear the Article 6(4) self-assessment burden and need a structured methodology for documenting their classification; deployers, who under Article 25(1)(c) risk reclassification as providers if they use a system outside its documented intended purpose, and need a basis for critically evaluating the provider's characterisation; and national market surveillance authorities, who need assessable dimensions for reviewing self-assessments.

5.3 *De lege ferenda*: improving the legislative framework

The analysis reveals weaknesses in the Article 6(3) framework that warrant legislative, Commission, or supervisory clarification. I offer four *de lege ferenda* reflections:

First, I refer to the analysis in Section 3.1.2, arguing that the relationship between the general “no significant risk of harm” condition and the four enumerated conditions is unclear, but that the better reading is that the four conditions are the sole gateway to the carve-out. Confirming this in Article 6(5) guidance or through textual amendment would improve legal certainty.

Second, the profiling override in the third subparagraph operates as a categorical bar: any AI task that evaluates personal aspects of a natural person triggers the high-risk regime, regardless of whether the evaluation operates to the person's detriment or to their benefit. This binary treatment may be appropriate for prototypical cases such as credit scoring or recidivism prediction, where the evaluation directly conditions an adverse outcome. It is less obviously appropriate for “Pareto-improving” deployments in which a profiling system clearly supports the person profiled without exposing them to any risk of harm – for example, an employee-

onboarding or upskilling tool that profiles a new hire's existing skills in order to recommend a personalised training pathway, where the recommendation cannot serve as the basis for any decision adverse to the employee and the employee retains the option to disregard it. Capturing such systems within the high-risk regime by virtue of the profiling alone is, in my view, disproportionate. A targeted Article 6(5) clarification, or a textual amendment to the third subparagraph, could permit profiling-based AI to fall within Article 6(3)(d) where the system's output is used only in the deployer's stated interest in supporting the person profiled, the output cannot be used as the basis for any adverse decision concerning that person, and the person retains a meaningful right to opt out. With such conditions I believe the cost — making it somewhat harder to police covert evaluative scoring of natural persons — is worth paying to remove the disincentive for genuinely beneficial profiling-based applications in Annex III sectors.

Third, the framework yields a paradoxical result in cases involving confidence reporting on otherwise preparatory tasks, as illustrated by the clinical documentation case study summarised in Section 4.3 and detailed in the Appendix. A system performing speaker labelling or statement labelling without confidence scores passes the derogation criteria, while a system returning the same outputs with confidence scores included does not, because the confidence scores risk steering the human quality assurance. It is somewhat counterintuitive that a system can change classification from "high-risk" to "not-highrisk" simply by offering less information relevant to the subsequent human review. In my view, the result is nevertheless defensible. Article 6(3) is concerned not with the volume of information an AI system produces, but with whether the form of that information meaningfully shifts cognitive effort from the human assessor to the system. A bare classification leaves the assessor in the position of judging whether to trust each label on its own evidentiary footing; a confidence-scored classification invites the assessor to partially delegate that judgment to the system. The Article 14 design requirements are calibrated to the latter case, and so is the high-risk regime as a whole. The open *de lege ferenda* question, in my view, is whether providers should nevertheless be incentivised to provide such transparency information – for example, through a graduated regime that admits transparent confidence reporting under stricter human-oversight conditions.

Fourth, the self-assessment model of Article 6(4) places the classification burden on the provider, but the analysis in Chapters 3 and 4 has shown that several of the operative criteria – particularly the transparency, verifiability, and automation bias dimensions – depend heavily on the conditions of deployment, on which the provider can only make assumptions. The Act's current architecture addresses this partly through the deployer's obligations under Article 26 and the reclassification mechanism of Article 25(1)(c), but it does not require the provider to specify the deployment conditions under which the Article 6(3) assessment was made beyond the intended purpose "specific context and conditions of use" of Article 6(4) read in conjunction with Article 49(2), Article 71, Annex VIII A(5), and Article 3(12). I submit that — as part of

the Article 6(3) and (4) self-assessment and registration — there should be an explicit requirement on providers documenting context and conditions of use to specify in detail the assumed roles and capabilities of the humans in/on the loop,¹⁰³ the availability of verification mechanisms, and the organisational context. This would make the contextual dependencies of the classification transparent and enable deployers and supervisory authorities to assess whether the provider's assumptions hold in their specific context.¹⁰⁴

5.4 Residual uncertainties and directions for future research

The boundary between syntactic and semantic processing, while analytically useful, will require further calibration for natural language processing tasks that inherently involve engagement with meaning. The Recital (53) examples of preparatory tasks – text processing, translation, indexing – demonstrate that some degree of semantic engagement is compatible with the Article 6(3)(d) derogation, but the threshold beyond which semantic processing becomes evaluative remains underspecified. The Commission's forthcoming guidelines under Article 6(5) could usefully address this question by providing a graduated set of examples illustrating where the boundary lies for common language-processing tasks such as summarisation, entity extraction, sentiment classification, topic classification, and relevance ranking.

The risk of mission creep in general-purpose-capable AI systems — where a system assessed as performing a procedural or preparatory task is gradually repurposed for more substantive functions — warrants attention. The Article 6(3) assessment is made at a point in time, anchored to the provider's documented intended purpose. But the capabilities of modern AI systems, particularly those based on LLMs, often extend well beyond any single documented purpose, and evolve quickly. The existing mechanism of Article 25(1)(c), which reclassifies deployers as providers when they use a system beyond its intended purpose, provides a formal safeguard, but its practical effectiveness depends on the willingness and capacity of supervisory authorities to monitor actual deployment practices.

The framework proposed here has been finalised in early May 2026. It is developed on the basis of doctrinal analysis and information theory and tested against illustrative cases. It has not been subjected to empirical validation against a comprehensive taxonomy of real-world AI deployments across all Annex III areas. As enforcement practice develops and the Commission's

¹⁰³ See Crootof, Kaminski, and Price (2023) 473–487 and thesis footnotes 81 and 100.

¹⁰⁴ As Mahler and Schwemer (2024) 1143 point out, a market surveillance authority's evaluation of a self-classified not-highrisk AI system under Article 80(1) would typically start from the documentation registered under Article 6(4) in the EU database. A more granular documentation requirement would correspondingly enrich the basis both for that downstream review, and for the notification to the Commission when adopting restrictive measures through the Article 81 mechanism, see Tosoni and Esayas (2024) 1154–1155.

forthcoming Article 6(5) guidelines are adopted, the criteria may need to be refined, supplemented, or reconsidered. The framework is offered as an initial contribution to a discourse that will continue to evolve through the interaction of academic analysis, regulatory guidance, supervisory practice, and — in time — the jurisprudence of the Court of Justice.

Table of references

Literature

- Couneson (2024) Couneson, Guillaume. 2024. "Article 6: Classification Rules for High-Risk AI Systems." In Pehlivan, Forgó, and Valcke (eds) (2024), 193–218.
- Crootof, Kaminski, and Price (2023) Crootof, Rebecca, Margot E. Kaminski, and W. Nicholson Price II. 2023. "Humans in the Loop." *Vanderbilt Law Review* 76 (2): 429–510. <https://scholarship.law.vanderbilt.edu/vlr/vol76/iss2/2>.
- Ebers (2024) Ebers, Martin. 2024. "Truly Risk-Based Regulation of Artificial Intelligence: How to Implement the EU's AI Act." *European Journal of Risk Regulation* 16 (2): 684–703. doi:[10.1017/err.2024.78](https://doi.org/10.1017/err.2024.78).
- Enarsson, Enqvist, and Naarttijärvi (2022) Enarsson, Therese, Lena Enqvist, and Markus Naarttijärvi. 2022. "Approaching the Human in the Loop: Legal Perspectives on Hybrid Human/Algorithmic Decision-Making in Three Contexts." *Information & Communications Technology Law* 31 (1): 123–153. doi:[10.1080/13600834.2021.1958860](https://doi.org/10.1080/13600834.2021.1958860)
- Floridi (2011) Floridi, Luciano. 2011. *The Philosophy of Information*. Oxford: Oxford University Press. doi:[10.1093/acprof:oso/9780199232383.001.0001](https://doi.org/10.1093/acprof:oso/9780199232383.001.0001)
- Gallagher (2024) Gallagher, Daniel. 2024. "Article 26: Obligations of Deployers of High-Risk AI Systems." In Pehlivan, Forgó, and Valcke (eds) (2024), 537–552.
- Gasiola (2025) Gasiola, Gustavo Gil. 2025. "Rebuilding the Pyramid: The AI Act's Risk-Based Approach Using a Binary Decision Diagram." *Computer Law & Security Review* 58: 106189. doi:[10.1016/j.clsr.2025.106189](https://doi.org/10.1016/j.clsr.2025.106189).
- Granmar (2024) Granmar, C.G. 2024. "AI-Based Decision-Making and the Human Oversight Requirement Under the AI Act." In:

- Gill-Pedro, E. and A. Moberg (eds). *YSEC Yearbook of Socio-Economic Constitutions*, vol 2023. Springer, Cham. doi:[10.1007/16495_2024_68](https://doi.org/10.1007/16495_2024_68)
- Hanif et al (2024) Hanif, Hilmy, Jorge Constantino, Marie-Therese Sekwenz, Michel van Eeten, Jolien Ubacht, Ben Wagner, and Yury Zhauniarovich. 2024. “Navigating the EU AI Act Maze Using a Decision-Tree Approach.” *ACM Journal on Responsible Computing* 1 (3): 1–16. doi:[10.1145/3677174](https://doi.org/10.1145/3677174)
- Laux and Ruschemeier (2025) Laux, Johann, and Hannah Ruschemeier. 2025. “Automation Bias in the AI Act: On the Legal Implications of Attempting to De-Bias Human Oversight of AI.” *European Journal of Risk Regulation* 16 (4): 1519–1534. doi:[10.1017/err.2025.10033](https://doi.org/10.1017/err.2025.10033)
- Lazcoz and de Hert (2023) Lazcoz, Guillermo, and Paul de Hert. 2023. “Humans in the GDPR and AIA Governance of Automated and Algorithmic Systems: Essential Pre-Requisites Against Abdicating Responsibilities.” *Computer Law & Security Review* 50: 105833. doi:[10.1016/j.clsr.2023.105833](https://doi.org/10.1016/j.clsr.2023.105833)
- Lenaerts and Gutiérrez-Fons (2013) Lenaerts, Koen, and José A. Gutiérrez-Fons. 2013. “To Say What the Law of the EU Is: Methods of Interpretation and the European Court of Justice.” EUI AEL Distinguished Lectures of the Academy 2013/09. Florence: European University Institute. <https://hdl.handle.net/1814/28339>
- Levitina (2024) Levitina, Anna. 2024. “Humans in Automated Decision-Making Under the GDPR and AI Act.” *Revista CIDOB d'Afers Internacionals* 138: 121–143. doi:[10.24241/rcai.2024.138.3.121/](https://doi.org/10.24241/rcai.2024.138.3.121/) English text: <https://www.cidob.org/en/publications/humans-automated-decision-making-under-gdpr-and-ai-act>
- Mahler (2022) Mahler, Tobias. 2022. “Between Risk Management and Proportionality: The Risk-Based Approach in the EU's Artificial Intelligence Act Proposal.” In *The Swedish Law and Informatics Research Institute*, 247–270.

doi:[10.53292/208f5901.38a67238](https://doi.org/10.53292/208f5901.38a67238). Also available at
SSRN: <https://ssrn.com/abstract=4001444>

- Mahler and Schwemer (2024) Mahler, Tobias, and Sebastian Felix Schwemer. 2024. “Article 80: Procedure for Dealing with AI Systems Classified by the Provider as Non-High-Risk in Application of Annex III.” In Pehlivan, Forgó, and Valcke (eds) (2024), 1140–1148.
- Tosoni and Esayas (2024) Tosoni, Luca, and Samson Y. Esayas. 2024. “Article 81: Union Safeguard Procedure.” In Pehlivan, Forgó, and Valcke (eds) (2024), 1149–1157.
- Panezi (2024) Panezi, Argyri. 2024. “Article 14: Human Oversight.” In Pehlivan, Forgó, and Valcke (eds) (2024), 351–371.
- Parasuraman and Manzey (2010) Parasuraman, Raja, and Dietrich H. Manzey. 2010. “Complacency and Bias in Human Use of Automation: An Attentional Integration.” *Human Factors* 52, no. 3: 381–410. doi:[10.1177/0018720810376055](https://doi.org/10.1177/0018720810376055)
- Pehlivan, Forgó, and Valcke (eds) (2024) Pehlivan, Ceyhun Necati, Nikolaus Forgó, and Peggy Valcke (eds.). 2024. *The EU Artificial Intelligence (AI) Act: A Commentary*. Alphen aan den Rijn: Wolters Kluwer. ISBN 978-94-035-3227-1.
- Riede and Talhoff (2024) Riede, Lutz, and Oliver Talhoff. 2024. “Article 25: Responsibilities Along the AI Value Chain.” In Pehlivan, Forgó, and Valcke (eds) (2024), 524–536.
- Smuha and Yeung (2025) Smuha, Nathalie A., and Karen Yeung. 2025. “The European Union's AI Act: Beyond Motherhood and Apple Pie?” In *The Cambridge Handbook of the Law, Ethics, Policy and Artificial Intelligence*, edited by Nathalie A. Smuha, 228–258. Cambridge: Cambridge University Press. doi:[10.1017/9781009367783.015](https://doi.org/10.1017/9781009367783.015)
- Veale and Borgesius (2021) Veale, Michael, and Frederik Zuiderveen Borgesius. 2021. “Demystifying the Draft EU Artificial Intelligence Act.” *Computer Law Review International* 22 (4): 97–112.

doi:[10.9785/cri-2021-220402](https://doi.org/10.9785/cri-2021-220402).

arXiv:<https://arxiv.org/abs/2107.03721>

Wachter, Mittelstadt, and Floridi (2017)

Wachter, Sandra, Brent Mittelstadt, and Luciano Floridi. 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:[10.2139/ssrn.2903469](https://doi.org/10.2139/ssrn.2903469)

Legal sources

Legislation and treaties

Treaty on European Union (TEU)

Consolidated Version of the Treaty on European Union. OJ C 326, 26.10.2012. https://eur-lex.europa.eu/eli/treaty/teu_2012/oj/eng

Treaty on the Functioning of the European Union (TFEU)

Consolidated Version of the Treaty on the Functioning of the European Union. OJ C 326, 26.10.2012. http://data.europa.eu/eli/treaty/tfeu_2012/oj

Regulation (EU) 2016/679

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 (General Data Protection Regulation). OJ L 119, 4.5.2016. <https://eur-lex.europa.eu/eli/reg/2016/679/oj/>

Regulation (EU) 2024/1689

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act). OJ L 2024/1689. <http://data.europa.eu/eli/reg/2024/1689/oj>

Case law

Case C-283/81 (CILFIT)	Case 283/81, <i>CILFIT and Lanificio di Gavardo SpA v Ministry of Health</i> , ECLI:EU:C:1982:335. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:61981CJ0283
Case C-66/85 (Lawrie-Blum)	Case 66/85, <i>Lawrie-Blum v Land Baden-Württemberg</i> , ECLI:EU:C:1986:284. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:61985CJ0066
Case C-162/97 (Nilsson)	Case C-162/97, <i>Criminal Proceedings Against Nilsson and Others</i> , ECLI:EU:C:1998:554. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:61997CJ0162
Case C-268/06 (Impact)	Case C-268/06, <i>Impact v Minister for Agriculture and Food and Others</i> , ECLI:EU:C:2008:223. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62006CJ0268
Case C-226/11 (Expedia)	Case C-226/11. <i>Expedia Inc. v Autorité de la concurrence and Others</i> , ECLI:EU:C:2012:795. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62011CJ0226
Case C-201/13 (Deckmyn)	Case C-201/13, <i>Deckmyn and Vrijheidsfonds v Vandersteen and Others</i> , ECLI:EU:C:2014:2132. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:62013CJ0201
Case C-634/21 (SCHUFA)	Case C-634/21, <i>OQ v Land Hessen (SCHUFA Holding)</i> , ECLI:EU:C:2023:957. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:62021CJ0634

Institutional documents and official guidance

Article 29 Working Party (2018)	Article 29 Data Protection Working Party. 2018. “Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679.”
---------------------------------	---

	WP251rev.01. Adopted 3 October 2017; last revised 6 February 2018. Brussels. https://ec.europa.eu/newsroom/article29/items/612053
Council of the European Union (2022)	Council of the European Union. 2022. “Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence—General Approach.” ST 14954/22. Brussels, 25 November 2022. https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf
Council of the European Union (2024)	Council of the European Union. 2024. “Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence—Analysis of the Final Compromise Text with a View to Agreement.” ST 5662/24. Brussels, 26 January 2024. https://data.consilium.europa.eu/doc/document/ST-5662-2024-INIT/en/pdf
Council of the European Union (2026)	Council of the European Union. 2026. “Proposal for a Regulation Amending Regulations (EU) 2024/1689 and (EU) 2018/1139 as Regards the Simplification of the Implementation of Harmonised Rules on Artificial Intelligence (Digital Omnibus on AI)—Mandate for Negotiations with the European Parliament.” ST 7322/26. Brussels, 13 March 2026. https://data.consilium.europa.eu/doc/document/ST-7322-2026-INIT/en/pdf
European Commission (2025) (Digital Omnibus on AI)	European Commission. 2025. “Proposal for a Regulation Amending Regulations (EU) 2024/1689 and (EU) 2018/1139 as Regards the Simplification of the Implementation of Harmonised Rules on Artificial Intelligence (Digital Omnibus on AI).” COM(2025) 836 final. Brussels. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52025PC0836
European Data Protection Board and European Data Protection Supervisor (2026)	European Data Protection Board and European Data Protection Supervisor. 2026. “EDPB–EDPS Joint Opinion 1/2026 on the Proposal for a Regulation as Regards the Simplification of the Implementation of Harmonised Rules on Artificial Intelligence (Digital Omnibus on AI).” 21 January 2026. https://www.edpb.europa.eu/our-work-

[tools/our-documents/edpbedps-joint-opinion/edpb-edps-joint-opinion-12026-proposal_en](https://www.european-council.europa.eu/media/assets/5/5F/Statement_of_Opinion_Edinburgh_2023.pdf)

- European Parliament (2023) European Parliament. 2023. “Draft Compromise Amendments on the Proposal for a Regulation on Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act).” Consolidated compromise text, IMCO-LIBE. Brussels, 9 May 2023. https://www.europarl.europa.eu/meet-docs/2014_2019/plmrep/COMMIT-TEES/CJ40/DV/2023/05-11/ConsolidatedCA_IMCOLIBE_AI_ACT_EN.pdf
- European Parliament (2026) European Parliament. 2026. "Amendments adopted by the European Parliament on 26 March 2026 on the proposal for a regulation of the European Parliament and of the Council amending Regulations (EU) 2024/1689 and (EU) 2018/1139 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI) (COM(2025)0836 – C10-0304/2025 – 2025/0359(COD)).” 26 March 2026. https://www.europarl.europa.eu/doceo/document/TA-10-2026-0098_EN.html

Websites and online materials

- Microsoft Enterprise AI Services Code of Conduct (2025) Microsoft. 2025. “Enterprise AI Services Code of Conduct.” <https://learn.microsoft.com/en-us/legal/ai-code-of-conduct>
- Oxford English Dictionary (OED) Oxford English Dictionary. 2025. Entries cited:
 “assessment (n.), sense 5.a,”
 doi:[10.1093/OED/8740411998](https://doi.org/10.1093/OED/8740411998)
 “narrow (adj.), sense 3.a,” doi:[10.1093/OED/8863877943](https://doi.org/10.1093/OED/8863877943)
 “preparatory (adj.), sense 1,”
 doi:[10.1093/OED/7056784451](https://doi.org/10.1093/OED/7056784451)

“procedure (n.), sense 1.a,” doi:[10.1093/OED/3805511077](https://doi.org/10.1093/OED/3805511077)

“task (n.), sense 2.a,” doi:[10.1093/OED/8723383921](https://doi.org/10.1093/OED/8723383921)

Appendix: Applying the Article 6(3)(a) and (d) criteria to illustrative examples

Application to illustrative examples

To demonstrate the practical application and test the robustness of the proposed criteria, this section applies the framework to illustrative cases drawn from Annex III sectors. These examples are not intended as empirical claims about industry practice but as analytical exercises designed to concretise and stress-test the criteria.

Document translation for asylum assessment

Case: *An AI system is provided to translate asylum applicants' written statements to the language of the receiving state. The translations serve as input to the eligibility assessment performed by the asylum case officer.*

Criterion		Assessment
0.	<i>Is the intended use within an Annex III area?</i>	Asylum application examination is part of Annex III area 7.
1.	<i>Does the task involve profiling of natural persons?</i>	The system processes personal data but does not evaluate personal aspects; translation is a linguistic operation, not an assessment of the person.
2.	<i>Is the task specific, defined, and bounded?</i>	The task is specific and bounded: input is a document in the applicant's native language; output is the same document in the target language.
3.	<i>Is the task syntactic or semantic? If semantic, is it preparatory to an assessment?</i>	Translation involves semantic processing (engagement with meaning) but does not generate evaluative conclusions about the applicant's credibility, the merits of their claim, or any other aspect relevant to the following assessment. It is not merely procedural, but it is purely preparatory.
4.	<i>Does the task output bear on the substance</i>	The output (a translated document) makes information accessible but does not alter its substantive content; the same information is available in both languages. At first glance the counterfactual test is met: minor differences in translation approach

	<i>of an assessment or decision?</i>	would not change the underlying facts. On the other hand, important linguistic nuances might, as they say, be “lost in translation”.
5.	<i>Is the task output transparent and verifiable?</i>	This is where the analysis requires contextual sensitivity. If the case officer has access to the native language proficiency – either personally or through a qualified interpreter – the translation is verifiable and the criterion is met if it can be demonstrated that such access is not illusory. If, however, the case officer must rely exclusively on the AI translation without any means of verification, the output becomes functionally determinative of the informational basis for the assessment. In the latter scenario, the criterion is not satisfied. ¹⁰⁵
6.	<i>Is the task output inherently resistant to uncritical reliance?</i>	A translated document presents factual content without evaluative framing and is inherently resistant to automation bias, provided the assessor is aware that the text is AI-translated, and has mandate and access to native language proficiency in cases of doubt.
Conclusion:		The system may qualify for the Article 6(3)(d) carve-out, but only where the conditions of deployment ensure practical verifiability of the translation. The same AI system may qualify in one deployment context and not in another – a result consistent with the task-centred and context-dependent nature of the Article 6(3) assessment.

Selecting job interview candidates: a multi-stage recruitment workflow

Job applicant data extraction

Case: *The AI system extracts structured data fields from CVs and cover letters: names, contact details, educational qualifications, dates of employment, certifications, and language proficiencies.*

Criterion	Assessment
-----------	------------

¹⁰⁵ See the discussion in Section 3.3.2, where the same example is used to illustrate the contextual dimension of the verifiability requirement.

0.	<i>Is the intended use within an Annex III area?</i>	Recruitment is part of Annex III area (4)(a).
1.	<i>Does the task involve profiling of natural persons?</i>	The system processes personal data but does not evaluate personal aspects; it performs data extraction and conversion, not assessment of the person. The profiling threshold is not engaged.
2.	<i>Is the task specific, defined, and bounded?</i>	The task is specific (extract specified data), bounded (input: CVs and cover letters; output: structured data fields), and narrow.
3.	<i>Is the task syntactic or semantic? If semantic, is it preparatory to an assessment?</i>	The operation is strictly syntactic: it maps unstructured text to structured data fields using formal parsing rules. It precedes a following, functionally separate assessment. It is both procedural and preparatory.
4.	<i>Does the task output bear on the substance of an assessment or decision?</i>	The output (a structured set of data) does not bear on the substance of the hiring decision; the same substantive information is available to the assessor in the original CV.
5.	<i>Is the task output transparent and verifiable?</i>	The output is transparent (the assessor understands that data has been extracted) and verifiable (the assessor can check each data point against the source documents).
6.	<i>Is the task output inherently resistant to uncritical reliance?</i>	A structured set of data does not invite uncritical reliance; it presents factual data without suggesting evaluative conclusions.
Conclusion:		The system satisfies all six criteria and qualifies for both Article 6(3) carve-out (a) and (d).

Score candidates against clear, rules-based criteria

Case: *The AI system assigns numerical scores to candidates based on transparent, predefined criteria: for example, awarding points for each year of relevant work experience, for specific educational degrees, and for listed language proficiencies. The scoring formula is crafted by natural persons, documented, fixed, and deterministic.*

Criterion	Assessment
1. <i>Does the task involve profiling of natural persons?</i>	Despite the mechanical nature of the scoring, the system uses personal data to evaluate aspects of the natural person – their qualifications and experience – in relation to the job requirements. This constitutes profiling within GDPR Article 4(4). The assessment terminates here.
Conclusion:	The Article 6(3) carve-outs do not apply. The fact that the scoring formula is transparent and rule-based does not change its characterisation as profiling in the context of Article 6(3): The relevant question under the profiling override is not whether the evaluation is sophisticated or opaque, but whether it constitutes an automated evaluation of personal aspects – which it plainly does. It might seem counterintuitive that such clear, rules-based processing should not be covered by a derogation. This use case could in fact be covered by another AI Act “safety valve”, that of the definition of “AI system” in itself: Recital (12) specifies that the notion of “AI system” should not cover systems “that are based on the rules defined solely by natural persons to automatically execute operations”. In other words, such scoring tasks can be automated outside the scope of the AI Act by assigning them to separate systems without machine learning components or other features covered by the “AI system” definition.

Score candidates against ambiguously defined criteria

Case: *The AI system evaluates candidates against soft criteria such as “leadership potential,” “cultural fit,” or “communication skills,” producing a qualitative or numerical assessment.*

The analysis is identical to the previous sub-task in its engagement of the profiling override, but the additional ambiguity of the criteria compounds the concerns at every subsequent level. Even setting aside the profiling threshold, the task fails at Criterion 3 (the operation is semantic, requiring evaluative judgment about the person), Criterion 4 (the output bears directly on the substance of the hiring decision), Criterion 5 (the output is opaque – the basis for a “cultural

fit” score cannot meaningfully be verified by the human assessor), and Criterion 6 (an evaluative score on ambiguous criteria is maximally susceptible to automation bias). Likewise, as such soft criteria are all but impossible to capture in formal rules designed solely by natural persons, the above-mentioned “AI system”-definition escape is closed.

Create a final ranking of candidates

Case: *The AI system produces a ranked list of candidates for the recruiter’s consideration.*

This is the composite output of the preceding scoring stages. Whether the ranking is based on rule-based scores, ambiguous criteria, or a combination, the system evaluates personal aspects of natural persons and produces an output that directly structures the hiring decision. The profiling override applies. Even if the profiling override were not applicable, the system would fail at Criterion 3 (the ranking involves semantic evaluation of candidates’ suitability), Criterion 4 (the ranking bears on the substance of the hiring decision by directing the recruiter’s attention), and Criterion 6 (a relevance ranking is precisely the kind of output that invites uncritical reliance).

Observations from applying the criteria to the recruitment workflow

The recruitment workflow example demonstrates a characteristic gradient: the earliest stages (data extraction) are genuinely procedural or preparatory and may qualify for a carve-out, while each successive stage introduces evaluative elements that move the system further from the procedural or preparatory domain and into the territory of substantive assessment. The example also illustrates the importance of the profiling override as a decisive threshold: the moment an AI system begins evaluating the suitability, qualifications, or personal characteristics of candidates – even through transparent, rule-based criteria – it constitutes profiling and is categorically high-risk.¹⁰⁶

Writing a medical journal: a multi-stage clinical documentation workflow

Transcribe audio recordings to text

Case: *An AI system converts the audio recording of a medical consultation into a written transcript.*

¹⁰⁶ The recruitment example thus confirms that the transition from carve-out eligible to non-eligible may occur at a very early stage in the workflow. When processing data about natural persons only the most mechanical extraction and structuring sub-tasks survive the framework’s initial “filter”: From the moment the system begins to summarise, rank, or evaluate candidates – even by reference to predetermined criteria – it crosses the threshold into evaluative territory, and the profiling override removes the Article 6(3) carve-out from the picture entirely.

Criterion		Assessment
0.	<i>Is the intended use within an Annex III area?</i>	Clinical documentation may contribute to the evaluation of the eligibility of natural persons for healthcare services – Annex III area (5)(a) – and in risk assessment in life and health insurance – Annex III area (5)(c). AI systems generating clinical documentation might also be products covered by Article 6(1) and Annex I A(11) – such systems are out of scope for the Article 6(3) derogations.
1.	<i>Does the task involve profiling of natural persons?</i>	The system processes personal data (the patient's voice, name, and health-related statements) but does not evaluate personal aspects; it converts speech to text without assessing the person's medical condition or any other personal characteristic.
2.	<i>Is the task specific, defined, and bounded?</i>	The task is specific and bounded (input: consultation audio recording; output: text transcription of the recording).
3.	<i>Is the task syntactic or semantic? If semantic, is it preparatory to an assessment?</i>	The operation involves semantic processing – converting spoken language to written form requires engagement with meaning – but does not generate evaluative conclusions about the patient's health or treatment.
4.	<i>Does the task output bear on the substance of an assessment or decision?</i>	The transcript makes information accessible in a different modality without altering its content; the same information is available in the original audio recording.
5.	<i>Is the task output transparent and verifiable?</i>	The output is transparent and verifiable: the physician can listen to the original recording and check the transcript. This verification capacity is critical; if the audio was not stored, or deleted before verification, the criterion would require reassessment.
6.	<i>Is the task output inherently resistant to uncritical reliance?</i>	A verbatim transcript does not invite uncritical reliance in the way that a summary or analysis would.

Conclusion:	The Article 6(3)(d) carve-out applies, provided that the original audio remains accessible for verification. This is analogous to the document translation example above.
--------------------	---

Classify speakers (doctor, patient(s))

Case: *The AI system identifies and labels which portions of the transcript were spoken by the doctor and which by the patient(s).*

Criterion		Assessment
1.	<i>Does the task involve profiling of natural persons?</i>	The system processes personal data (voice characteristics) but does not evaluate personal aspects of the patient in any meaningful sense; speaker identification is a formal classification based on acoustic or contextual features, not an assessment of the person's health, behaviour, or characteristics.
2.	<i>Is the task specific, defined, and bounded?</i>	The task is specific and bounded (input: consultation audio recording; output: text transcription of the recording with speaker labels).
3.	<i>Is the task syntactic or semantic? If semantic, is it preparatory to an assessment?</i>	The operation involves a significant degree of semantic processing (audio to text, distinguishing speakers by vocal characteristics) but the classification is formal (who spoke) rather than evaluative (what the speech means for the patient's condition).
4.	<i>Does the task output bear on the substance of an assessment or decision?</i>	Speaker labels do not alter the substantive content of the transcript; they merely make it more accessible for a later assessment.
5.	<i>Is the task output transparent and verifiable?</i>	The output is verifiable – the physician, who was present at the consultation, can readily confirm speaker attributions.
6.	<i>Is the task output inherently resistant to uncritical reliance?</i>	Speaker labels do not invite uncritical reliance in the medical assessment.

Conclusion:	While the degree of semantic processing is larger than for the transcription subtask, the Article 6(3)(d) carve-out still applies, again provided that the original audio remains accessible for verification.
--------------------	--

Classify message content (symptoms, diagnoses, medical advice, prescription references)

Case: *The AI system assigns semantic labels to segments of the transcript, categorising statements as descriptions of symptoms, diagnostic conclusions, therapeutic advice, medication references, and so on.*

Criterion		Assessment
1.	<i>Does the task involve profiling of natural persons?</i>	The system uses the patient’s personal data to classify statements about their medical condition. Categorising a patient’s statement as a “symptom” involves some evaluation of the medical significance of what the patient has said – depending on the level of detail this could approach the threshold of assessing a person’s “health” within the meaning of GDPR Article 4(4). A binary “symptom: True/False” classification should pass below the threshold, for more detailed “symptom of X” classifications the profiling override will soon apply. ¹⁰⁷
2.	<i>Is the task specific, defined, and bounded?</i>	The task is specific and bounded (input: consultation audio recording; output: text transcription of the recording labelled by speaker and statement type).
3.	<i>Is the task syntactic or semantic? If semantic, is it preparatory to an assessment?</i>	Classifying whether a physician’s statement constitutes a “diagnosis” versus “medical advice” requires significant semantic understanding of medical content – a judgment about the evaluative significance of information. Still, the task is at its core preparatory in nature, making no evaluation of the significance or correctness of the labelled statements.

¹⁰⁷ Even if such classification is based on keywords rather than clinical reasoning, such a system would evaluate information about a patient’s health. GDPR Article 4(4) specifies this as an example of profiling when such processing is used to analyse or predict aspects concerning a natural person’s health.

4.	<i>Does the task output bear on the substance of an assessment or decision?</i>	Like speaker labels, statement classifications primarily make the journal transcript more accessible for a following assessment. The theoretical possibility of scenarios where such classification structures the journal in a way that shapes how subsequent readers (including the consulting physician themselves) interpret the consultation must be balanced against <i>effet utile</i> .
5.	<i>Is the task output transparent and verifiable?</i>	The output is verifiable – the physician, who was present at the consultation, can confirm or reject the statement classifications.
6.	<i>Is the task output inherently resistant to uncritical reliance?</i>	Again, like speaker labels, high level statement classification (“symptom”, “diagnosis”, “medication”) does not replace medical assessment of the statement’s content.
Conclusion:		The task is clearly not procedural and challenges the boundaries of the Article 6(3)(d) “preparatory” derogation. Still, given sufficiently high-level classification categories (“symptom”, “diagnosis”, “medication”) and responsible users, an assessment of no significant risk of harm is defensible. On the other hand, a system performing more detailed classification (“symptom of X”, “diagnosis of illness Y”, “recommendation of medication Z”) would fail multiple criteria, including the initial “profiling” filter.

Remove non-relevant content (e.g. interruptions, off-topic discussion)

Case: *The AI system identifies and removes portions of the transcript deemed irrelevant to the medical record – social pleasantries, interruptions, digressions about unrelated topics.*

Criterion		Assessment
1.	<i>Does the task involve profiling of natural persons?</i>	The profiling filter may or may not be engaged depending on how the system determines “non-relevance”: if the system merely predicts the degree of a statement’s general “medical relevance”, making no evaluation of the medical content of the statement, it should avoid the profiling threshold.

2.	<i>Is the task specific, defined, and bounded?</i>	The task is specific and bounded (input: consultation audio recording; output: text transcription of the parts of the recording classified as medically relevant).
3.	<i>Is the task syntactic or semantic? If semantic, is it preparatory to an assessment?</i>	Predicting medical relevance is a non-procedural task requiring significant semantic understanding of medical content. Furthermore, non-relevance classification includes an evaluation of the significance of statements that challenges the upper bounds of what can be defended as purely preparatory.
4.	<i>Does the task output bear on the substance of an assessment or decision?</i>	The output is not a neutral data transformation; it is an editorial act that determines what information the physician and subsequent readers will and will not see in the final transcript. If the system incorrectly classifies a patient's mention of a symptom as "off-topic" and removes it, that information is lost from the medical record – a consequence that bears directly on the substance of subsequent assessments. The counterfactual test of Criterion 4 is met: a time-pressed physician's assessment could differ depending on what the system chose to include or exclude, what the physician themselves remembers from the consultation, and how much time they feel they can afford to spend comparing the edited transcription to the raw transcription. This terminates the case assessment.
	Conclusion:	Article 6(3) carve-outs do not apply. Selective removal of content from a draft medical record is an editorial function that bears on the substance of the medical documentation and, by extension, of the assessments based on it.

Score confidence of classifications to prioritise human review

Case: *The AI system assigns a confidence score to each of its classifications (from sub-tasks above) and uses these scores to flag portions of the journal for priority human review.*

This sub-task has a dual character. On one hand, the confidence scoring is a meta-operation on the system's own outputs – not a direct evaluation of the patient, or the medical content of transcribed statements. On the other hand, the practical effect of the confidence scoring is to direct the physician's attention: high-confidence classifications may receive cursory review, while low-confidence classifications are flagged for closer scrutiny. This attention-directing

function engages Criterion 6: the confidence score creates a cognitive shortcut that incentivises the physician to defer to the system's own assessment of which parts of the journal require careful review. The system is telling the human overseer where to look – and, implicitly, where not to look. This is precisely the kind of cognitive pre-structuring that the Article 14 safeguards are designed to counteract.¹⁰⁸

Conclusion: Article 6(3) carve-outs do not apply. The confidence scoring directs the human assessor's attention in a manner that bears on the quality and thoroughness of the medical review.

Observations from applying the criteria to the clinical documentation workflow

The speaker-labelling and statement-classification subtasks illustrate that an AI system may perform semantically rich operations and still qualify under Article 6(3)(d), provided that the task remains genuinely preparatory and does not steer the subsequent medical assessment. By contrast, the confidence-scoring subtask shows the boundary of that conclusion. A system that labels speakers or statement types without confidence scores may remain within the carve-out, while a system returning the same labels with confidence scores may fall outside it, because the scores direct the physician's attention and invite reliance on the system's own assessment of reliability. The result is counterintuitive: providing the physician more information shifts the system into a higher risk category. In my view, *de lege lata* the distinction is defensible. Article 6(3) is concerned not with the volume of information an AI system produces, but with whether the form of that information meaningfully shifts cognitive effort from the human assessor to the system. A bare classification leaves the assessor in the position of judging whether to trust each label on its own evidentiary footing; a confidence-scored classification invites the assessor to partially delegate that judgment to the system. The Article 14 design requirements are calibrated to the latter case, and so is the high-risk regime as a whole. The *de lege ferenda* question of whether providers should be incentivised differently, for example, by allowing a graduated regime that admits transparent confidence reporting under stricter human-oversight conditions, is taken up in Section 5.4.

¹⁰⁸ The confidence score illustrates the Article 14 paradox discussed in Section 4.2.6: a system output that is nominally designed to support human oversight may in practice structure and direct the human decision-maker's attention. By flagging which sections require review, the confidence score defines the scope of the physician's scrutiny, creating the very automation bias it purports to mitigate.