CLARA: A Hybrid LLM-Rule System for EU AI Act Risk Classification

Vasco Alves[1], **André Carreiro**[1], **Iakovina Kindylidi**[2, 3]
[1]Fraunhofer AICOS, Portugal
[2]NOVA School of Law, Portugal
[3]Vieira de Almeida & Associados, Portugal
{vasco.gomes,andre.carreiro}@fraunhofer.pt, iakovina.kindylidi@novalaw.unl.pt

Abstract

Interpreting and operationalizing the European Union Artificial Intelligence Act poses a dual challenge: it demands both technical understanding of AI systems and legal interpretation of a complex regulatory framework. Manual risk classification is costly, inconsistent, and difficult to scale. This work presents CLARA, a hybrid system that combines large language models (LLMs) with symbolic rule reasoning to support organizations in the first step of self-assessing the regulatory risk tier of their AI systems. CLARA processes free-form textual descriptions of AI systems, retrieves relevant provisions and guidelines, and evaluates them through two complementary approaches: (1) an LLM-only semantic reasoning pipeline, and (2) a neurosymbolic pipeline where LLMs identify condition matches and a deterministic rule engine produces the final decision. We demonstrate CLARA through an interactive web interface that allows the user to visualize evidence retrieval, rule evaluation, and explainable classifications. Preliminary experiments using a set of example system descriptions constructed by a legal expert suggest that hybrid reasoning improves interpretability and robustness compared to purely generative approaches. The demo highlights how AI and Law can be effectively bridged through transparent, legally-grounded reasoning.

Introduction

The European Union Artificial Intelligence Act (AI Act) is the world's first comprehensive horizontal legal framework for regulating artificial intelligence (AI). It applies across sectors, covering the entire AI value chain, from model providers and providers of AI systems to deployers (users of the technology in their professional capacity), distributors, and importers, and does so with extraterritorial reach, extending its obligations to actors outside the EU when their AI systems are placed on, or their outputs are used in the European market (European Union 2024).

Rather than regulating AI technologies as a whole, the AI Act adopts a risk-based approach, tailoring legal obligations to the level of risk an AI system poses to health, safety, or fundamental rights. In this context, the AI Act regulates systems falling in one of the risk tiers illustrated in Figure 1.

The Act places its heaviest compliance burden on providers and deployers of high-risk AI systems, while setting transparency and cooperation duties for all other actors. Its ultimate objective is to foster trust in AI by establishing obligations and best practices, supporting innovation through legal certainty rather than prescriptive control of technology design.

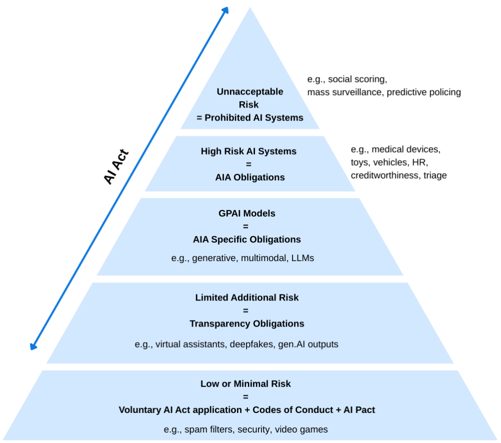

AI Act
Unacceptable Risk = Prohibited AI Systems
e.g., social scoring, mass surveillance, predictive policing
High Risk AI Systems = AIA Obligations
e.g., medical devices, toys, vehicles, HR, creditworthiness, triage
GPAI Models = AIA Specific Obligations
e.g., generative, multimodal, LLMs
Limited Additional Risk = Transparency Obligations
e.g., virtual assistants, deepfakes, gen.AI outputs
Low or Minimal Risk = Voluntary AI Act application + Codes of Conduct + AI Pact
e.g., spam filters, security, video games

Figure 1: AI Act risk levels

The AI Act is a complex legal instrument, designed for heterogeneous audiences, but it can often be difficult for engineers, researchers, legal practitioners, and companies to apply in practice. Hence, although the Act's text provides the legal scaffolding, organizations face challenges when seeking to determine, for a given AI system, (i) whether the Act applies, (ii) which risk tier it falls under, and (iii) which obligations must be implemented (or which prohibitions avoided).

To enable meaningful compliance and informed design choices, there is a growing need for simplified and, in some cases, machine-interpretable representations of its rules, capable of translating dense legal requirements into actionable, technically grounded guidance. Such tools can also support the effective implementation and interpretability of the Regulation, reinforcing its potential to evolve into a global standard for trustworthy AI governance.

To date, most AI Act compliance tools rely on Natural Language Processing (NLP) models, which probabilistically infer answers from text (such as the AI Act and the Commission's guidelines), but frequently lack the accuracy, explainability, or verifiability required for legal reasoning. In contrast, rule-based or symbolic models provide determinism and traceable logic, but often struggle with the wide variety of system descriptions and free-text inputs encountered in practice.

Here, we present CLARA, a hybrid reasoning framework combining LLM-based semantic understanding with symbolic rule-based legal logic. CLARA enables organizations to upload free-form descriptions of AI systems and classify the system into risk tiers under two complementary pipelines: (1) an LLM+RAG pipeline relying on retrieval-based reasoning using relevant provisions of the Act and guidelines, and (2) a neurosymbolic pipeline in which encoded legal rules are matched via LLM-semantic evaluation and combined via a deterministic rule engine for final classification and explanation. The system thereby bridges the gap between legal interpretation and computational reasoning, offering a transparent, verifiable, and scalable approach to self-assessment under the AI Act.

Background and Related Work

Several tools have recently emerged to support organizations in interpreting and complying with the AI Act. The AI Act Service Desk, launched by the European Commission, offers an official Compliance Checker based on a fixed decision tree of predefined questions to determine whether and how the Act applies to a given system (European Commission 2025). Similar questionnaire-driven approaches, such as the Future of Life Institute's web-based compliance checker (Future of Life Institute 2025) and commercial solutions like Securiti's EU AI Act assessment tool (Baig, Qayyum, and Khan 2025), provide straightforward outcomes but require users to interpret legal terms and self-categorize (e.g., as "provider" or "deployer"), which is often difficult in complex AI value chains (Leong and Judge-Raza 2025).

These static, rule-based paths are easy to follow but limited in adaptability. They struggle to handle ambiguous cases or explain borderline classifications, and they require frequent manual updates as the regulatory landscape evolves. To address such limitations, recent research prototypes leverage LLMs for dynamic legal reasoning. For example, Kovari et al. (2025) introduce an AI-powered chatbot that conversationally interprets the AI Act, while Davvetas et al. (2025) present a retrieval-augmented tool that automatically determines risk categories with minimal user input.

CLARA positions itself as a next-generation compliance assistant that unifies both paradigms. Unlike fixed questionnaires or purely generative chatbots, it combines semantic LLM reasoning with deterministic, expert-authored legal rules, producing explainable and verifiable classifications directly from free-form AI system descriptions.

CLARA Overview

CLARA follows a three-step reasoning pattern aligned with the AI Act: (i) decide whether the described artifact qualifies as an "AI system" under the Act; (ii) determine whether it falls within or outside the material scope of the Regulation (including specific exceptions and exclusions); and (iii) for in-scope systems, suggest a risk tier (prohibited, high, limited, minimal/none, or undetermined), together with a short explanation and links to the relevant legal provisions. The two pipelines differ in how they implement this workflow: the first relies entirely on LLM-based reasoning with RAG, while the second combines LLM semantic matching with a deterministic rule engine. Figure 2 illustrates both workflows until an assessment is reached.

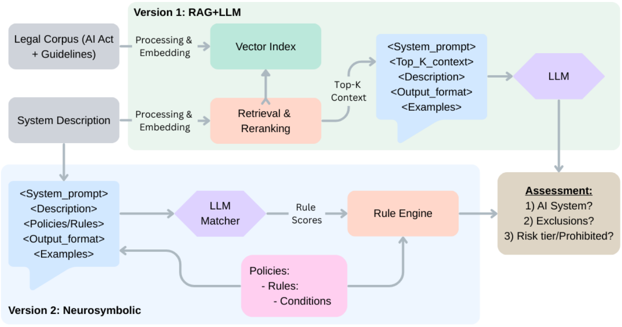


Figure 2: Diagram illustrating CLARA's workflow for the two versions considered: 1) LLM+RAG, 2) Neurosymbolic

Input and Preprocessing

The entry point to our system is a free-form description of an AI system, which can be inserted manually as text or submitted as a PDF or DOCX file. Although the system can ask follow-up questions, the description should be as complete and detailed as possible, including important aspects such as the system's intended purpose and the context in which it will be used.

Version 1: LLM+RAG Pipeline

Corpus preparation. The legal corpus comprises the AI Act (European Union 2024), the Guidelines on the definition of an AI system (European Comission 2025b), and the Guidelines on prohibited AI practices (European Comission 2025a), all of which are available in English. We preprocess the consolidated HTML of the AI Act into paragraph-level LangChain `Documents` enriched with legal metadata (source, article/annex, paragraph, section). Articles and annexes are segmented with lightweight heuristics (e.g., numeric/lettered enumerations, table rows) and softly split into N-character chunks to preserve local coherence while remaining LLM-friendly. For the EC guidelines, the JSON objects representing the sections are also parsed and converted into `Documents`.

Retrieval and reranking. We build an in-memory dense index and retrieve a candidate pool with similarity search. An optional cross-encoder reranker orders passages by query relevance. To preserve legal context, selected chunks are bundled by reference and paragraph and concatenated as contiguous "context blocks."

Prompting and structured output. The LLM is prompted with (i) the bundled legal context and (ii) the user description, and must emit a JSON object conforming to a typed schema: an `Assessment` with a single best tier, a short summary, and a list of `EvidenceItems` (citation ID, user-span, risk candidate, and an optional follow-up question). The prompt enforces conservatism (return `undetermined` at low confidence), forbids inventing legal references, and encourages follow-up when the context is insufficient. Additionally, a timestamped PDF report can be exported for record-keeping.

Scope and limitations. This pipeline offers breadth without requiring pre-encoded rules, but inherits LLM/RAG limitations (context sensitivity, potential over-reliance on retrieved phrasing, hallucination risk). Accordingly, it is positioned as a strong baseline and an interactive aide for early scoping. The neurosymbolic pipeline below adds determinism and explainability via expert-authored rules.

Version 2: Neurosymbolic Pipeline

Rules definition. The rule-based version of CLARA operationalizes the same three-step logic through an explicit rule layer co-designed with legal experts. Instead of asking the LLM to directly output a global risk tier, CLARA encodes legal tests (including definitions, scope exclusions, prohibited practices, and later, the remaining risk categories) as rules. Each rule specifies: (i) a set of conditions, each with a natural-language description, match/anti-hints and optional follow-up question; (ii) a mapping from combinations of condition states to a five-way verdict (`definitely_yes`, `probably_yes`, `undetermined`, `probably_not`, `definitely_not`); and (iii) a class label indicating the rule's implication (e.g., `true` for binary checks, or `prohibited`, `high`, `limited`). Rules are bundled into Policies, which have a scope (e.g.,"AI System Definition") and can define the interactions between the rules to achieve the final decision (e.g., boolean combinations, overriding/blocking conditions, etc.).

LLM-based triggering. When a user supplies a system description, we instantiate an LLM call per policy (rule set). The prompt lists all rules' conditions, their hints (in favor or against), and the processed user's description. The model returns a structured output for each condition, including a trivalent state (true/false/unknown), a brief rationale, evidence spans, and, when necessary, a follow-up question. The outputs are consumed by the Rule Engine, which evaluates the verdict expressions of each rule. For each rule, the engine produces an output summarizing the final verdict and the condition states that led to it.

The policy then aggregates all rule hits into a single result comprising the final decision and a certainty level (definite/probable/none). This yields a neurosymbolic classification that is both reproducible (from the point of triggered conditions) and inspectable: the LLM is used only to score the satisfaction of individual conditions, while the combination of those scores into a risk tier is fully determined by the explicit rules and policy configuration. Clarification via targeted follow-up questions improves the user experience, lowers the entry barrier, and enables the performance of "what-if" scenarios by switching specific conditions (e.g., `true` → `false`) and observing how they propagate.

Implementation and User Experience

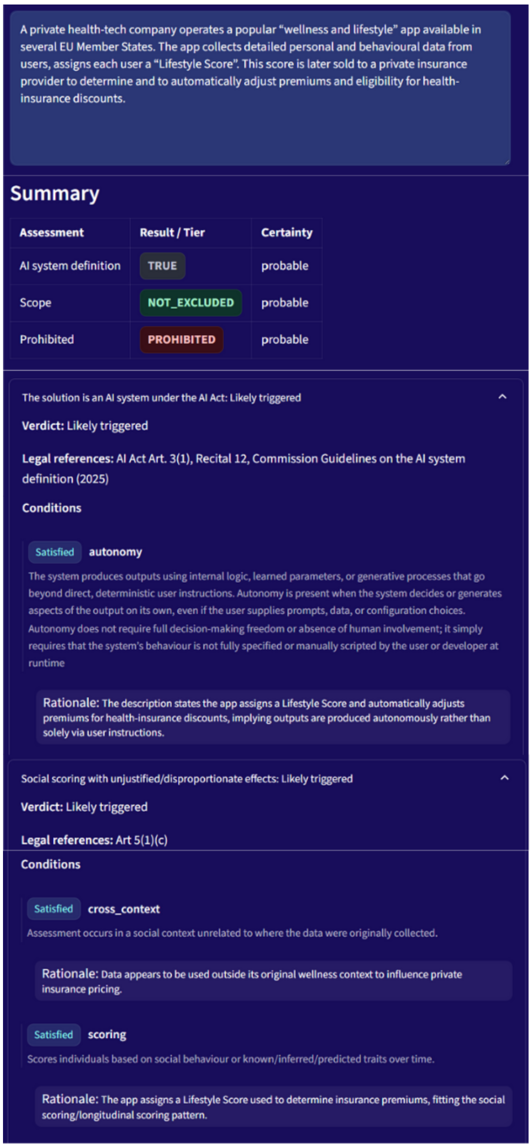
CLARA's prototype is delivered as an interactive web application built with *Streamlit*. The backend is modular, implemented in Python. The interface is organized into a small number of tabs that mirror the three-stage workflow and allow users to inspect the system's reasoning in detail:

- **Input:** users can type a free-form description or upload a PDF/DOCX file. The system performs automatic text extraction and pre-processing.
- **Retrieved Evidence:** for the LLM+RAG pipeline, this tab shows the AI Act and guideline paragraphs selected by the RAG component (sorted and grouped by article/annex), together with relevance scores and provenance metadata.
- **Rule Evaluation:** for the neurosymbolic pipeline, this tab visualizes each rule's summary. Users can expand each rule to inspect the conditions' states, rationale, and follow-up questions for undetermined conditions.
- **Final Assessment:** both pipelines ultimately produce a concise summary: (i) whether the artifact qualifies as an AI system, (ii) whether any scope exclusion applies, and (iii) the resulting risk tier (or "undetermined") - although the current version of the rules only covers prohibited cases, we rely on the LLM route for non-prohibited systems while we work on other tiers' rules. Explanations and cross-references to the supporting legal provisions are shown inline for transparency.

For audit and reproducibility purposes, CLARA can generate a timestamped PDF report. The report includes the input description, the retrieved legal context, the pipeline's reasoning trace (LLM evidence items or rule hits), the final classification, and any outstanding clarification questions. This feature is designed to support internal compliance documentation and facilitate human review.

Evaluation

To get an initial indication of CLARA's behaviour and to validate its usefulness for legal practitioners and entities developing or using AI systems, we conducted a small-scale qualitative evaluation using examples developed by a legal expert and assessed according to the AI Act (see Table 1 in Appendix). For each selected case, we compared four approaches: (i) the independent assessment by the legal expert, (ii) a ChatGPT (5.1) assessment based on a well-crafted prompt, (iii) CLARA's LLM+RAG pipeline (with GPT-5-nano), and (iv) CLARA's neurosymbolic pipeline. Because the descriptions vary widely in detail and completeness, we focused on consistency of reasoning and validity of legal references and follow-up questions, rather than numerical accuracy. Due to space constraints, we restrict the in-depth discussion of a single case, with a few more illustrative results shown in Table 2 in the Appendix. Figure 3 shows the use of CLARA for a user-provided description of a system that, in a nutshell, collects personal and behavioral data from users and scores them with a "Lifestyle" score, which is later sold to a private insurance provider to support automatic premium adjustments and eligibility for discounts.



A private health-tech company operates a popular “wellness and lifestyle” app available in several EU Member States. The app collects detailed personal and behavioural data from users, assigns each user a “Lifestyle Score”. This score is later sold to a private insurance provider to determine and to automatically adjust premiums and eligibility for health-insurance discounts.

Summary

Assessment	Result / Tier	Certainty
AI system definition	TRUE	probable
Scope	NOT_EXCLUDED	probable
Prohibited	PROHIBITED	probable

The solution is an AI system under the AI Act: Likely triggered

Verdict: Likely triggered

Legal references: AI Act Art. 3(1), Recital 12, Commission Guidelines on the AI system definition (2025)

Conditions

Satisfied **autonomy**

The system produces outputs using internal logic, learned parameters, or generative processes that go beyond direct, deterministic user instructions. Autonomy is present when the system decides or generates aspects of the output on its own, even if the user supplies prompts, data, or configuration choices. Autonomy does not require full decision-making freedom or absence of human involvement; it simply requires that the system's behaviour is not fully specified or manually scripted by the user or developer at runtime

Rationale: The description states the app assigns a Lifestyle Score and automatically adjusts premiums for health-insurance discounts, implying outputs are produced autonomously rather than solely via user instructions.

Social scoring with unjustified/disproportionate effects: Likely triggered

Verdict: Likely triggered

Legal references: Art 5(1)(c)

Conditions

Satisfied **cross_context**

Assessment occurs in a social context unrelated to where the data were originally collected.

Rationale: Data appears to be used outside its original wellness context to influence private insurance pricing.

Satisfied **scoring**

Scores individuals based on social behaviour or known/inferred/predicted traits over time.

Rationale: The app assigns a Lifestyle Score used to determine insurance premiums, fitting the social scoring/longitudinal scoring pattern.

Figure 3: Screenshot of version 2. Some intermediate information is cut to highlight some conditions to check if it is an AI System, and the rule triggered for a prohibited case.

CLARA’s neurosymbolic version identifies the system as “Probably an AI System”, given that some of the conditions are satisfied, including “autonomy” (shown in Figure 3, including the rationale) and “inference”, but, for instance, “adaptiveness” is unclear, with the system prompting a follow-up question: “Does the system update or adapt its scoring model after deployment based on new data?”. Then, regarding the material scope of the AI Act, no exclusion criteria were found, leading CLARA to confirm that the system is within AI Act’s material scope. Finally, CLARA deemed the system as probably Prohibited, due to a rule corresponding to Art. 5(1)(c) on social scoring, triggering the conditions of “cross-context” and “scoring”, with the rationale shown in Figure 3, aligning closely with the expert assessment. In comparison, ChatGPT (5.1) also assessed the solution as an AI system and found no exclusions from the material scope. However, it classified the system as High-risk, based on Annex III 5(c), focusing on risk assessment and pricing relating to natural persons in life and health insurance, but ignoring the prohibited scoring and cross-context use. In turn, the LLM+RAG version of CLARA returned “Undetermined” for the AI system check, posing follow-up questions like “Are the scoring rules fixed or do they adapt over time based on observed data?”. It found a possible exception based on “legitimate insurance underwriting (...) when lawful and proportionate under Union and national law”, but stating it hinges on “sector-specific data use and safeguards being in place”. The suggested risk tier is “High”, again due to Annex III 5(c), not considering the scoring and cross-context prohibition.

From our limited analysis, we could still draw some intuitive insights. Although the prompt given to ChatGPT included a request to be conservative, it returned fewer “Undetermined” results for the AI system check, even though some AI system criteria, such as autonomy, were not explicitly mentioned in the prompt. On the other hand, ChatGPT returned more “Undetermined” when checking for exclusions from the material scope and for the risk classification, failing to identify some exceptions that seemed obvious under the EC Guidelines (which CLARA considers, in both the LLM+RAG and neurosymbolic pipeline, as context or rule conditions). Comparing both CLARA versions, there are three main apparent differences (besides the rule-based not being able to classify other risk tiers beyond Prohibited for now): the LLM+RAG seems to return more “Undetermined” results, probably because rules are grounded in more fine-grained conditions; and the former is also more prone to find exclusions related to the Minimal/no-risk tier (not our material scope question). Additionally, in a small experiment, where we ran the descriptions several times, the rules-based pipeline produced more stable results, as its dependence on the LLM is limited to condition checks. Finally, Rules provide a more interpretable and controllable assessment, allowing for “what-if” scenarios or quick corrections.

Conclusion

We demonstrated CLARA, a tool that combines LLM-based semantic reasoning with deterministic rule logic to support early-stage AI Act self-assessment. The system accepts free-form descriptions, retrieves relevant provisions of the Act and Commission guidelines, and applies either a fully LLM-driven or a neurosymbolic pipeline to determine applicability, scope exceptions, and risk tier. Our initial experiments indicate that the hybrid approach improves interpretability and stabilizes classification outcomes compared to LLM-only reasoning. Nevertheless, several limitations remain: the interpretation of the AI Act is evolving; system descriptions vary in clarity and granularity; and semantic matching depends on the capabilities of current LLMs. Future work includes expanding the rule base (e.g., high-risk Annex III categories) and supporting human-in-the-loop verification (e.g. prioritizing follow-up questions with the most impact on the result), moving towards a conversational framework where each clarification results in a quick update. More broadly, CLARA illustrates how multidisciplinary collaboration towards hybrid AI–law systems can provide transparent, trustworthy, and practically usable compliance support as organizations adapt to emerging regulatory frameworks.

Appendix

Table 1 shows the full descriptions of the chosen use cases for the assessment. In turn, the results from the comparative assessment are summarized in Table 2. Finally, Table 3 presents the rationale behind three illustrative use cases, accompanied by relevant legal references and follow-up questions as needed. We note, however, that this is a simplified version, as every model is more explanatory in its results.

Ethical Statement

This research does not involve human participants, personal data, or sensitive or biometric information. All use-case descriptions are synthetic or publicly available examples drawn from non-personal datasets. The CLARA system is intended solely as a technical demonstrator to support understanding of the EU AI Act and should not be used as a source of legal advice or regulatory determination. We highlight that automated legal reasoning can yield uncertainty or inaccuracies. Therefore, any compliance assessment must be reviewed by qualified legal professionals. We also emphasize transparency: all model prompts, rule logic, and evaluation data are documented to support reproducibility and responsible use.

Acknowledgments

This work was supported by European funds through the Recovery and Resilience Plan, via the project ”Center for Responsible AI”, with identification number C645008882-00000055.

References

Baig, A.; Qayyum, R. F.; and Khan, S. 2025. EU AI Act Applicability Assessment & Compliance Checker. Securiti Blog, https://securiti.ai/eu-ai-act/assessments/.

Davvetas, A.; Ziouvelou, X.; Dami, Y.; Kaponis, A.; Giouvanopoulou, K.; and Papademas, M. 2025. TAI Scan Tool: A RAG-Based Tool With Minimalistic Input for Trustworthy AI Self-Assessment. arXiv:2507.17514.

European Comission. 2025a. Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act). https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act.

European Comission. 2025b. Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act). https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application.

European Commission. 2025. Commission launches AI Act Service Desk and Single Information Platform to support AI Act implementation. https://digital-strategy.ec.europa.eu/en/news/commission-launches-ai-act-service-desk-and-single-information-platform-support-ai-act.

European Union. 2024. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689.

Future of Life Institute. 2025. EU AI Act Compliance Checker. https://artificialintelligenceact.eu/assessment/eu-ai-act-compliance-checker/. Accessed Nov. 2025.

Kovari, A.; Ghafourian, Y.; Hegedus, C.; Naim, B. A.; Mezei, K.; Varga, P.; and Tauber, M. 2025. Let’s have a chat with the EU AI Act.

Leong, B.; and Judge-Raza, Z. 2025. The EU’s AI Act Compliance Checker & Explorer—What’s Useful Today (And What’s Still to Come). ZwillGen Blog, https://www.zwillgen.com/artificial-intelligence/the-eus-ai-act-compliance-checker-explorer-whats-useful-today-whats-still-come/.

Table 1: Full use-case descriptions for CLARA's qualitative assessment.

ID	Full Description
1	A U.S. company offers a large language system capable of generating text and code through a public API. Subscriptions are priced in euros and advertised in EU languages. All computation occurs on servers in North America but is used in Europe.
2	A European car manufacturer integrates a computer-vision lane-keeping module into a passenger-vehicle line sold across the EU under its own brand. The module interprets camera input and issues steering corrections autonomously.
3	A mobile app publisher offers a generative photo-enhancement application. The app's AI model applies generative filters in response to user prompts and images, producing creative outputs that are not used for professional or commercial activities.
4	A university in the Netherlands develops a reinforcement-learning controller to optimise laboratory robotics for experimental chemistry. The system runs only on lab equipment within the university and is put into service exclusively for scientific research under a public grant. There is no provision of the system to third parties
5	A software company builds a demand-forecasting AI for supermarkets. During development, the team performs offline modelling and synthetic simulations that have no effect on real operations. Later, the company runs a live pilot in several supermarkets, where the model's forecasts directly change replenishment orders and shelf availability for actual customers.
6	A defence contractor develops an AI system for target recognition on surveillance drones, used for national defence missions under government contracts.
7	A consortium publishes an image generation system under a free and open-source licence, releasing weights and code on a public repository at no charge.
8	An aerospace company provides a software component that uses pre-programmed mathematical optimisation methods to allocate satellite bandwidth efficiently. The component employs regression and heuristic approximations to speed up calculations but does not learn or modify its model beyond these fixed rules.
9	An EU bank operates a loan-risk scoring engine developed in-house. The system calculates a risk score by applying a fixed, manually designed formula that weights factors such as income, credit history, and outstanding debt. The coefficients were set by credit analysts and are updated annually through manual review.
10	A video-game studio releases a fantasy role-playing game in which non-player characters (NPCs) react and adapt to player actions using a set of guidelines.
11	A mobile learning app for children tracks facial micro-expressions through the device camera and deploys imperceptible colour shifts and sound frequencies that trigger emotional excitement to increase engagement.
12	A digital-assistant service for seniors uses behavioural profiling to detect loneliness and suggests relevant services.
13	A law firm is launching an emotion recognition AI tool to detect early signs of burnout in its lawyers and employees.
14	A private health-tech company operates a popular "wellness and lifestyle" app available in several EU Member States. The app collects detailed personal and behavioural data from users, assigns each user a "Lifestyle Score". This score is later sold to a private insurance provider to determine and to automatically adjust premiums and eligibility for health-insurance discounts.
15	Our city's police department will operate a network of street cameras and use an AI model to run live face recognition to spot suspects as people pass through subway stations.
16	Police perform real-time face recognition using on street CCTV using AI, but only under a specific court order to locate a missing child during a 24-hour authorised window

Table 2: Case comparison across expert assessment, GPT baseline, CLARA LLM+RAG, and CLARA rules-based system. Columns report whether the system is an AI system (**AI**), whether it is excluded or out of scope under Articles 2(3)–2(12) (**Excl**), and the assigned risk tier (**Tier**). Symbol legend: **AI** - T=AI system, F=not AI, U=undetermined; **Excl** - T=excluded/out of scope, F=not excluded, U=undetermined; **Tier** - P=Prohibited, H=High risk, L=Limited, M=Minimal/no risk, -=not applicable, U=undetermined. Expert assessments are shown without color (baseline). Model outputs (ChatGPT baseline, CLARA LLM+RAG, CLARA Rules) are color-coded: green = match with expert; orange = undetermined; red = mismatch.

		Expert			**ChatGPT**			**CLARA LLM+RAG**			**CLARA Rules**		
ID	**Use Case**	**AI**	**Excl**	**Tier**	**AI**	**Excl**	**Tier**	**AI**	**Excl**	**Tier**	**AI**	**Excl**	**Tier**
1	US LLM API in EU	T	F	L	T	F	L	U	F	U	T	U	U
2	Lane-keeping module	T	F	H	T	F	H	T	U	H	T	U	H
3	Photo enhancement app	T	F	L	T	F	L	T	U	M	T	F	M
4	University RL controller (research)	T	T	–	T	T	-	T	T	-	T	T	-
5	Demand forecasting pilot	T	F	M	T	F	M	T	T	-	T	F	H
6	Defence target recognition	T	T	–	T	T	-	U	T	-	U	T	-
7	Open-source generative model	T	T	L (if in-scope)	T	U	U	U	T	-	U	T	-
8	Satellite optimiser (no learning)	F	–	-	F	-	-	F	-	-	F	-	-
9	Loan-risk scoring (fixed formula)	F	-	-	F	-	-	F	-	-	F	-	-
10	RPG NPC behaviours	T	F	M	T	F	M	T	U	M	T	F	M
11	Children micro-expression manipulation	T	F	P	T	F	H	U	F	H	T	F	P
12	Elderly behavioural profiling	T	F	P	T	F	L	U	U	H	T	F	H
13	Burnout emotion-recognition tool	T	F	P	T	F	P	T	F	P	U	F	P
14	Lifestyle-scoring for insurers	T	F	P	T	F	H	U	U	H	T	F	P
15	Live face recognition (public spaces)	T	F	P	T	F	P	T	U	H	T	F	P
16	Court-ordered face recognition (missing child)	T	F	H (exception applies)	T	F	H	T	U	H	T	F	H

Table 3: Detailed rationales, legal references, and follow-up questions for three representative use cases (6, 11, 14). AI? = AI system classification; Excl? = excluded/out-of-scope determination; Tier = assigned risk category. Symbol legend: **AI** - T=AI system, F=not AI, U=undetermined; **Excl** - T=excluded/out of scope, F=not excluded, U=undetermined; **Tier** - P=Prohibited, H=High risk, L=Limited, M=Minimal/no risk, -=not applicable, U=undetermined.

Use Case	Source	AI?	Excl?	Tier	Rationale (R), Legal References, Follow-up Questions (Q)
6	Expert	T	T	-	Art 2(3). Could be high-risk if considered in-scope.
	ChatGPT	T	T	-	R: Article 2(3) excludes AI systems used exclusively for military, defence or national security purposes. Q: Is the system used exclusively in defence contexts, or are there any civilian applications?
	CLARA LLM+RAG	U	T	-	AI - R:While target recognition implies inference from input to outputs, explicit confirmation of adaptive learning or autonomy (Guideline II §3) is not provided in the description. Q: What is the intended level of autonomy during operation (fully autonomous, semi-autonomous, or manual override)? Excl - R: National security exclusion appears to apply based on the described use for defence/military purposes. The description indicates exclusive use for national defence, which would fall outside the AI Act's scope per the national security carve-out. However, any non-defence or dual-use deployment would require compliance. Q: Will the system ever be marketed or deployed in non-defence contexts that would bring it under the AI Act's scope?
	CLARA Rules	U	T	-	AI - R: Autonomy and Adaptiveness unclear - The description does not specify autonomous, self-generated outputs beyond direct user instructions. No explicit statement about self-learning or adaptation after deployment. Q: Does the system generate outputs autonomously without real-time human control? Does the system adapt or self-learn after deployment? Excl - R: Description references defence and national security related use but does not explicitly state exclusivity to military/national security purposes; although some ambiguity remains, it is probably excluded. Q: Is the system used exclusively for military, defence, or national security purposes?
11	Expert	T	F	P	Art 5(1)(a)
	ChatGPT	T	F	H	R: Annex III lists AI systems intended for emotion recognition among high-risk uses when permitted by law. This app processes children's biometric data to infer emotions and influence their engagement, raising significant fundamental-rights concerns. It may approach the boundary of prohibited practices exploiting vulnerabilities of children, but on the information given, it is more appropriate to classify it as high-risk rather than clearly prohibited. Q: Does the design risk materially distort children's behaviour in a way likely to cause significant harm?
	CLARA LLM+RAG	U	F	H	AI - R: The guidelines require AI techniques (e.g., ML or logic-based inference) and potential adaptiveness after deployment to qualify as an AI system under Article 3(1). The description does not specify the use of AI methods or post-deployment adaptiveness, so a firm conclusion cannot be drawn. Q: Does the app use AI techniques (e.g., machine learning or logic-based inference) to analyze facial expressions? Risk - R: The app uses camera-based emotion recognition on children and then subtly modulates colours and sounds to influence engagement. That combination aligns with a high-risk use under the EU AI Act, triggering requirements for risk management, transparency, and protections for minors.
	CLARA Rules	T	F	P	R: Triggers Art 5(1)(a), Recitals 28-29 regarding subliminal or manipulative techniques beyond consciousness. The stimuli are described as imperceptible, suggesting subconscious influence. The system uses imperceptible stimuli to influence engagement. Furthermore, likely triggers Art 5(1)(b) on the exploitation of vulnerabilities since the technique aims to increase engagement among users, which could distort the behavior of vulnerable users, children.
14	Expert	T	F	P	Art. 5(1)(c)(ii), secondary use
	ChatGPT	T	F	H	R: Annex III point 5(c) designates as high-risk AI systems used for risk assessment and pricing relating to natural persons in life and health insurance. Q: Are policyholders clearly informed that AI-based profiling affects their insurance pricing?
	CLARA LLM+RAG	U	U	H	AI - R: The description involves generating a Lifestyle Score from user data to influence insurance premiums, which could be an AI-enabled inference with some autonomy. However, the text does not specify whether AI techniques, learning/adaptation, or autonomous decision-making are actually used, leaving room for simple data processing. Q: Are the scoring rules fixed or do they adapt over time based on observed data? Risk - R: The system maps to risk assessment and pricing in life/health insurance, a high-risk area under Annex III. The setup also suggests profiling of users, which raises data governance and privacy concerns that would require safeguards and compliance.
	CLARA Rules	T	F	P	Risk - R: Likely triggers Art 5(1)(c) on social scoring. Data collected via the wellness app is used by a private insurer to make pricing decisions, indicating context shift from original collection context (cross-context). The app uses a Lifestyle Score to determine and adjust health-insurance premiums/discounts, fitting a scoring system based on behaviour/traits over time (scoring). Q: Is the outcome disproportionate to the person's behaviour/traits?