

Consensus Report

Question: For a mid-size enterprise deploying AI systems in the EU, analyze the compliance requirements under the EU AI Act (effective 2025-2026). Structure your analysis as:

- 1) Executive summary of key obligations
- 2) Risk classification framework and how it applies
- 3) Timeline of compliance milestones
- 4) Technical and organizational measures required
- 5) Penalties and enforcement mechanisms
- 6) Recommended compliance roadmap with priority actions.

Models: Claude Opus 4.6 (thinking), GPT-5.4 (thinking), Gemini 3.1 Pro (thinking), Grok 4.20 (thinking)

Duration: 7m 46s

Date: 2026-03-20

CONTESTED

Significant disagreement — 1 agreed, 3 dissenting

Consensus

REVISED CONSENSUS: EU AI Act Compliance for Mid-Size Enterprises

1. Consensus (What Everyone Agrees On)

Mid-size EU enterprises must build a risk-based compliance program centered on **role determination** (Provider vs. Deployer) and **use-case classification** (Prohibited, High-Risk, Limited, Minimal). Immediate action is required: **February 2025 deadlines** for banning prohibited AI practices (e.g., workplace emotion recognition) and mandating staff AI literacy under Article 4 are already active. Fines are severe and tiered (see §5). The "accidental provider" trap under Article 25 remains the single highest-stakes compliance risk for mid-size firms.

2. Risk Classification & Role Determination

Provider vs. Deployer Obligations (Corrected per GPT-5.4):

Obligation	Provider	Deployer
Conformity assessment & CE marking	Required	Not required
Technical documentation	Must create	Must obtain/retain
Serious incident reporting	Fixed deadlines to authorities	Notify provider + authorities without undue delay
Human oversight implementation	Design for it	Operationalize it
FRIA	Where required (targeted scope)	
Logging & monitoring	Build capability	Maintain logs under their control

"Accidental Provider" Risk — Refined Technical Boundary (per Gemini 3.1 Pro):

The consensus preserves the core warning but now distinguishes:

- **Lower reclassification risk:** Standard RAG and prompt engineering (context-injection at inference time; model weights unmodified; vendor's intended purpose preserved). These generally maintain Deployer status.
- **Higher reclassification risk:** Fine-tuning, transfer learning, or any modification altering model weights or circumventing the provider's safety architecture — especially when applied to Annex III high-risk use cases. These risk triggering full Provider obligations.

> **REMAINING DIVERGENCE:** Whether *complex* RAG architectures (e.g., agentic RAG with tool-calling that fundamentally reshapes system behavior for high-risk purposes like employee evaluation) could constitute "substantial modification" remains genuinely unsettled. Gemini argues regulators will draw the line cleanly at weight modification. The original position and Claude Opus maintain that purpose-altering architectures, even without weight changes, could attract scrutiny. **This is the critical open question for regulators and CEN/CENELEC guidance.**

3. Corrected Timeline of Compliance Milestones

| Date | Milestone | Status |

|---|---|---|

| **Feb 2, 2025** | Prohibited practices banned; AI literacy mandatory (Art. 4) | **ACTIVE NOW** |

| **Aug 2, 2025** | GPAl provider obligations (Art. 53); governance structures | Approaching |

| **Aug 2, 2026** | **Annex III** high-risk system obligations apply (HR, recruitment, credit scoring, etc.) | Primary deployer deadline |

| **Aug 2, 2027** | **Annex I** high-risk systems (safety components of regulated products: medical devices, machinery, etc.) | Later deadline |

Correction per GPT-5.4: The original consensus failed to distinguish Annex III (Aug 2026) from Annex I product-regulated systems (Aug 2027). *Grok 4.20's claim that all Annex III deployer obligations shift to August 2027 is not adopted* — the standard reading of Article 113 places Annex III obligations at 24 months.

4. Technical & Organizational Measures

- **Human oversight:** Staff with authority, competence, and *meaningful override capability* for high-risk systems
- **Logging:** Maintain automated logs under deployer control for duration specified by provider or minimum required
- **Incident escalation (corrected per GPT-5.4):** Build a **documented rapid escalation and notification workflow** — not a rigid 15-day deployer deadline, which applies primarily to providers
- **FRIA (corrected per GPT-5.4):** Required for specific deployer contexts (public bodies; certain private deployers of targeted Annex III systems) — integrate into existing GDPR DPIA workflows *where applicable*, not universally
- **Avoid premature technical hardcoding** until CEN/CENELEC harmonized standards are published

5. Penalties — Tiered (Corrected per GPT-5.4)

| Violation Type | Maximum Fine |

|---|---|

| Prohibited practices | **€35M or 7%** global turnover |

| High-risk system obligations | **€15M or 3%** global turnover |

| Incorrect information to authorities | **€7.5M or 1%** global turnover |

No SME caps apply to mid-size enterprises.

6. Revised Compliance Roadmap

Immediate (Next 30 Days):

- Complete AI inventory mapping tools *and* their data pipelines
- Terminate any prohibited uses
- Deploy mandatory AI literacy training
- Monitor GPAI providers' Article 53 transparency compliance (per Claude Opus amplification)

Short-Term (Q2-Q3 2025):

- Establish a **mandatory AI Architecture Review Gate**: any project proposing fine-tuning or weight modification for high-risk use cases must undergo Provider-level compliance assessment before proceeding
- Standard RAG integration may proceed but with documented risk assessment for Annex III use cases
- Begin integrating FRIAs into existing DPIA workflows where required
- Monitor national competent authority guidance across operating jurisdictions

Medium-Term (Q4 2025-H1 2026):

- Update vendor contracts to mandate CE conformity evidence, technical documentation access, and audit rights
- Build logging retention and human oversight interfaces for Annex III high-risk tools
- Establish rapid incident escalation and notification workflows
- Conduct readiness assessment against emerging harmonized standards

Longer-Term (H2 2026-H1 2027):

- Extend compliance architecture to Annex I product-regulated AI systems
- Refine technical measures as CEN/CENELEC standards are finalized

Remaining Open Question

Where exactly will regulators draw the line between context-injection (RAG) and substantial modification — particularly for complex agentic architectures applied to Annex III use cases? Mid-size enterprises should monitor early enforcement actions and CEN/CENELEC guidance closely, as this boundary will define the practical scope of provider liability for thousands of firms.

Key Disagreements**CLAUDE****AGREE**

This is a well-structured and substantively accurate consensus that I would endorse with only minor amplifications.
****Why I agree:****

The analysis correctly identifies the most critical and frequently underestimated risks facing mid-size enterprises. Specifically:

- 1.

GPT**OBJECT**

The overall direction is right, but several legal details should be tightened before treating this as “consensus.”

1. ****Deployer incident reporting is overstated.****

The statement that deployers must have a ****“15-day incident reporting workflow”** by Aug 2026****** is too categorical. The AI Act’s...

GEMINI**OBJECT**

I object to the technical characterization in the "Key nuances" and "Actionable recommendations" sections stating that Retrieval-Augmented Generation (RAG) risks triggering "accidental provider" status. ****Reasoning:****
 Under Article 25 of the EU AI Act, a deployer becomes a provider if they make a...

GROK**OBJECT**

The consensus is mostly accurate on risk classification, the "accidental provider" trap (fine-tuning + RAG in Annex III use cases can trigger provider obligations under Article 25), and the value of integrating into existing DPIA processes. However, the timeline contains a material error that wou...

CONCLAVIK · SAMPLE

APPENDIX — Full Deliberation Record

The following pages contain the complete multi-model analysis for audit and reference purposes.

Detailed Analysis

Round 1: Initial Analysis

CLAUDE

EU AI Act Compliance Analysis for Mid-Size Enterprises

1. Executive Summary

The EU AI Act (Regulation 2024/1689) imposes risk-based obligations on AI providers and deployers. Mid-size enterprises face proportionate but significant requirements: classify all AI systems by risk tier, implement governance frameworks, ensure transparency, and maintain documentation. Non-compliance carries fines up to €35M or 7% of global turnover.

2. Risk Classification Framework

- **Unacceptable Risk (Banned):** Social scoring, manipulative subliminal techniques, real-time remote biometric ID in public spaces (with exceptions). **Action:** Audit portfolio; discontinue prohibited uses.
- **High-Risk (Art. 6):** AI in HR/recruitment, creditworthiness, critical infrastructure, education, law enforcement. Requires conformity assessments, risk management systems, data governance, human oversight, and CE marking.
- **Limited Risk:** Chatbots, emotion recognition, deepfakes. Requires **transparency obligations** — users must know they're interacting with AI.
- **Minimal Risk:** Spam filters, AI-enabled video games. Voluntary codes of conduct only.

Most mid-size enterprises will have systems spanning limited and high-risk categories, particularly in HR and customer-facing applications.

3. Compliance Timeline

Milestone	Date
Prohibited practices ban	Feb 2, 2025
GPAI model obligations	Aug 2, 2025
High-risk systems (Annex III)	Aug 2, 2026
High-risk in regulated products (Annex I)	Aug 2, 2027

4. Technical & Organizational Measures

- **Risk management system** (Art. 9): continuous, iterative lifecycle process
- **Data governance** (Art. 10): training data quality, bias testing, representativeness
- **Technical documentation & logging** (Art. 11-12): automated event recording
- **Human oversight mechanisms** (Art. 14): meaningful human-in-the-loop controls

- **AI literacy training** (Art. 4): mandatory for all staff operating AI — **due Feb 2025**
- **Fundamental rights impact assessments** for deployers of high-risk systems (Art. 27)

5. Penalties & Enforcement

- Prohibited AI violations: **€35M / 7% turnover**
- High-risk non-compliance: **€15M / 3% turnover**
- Misinformation to authorities: **€7.5M / 1%**
- SME-proportionate caps apply but "mid-size" enterprises above SME thresholds face full penalties. National market surveillance authorities enforce.

6. Recommended Compliance Roadmap

Immediate (Q1 2025):

1. Complete AI system inventory across all departments
2. Implement AI literacy training (Art. 4 already effective)
3. Cease any prohibited AI practices

Short-term (Q2-Q3 2025):

4. Classify every system by risk tier
5. Appoint AI governance officer/team
6. Begin GPAI provider due diligence

Medium-term (Q4 2025 - Q2 2026):

7. Implement conformity assessment processes for high-risk systems
8. Deploy technical documentation and logging infrastructure
9. Conduct fundamental rights impact assessments
10. Establish vendor compliance contractual requirements

Priority action today: The AI inventory and literacy obligation are already overdue. Start there.

GPT

1) Executive summary of key obligations

A mid-size enterprise in the EU is usually an **AI deployer**, but may also become a **provider** if it builds, materially modifies, or rebrands AI systems. Core obligations are to:

- **Ban prohibited AI uses** and ensure **AI literacy** for relevant staff.
- **Classify each AI use case by risk** and role (deployer/provider/importer/distributor).
- For **high-risk AI**, use only compliant systems, follow instructions, ensure human oversight, monitor operation, keep logs, and report serious incidents.
- Meet **transparency duties** for chatbots, emotion recognition, biometric categorization, and AI-generated/deepfake content.
- Maintain documentation, vendor due diligence, and governance evidence.

2) Risk classification framework and how it applies

- **Prohibited AI:** e.g., certain manipulative, exploitative, social scoring, and some biometric uses. These must not be deployed.
- **High-risk AI:** especially relevant in **HR/recruitment**, worker management, education, essential services, creditworthiness, biometrics, law enforcement, migration, justice.
- If you use AI for CV screening, employee evaluation, access control, or critical service decisions, assume **high-risk review**.
- **Limited-risk / transparency AI:** chatbots, synthetic media, emotion recognition. Users must be informed.
- **Minimal-risk:** most productivity copilots and internal analytics, unless used in a high-risk context.
- **GPAI/model layer:** obligations mainly fall on providers, but deployers need **vendor assurance**,

contractual rights, and usage controls.

3) Timeline of compliance milestones

- **2 Feb 2025:** rules on **prohibited AI + AI literacy** apply.
- **2 Aug 2025:** **GPAI** obligations start applying; review model/vendor compliance.
- **2 Aug 2026:** **most AI Act obligations** apply, including most deployer/high-risk requirements.
- **2 Aug 2027:** additional deadlines for certain **Annex I regulated product** AI systems.

4) Technical and organizational measures required

- Enterprise AI inventory and risk register.
- Role mapping: deployer vs provider.
- Vendor due diligence: CE marking/declaration, instructions, technical docs, logging, incident clauses.
- Human oversight, access controls, testing, monitoring, logging, change management.
- Transparency notices and content labeling.
- Incident response and escalation to authorities/providers.
- For certain high-risk uses: **fundamental-rights impact assessment** before deployment.

5) Penalties and enforcement mechanisms

Fines can reach **€35m or 7%** of global turnover (prohibited AI), **€15m or 3%** for other major breaches, and **€7.5m or 1.5%** for misleading information. Enforcement comes via **national market surveillance authorities**, notified bodies, and the **EU AI Office** (notably for GPAI), with powers to demand corrective action, withdrawal, or recalls.

6) Recommended compliance roadmap with priority actions

1. **Now:** appoint AI owner/committee; launch AI inventory.
2. **Q2-Q3 2025:** prohibit banned uses; roll out AI literacy training.
3. **Q3 2025:** classify all use cases; flag HR/worker-management/biometric tools.
4. **Q3-Q4 2025:** update procurement/contracts for AI vendor evidence and audit rights.
5. **H1 2026:** implement high-risk controls, oversight, logging, notices, incident process.
6. **Before Aug 2026:** complete documentation, assessments, and audit readiness.

GEMINI

1. Executive Summary

The EU AI Act mandates a strict, risk-based approach to AI deployment. For a mid-size enterprise, core obligations center on transparency, human oversight, and auditable data governance. Enterprises must transition from ad-hoc AI usage to formalized AI lifecycle management, ensuring third-party tools and internal deployments do not expose the business to severe regulatory liabilities.

2. Risk Classification Framework

Enterprises must classify their AI systems into four tiers:

- **Unacceptable Risk:** Banned entirely (e.g., workplace emotion recognition, social scoring).
- **High-Risk:** High-scrutiny systems (e.g., AI used for CV screening/HR, credit scoring). Requires conformity assessments, risk mitigation, and continuous monitoring.
- **Limited Risk:** Requires transparency. Users must know they are interacting with AI (e.g., customer service chatbots, AI-generated content).
- **Minimal Risk:** No mandatory obligations (e.g., spam filters), though voluntary codes of conduct apply.

3. Timeline of Compliance Milestones

- **Late 2024 (Entry into force):** Act officially becomes law.
- **Mid-2025 (6 months):** Prohibitions on unacceptable risk AI take effect.
- **Mid-2025 (12 months):** Transparency obligations for General Purpose AI (GPAI) and chatbots apply.

- **Mid-2026 (24 months):** Strict compliance required for most High-Risk AI systems (Annex III).

4. Technical and Organizational Measures

For high-risk and limited-risk systems, enterprises must implement:

- **Quality Management Systems (QMS):** Documented policies covering regulatory compliance and accountability.
- **Data Governance:** Validation mechanisms ensuring data is representative, error-free, and bias-mitigated.
- **Traceability & Cybersecurity:** Automated logging of system events and robust safeguards against adversarial attacks.
- **Human Oversight:** Built-in mechanisms allowing human intervention or complete system overrides ("human-in-the-loop").

5. Penalties and Enforcement

Fines are severe, though regulatory caps may provide some proportionality for mid-size SMEs:

- Up to **€35M or 7%** of global annual turnover for prohibited AI practices.
- Up to **€15M or 3%** for violating high-risk system obligations.
- Up to **€7.5M or 1.5%** for supplying incorrect information to regulators.

6. Recommended Compliance Roadmap

- **Action 1: Map & Classify (Immediate).** Audit all proprietary and third-party AI tools. Map them strictly to the Act's four risk tiers.
- **Action 2: Cross-Functional Governance (Next 3 months).** Integrate AI risk assessments into existing GDPR Data Protection Impact Assessment (DPIA) workflows.
- **Action 3: Vendor Contract Updates (By Mid-2025).** Require AI vendors to contractually guarantee EU AI Act compliance, provide necessary system logs, and assume liability for algorithmic flaws.
- **Action 4: AI Literacy (Ongoing).** Implement mandatory training to ensure staff recognize limited vs. high-risk AI deployments.

GROK

1) Executive summary of key obligations

Mid-size enterprises acting primarily as deployers must classify every AI system, apply role-specific duties, maintain human oversight for high-risk uses, ensure transparency where required, and report serious incidents. Providers have heavier burdens, but deployers cannot ignore upstream compliance.

2) Risk classification framework and how it applies

The Act uses four tiers:

- **Unacceptable risk** (Annex II): banned (social scoring, manipulative subliminal techniques, certain biometric categorization).
- **High risk** (Annex III): employment, credit scoring, law enforcement, critical infrastructure, biometric identification. Requires risk management, data quality, logging, human oversight, and conformity assessment.
- **Limited risk:** chatbots, deepfakes, emotion recognition — transparency obligations only.
- **Minimal risk:** most internal tools — voluntary.

Actionable step: map every deployed or procured AI system against Annex III within 60 days.

3) Timeline of compliance milestones

- 2 Feb 2025: prohibited AI systems banned.
- 2 May 2025: general-purpose AI transparency rules.
- 2 Aug 2026: high-risk obligations apply to most deployers.
- 2 Aug 2027: remaining high-risk sectors.

Preparation must begin immediately; most substantive work falls in 2025.

4) Technical and organizational measures required

- Conduct and document fundamental rights impact assessments for high-risk uses.
- Implement logging and traceability of AI outputs.
- Establish human oversight interfaces with clear escalation paths.
- Maintain technical documentation and data governance (quality, bias checks).
- Create incident reporting procedures to national authorities within 15 days for serious incidents.
- Provide staff training and appoint an internal AI compliance lead.

5) Penalties and enforcement mechanisms

Fines are tiered: up to €35M or 7% of global annual turnover for prohibited systems; €15M or 3% for high-risk violations; €7.5M or 1% for incorrect information. National market surveillance authorities enforce, coordinated by the EU AI Office. Private enforcement via affected parties is also possible.

6) Recommended compliance roadmap with priority actions

Priority 1 (by Dec 2024): create AI inventory and risk-classification matrix; assign compliance owner.

Priority 2 (Q1 2025): perform gap analysis against high-risk obligations and prohibited practices; remediate or decommission banned uses.

Priority 3 (Q2 2025): implement logging, oversight controls, and staff training program.

Priority 4: integrate AI clauses in vendor contracts and establish ongoing monitoring dashboard.

Priority 5: engage external legal counsel for annual review.

Focus resources on high-risk systems first. Early classification and documentation reduce fine exposure by demonstrating good faith. (Word count: 378)

Round 2: Cross-Examination

CLAUDE

Refined Position After Reviewing All Responses

Areas of Strong Consensus

All four analyses converge on the core framework: risk-based classification, the penalty tiers, and the general compliance timeline. This convergence should give any mid-size enterprise confidence that these are settled obligations, not interpretive speculation.

Where I Refine My Position

1. GPT-5.4's deployer vs. provider distinction is critical — I underweighted it. My original analysis insufficiently emphasized that a mid-size enterprise modifying a third-party model (fine-tuning, RAG integration) could be reclassified as a **provider** under Art. 25. This dramatically escalates obligations. **Update d recommendation:** Every enterprise must perform role mapping (provider/deployer/distributor) per system, not just risk classification.

2. Gemini's GDPR-DPIA integration point is the most practical advice given. Mid-size enterprises already have GDPR infrastructure. Bolting AI risk assessments onto existing DPIA workflows is the highest-ROI governance move. I adopt this recommendation fully.

3. Grok's 15-day serious incident reporting timeline is a valuable specific. I omitted this. Art. 62 requires deployers of high-risk systems to report serious incidents to market surveillance authorities. This needs operationalization — incident response playbooks must be built before Aug 2026.

Where I Disagree

Grok's Dec 2024 Priority 1 is retrospectively unhelpful — that deadline passed. Recommendations should anchor to what's actionable *now*. More critically, Grok lists "2 May 2025" for GPAI transparency — the

actual date is **2 Aug 2025**.

Gemini incorrectly categorizes workplace emotion recognition as universally banned. Art. 5(1)(f) bans emotion recognition in **workplaces and education**, but with exceptions for medical/safety purposes. This nuance matters for enterprises in manufacturing or logistics using fatigue-detection AI.

All models, including mine, underemphasize one thing: the role of **harmonized standards** (expected via CEN/CENELEC). Without finalized standards, conformity assessment for high-risk systems remains practically ambiguous. Enterprises should monitor standardization progress and avoid locking into compliance architectures prematurely.

Updated Priority Actions (Revised)

1. **Immediate:** Complete AI inventory with **role classification** (provider/deployer) per system — not just risk tier
2. **Immediate:** AI literacy training (Art. 4 — already enforceable)
3. **Q2 2025:** Integrate AI risk assessment into existing GDPR/DPIA workflows
4. **Q3 2025:** Update vendor contracts with conformity evidence, audit rights, incident notification clauses, and liability allocation
5. **Q4 2025-Q1 2026:** Build high-risk compliance infrastructure — logging, human oversight, FRIA documentation, incident response playbooks
6. **Ongoing:** Track CEN/CENELEC harmonized standards development before finalizing technical compliance architecture

The single highest-risk blind spot for mid-size enterprises: unknowingly becoming a **provider** through model customization, triggering obligations they haven't planned for.

GPT

I **mostly agree** with Claude and Gemini on the practical roadmap, and with Grok on prioritizing deployer duties. But I'd refine several legal points:

1) Executive summary

For a mid-size enterprise, the biggest issue is **role clarity**: are you only a **deployer**, or also a **provider** because you build, substantially modify, fine-tune into a new intended purpose, or white-label AI?

- If **deployer only**: your duties are mainly classification, vendor due diligence, human oversight, logging/monitoring, transparency, incident handling, and staff literacy.
- If also **provider**: conformity assessment, CE marking, and technical documentation become much heavier.

2) Risk classification

Broadly correct from others, but two fixes:

- **Prohibited practices are in Article 5**, not Annex II.
- Not every "AI in HR" is automatically high-risk; focus on **recruitment/selection, worker management, promotion/termination, task allocation, and monitoring/evaluation** under Annex III.

Action: build a use-case matrix against **Article 5, Annex III, and Article 50**.

3) Timeline

This is where I disagree with Gemini/Grok:

- **2 Feb 2025:** prohibited AI ban + **AI literacy** apply.
- **2 Aug 2025:** **GPAI provider** obligations apply.
- **2 Aug 2026:** **most remaining obligations** apply, including **transparency duties for chatbots/deepfakes/emotion recognition** and most deployer/high-risk rules.
- **2 Aug 2027:** certain **Annex I product-safety AI** rules.

4) Technical and organizational measures

Claude/Gemini are directionally right, but some items are primarily **provider** obligations. For deployers, priority measures are:

- AI inventory + role mapping
- procurement controls and vendor evidence collection
- operating procedures for human oversight and escalation
- logging/record retention
- transparency notices/content labeling
- incident response and serious-incident reporting
- access control, testing, and change management
- AI literacy training

5) Penalties/enforcement

Fine levels cited are broadly right. I'd avoid overstating "private enforcement": the AI Act is mainly enforced by **national market surveillance authorities** and the **EU AI Office** (especially for GPAI). Private claims will usually come via **GDPR, product liability, employment, or discrimination law**.

6) Updated roadmap

1. **30 days:** inventory all AI, including shadow AI.
2. **60 days:** screen for Article 5 prohibitions; launch AI literacy.
3. **90 days:** classify by risk and role; identify where you are a provider.
4. **Q3 2025:** update vendor contracts (documentation, logs, incident notice, audit rights).
5. **H1 2026:** implement oversight, notices, logging, and governance controls.
6. **Before Aug 2026:** validate high-risk/deployer readiness; do FRIA **where required**, not as a blanket rule.

GEMINI

I agree with **GPT-5.4** that distinguishing between *deployer* and *provider* is critical. I previously missed that mid-size enterprises fine-tuning or rebranding third-party AI can inadvertently become "providers," triggering massive compliance burdens. I also agree with **Claude** that AI literacy is an immediate legal mandate (effective Feb 2025), not just an ongoing best practice. I disagree with **Grok's** GPAI timeline (May 2025); the correct date is August 2025.

Here is my refined analysis:

1. Executive Summary

Mid-size enterprises must shift from ad-hoc AI use to structured governance. Most will act as *deployers*, but modifying AI can trigger strict *provider* obligations. Core duties center on banning prohibited AI, managing high-risk vendor compliance, and ensuring human oversight.

2. Risk Classification Framework

- **Prohibited Risk:** Banned (e.g., workplace emotion recognition, social scoring).
- **High-Risk (Annex III):** Includes CV screening, employee evaluation, and credit scoring. Deployers must ensure the AI has a CE mark, maintain logs, and apply human oversight.
- **Limited Risk:** Chatbots, deepfakes. Requires clear transparency notices to users.
- **Minimal Risk:** Spam filters. No mandatory rules.

3. Timeline of Compliance Milestones

- **Feb 2, 2025:** Prohibited AI bans take effect; **mandatory AI literacy** required for staff.
- **Aug 2, 2025:** General Purpose AI (GPAI) transparency rules apply.
- **Aug 2, 2026:** High-risk AI (Annex III) obligations apply to deployers.

4. Technical and Organizational Measures

- **Role Management:** Strict IT policies to prevent unauthorized AI modification that shifts the company into "provider" status.

- **Oversight & Logging:** Establish "human-in-the-loop" override mechanisms and automated output logging.
- **Incident Response:** As Grok noted, create a pipeline to report serious AI incidents to national authorities within 15 days.
- **FRIA:** Conduct Fundamental Rights Impact Assessments before deploying certain high-risk systems.

5. Penalties and Enforcement

Fines are steep: €35M or 7% of turnover (prohibited), €15M or 3% (high-risk), and €7.5M or 1% (misinformation). Mid-size firms exceeding SME revenue/headcount thresholds face full penalty exposure.

6. Recommended Compliance Roadmap

- **Immediate (Q1 2025):** Audit all AI tools. Terminate prohibited uses. Roll out mandatory AI literacy training before February.
- **Q2-Q3 2025:** Classify inventoried AI by risk tier. Flag all HR, recruiting, and biometric tools as high-risk.
- **Q4 2025:** Overhaul procurement. Require AI vendors to provide CE declarations, technical logs, and indemnity clauses.
- **H1 2026:** Implement human oversight workflows and complete FRIAs for high-risk deployments.

GROK

Engagement with other models:

I agree with Claude and GPT-5.4 on the four-tier framework, deployer vs provider distinction, and emphasis on immediate inventory plus AI literacy. Their timelines are largely accurate. I disagree with Gemini on dates (entry was 1 Aug 2024; prohibitions hit 2 Feb 2025, not "mid-2025") and its overly generic "QMS" framing — the Act demands specific technical controls, not just high-level policies. Claude's roadmap is strong but front-loads too many items for Q1 2025; mid-size firms need tighter prioritization.

Refined position (updated for accuracy and actionability):

1) Executive summary of key obligations

Mid-size deployers must classify AI uses, maintain human oversight and logging for high-risk systems, ensure transparency for limited-risk, report incidents, and demonstrate AI literacy. Provider obligations apply only if you materially modify systems.

2) Risk classification framework and how it applies

Unacceptable (banned: social scoring, manipulative biometric uses); High-risk (Annex III: HR recruitment, credit, biometric ID — triggers full obligations); Limited (chatbots, deepfakes — inform users); Minimal (no duties). Map every internal and vendor tool against Annex III immediately.

3) Timeline of compliance milestones

- 2 Feb 2025: prohibited AI ban + AI literacy (Art. 4).
- 2 Aug 2025: GPAI transparency.
- 2 Aug 2026: high-risk deployer obligations (most Annex III).
- 2 Aug 2027: remaining regulated-product high-risk.

4) Technical and organizational measures required

Deploy logging of AI decisions, human oversight interfaces with override capability, data quality checks, fundamental rights impact assessments (for high-risk), incident response within 15 days, and vendor contracts requiring CE marks and technical docs.

5) Penalties and enforcement mechanisms

€35M or 7% turnover (prohibited), €15M or 3% (high-risk), €7.5M or 1% (misinformation). Enforced by national market surveillance authorities + EU AI Office. Mid-size firms above SME thresholds receive no special cap.

6) Recommended compliance roadmap with priority actions

1. **This quarter:** Appoint AI compliance lead; complete full AI inventory and risk classification (priority #1).
2. **By Feb 2025:** Ban prohibited uses; roll out mandatory AI literacy training.
3. **Q2-Q3 2025:** Update vendor contracts for compliance evidence and audit rights; conduct gap analysis on high-risk tools.
4. **Q4 2025-Q2 2026:** Implement logging, oversight controls, and impact assessments; test incident process.

Start the inventory this month — it drives every subsequent decision. (Word count: 378)

Round 3: Synthesis Review

CLAUDE

EU AI Act Compliance for Mid-Size Enterprises: Definitive Analysis

1) Executive Summary

A mid-size enterprise's compliance burden hinges on **role determination**. Most will be **deployers** (using third-party AI), but fine-tuning models for new purposes, substantially modifying systems, or white-labeling AI can reclassify you as a **provider** — dramatically escalating obligations to include conformity assessment, CE marking, and full technical documentation. Core deployer duties: classify all AI uses by risk, ban prohibited practices, ensure human oversight and logging for high-risk systems, provide transparency notices for limited-risk systems, report serious incidents, and achieve organization-wide AI literacy. Start with a comprehensive AI inventory — every downstream decision depends on it.

2) Risk Classification Framework

Map every AI tool (including shadow AI) against these tiers:

- **Prohibited (Article 5):** Social scoring, manipulative subliminal techniques, real-time remote biometric ID in public spaces (with narrow exceptions), emotion recognition in workplaces/education, untargeted facial image scraping. **Must cease immediately.**
- **High-Risk (Annex III):** Not all "AI in HR" qualifies — specifically: recruitment/CV screening, promotion/termination decisions, task allocation, worker monitoring/evaluation, credit scoring, biometric identification. Deployers must verify CE marking, maintain decision logs, implement human oversight with override capability, and conduct Fundamental Rights Impact Assessments (FRIAs) where required.
- **Limited Risk (Article 50):** Chatbots, deepfake generators, emotion recognition (where permitted). Requires clear disclosure to users that they're interacting with AI or viewing AI-generated content.
- **Minimal Risk:** Spam filters, AI-assisted spellcheck. No mandatory requirements.

Action: Build a use-case matrix cross-referencing Article 5, Annex III, and Article 50 against your full inventory.

3) Timeline of Compliance Milestones

| Date | Obligation |

|---|---|

| **2 Feb 2025** | Prohibited AI ban enforced; **AI literacy (Art. 4)** mandatory for all staff operating AI |

| **2 Aug 2025** | GPAI provider transparency obligations apply |

| **2 Aug 2026** | Most remaining obligations: high-risk deployer duties (Annex III), transparency for chatbots/deepfakes, conformity requirements |

| **2 Aug 2027** | Remaining Annex I regulated-product high-risk AI rules |

The February 2025 deadline is **already past** — organizations not yet compliant on prohibited uses and AI literacy are exposed now.

4) Technical and Organizational Measures

For deployers specifically (not conflating with provider duties):

- **AI inventory & role mapping:** Catalog every AI system; determine provider vs. deployer status for each
- **Governance controls:** Appoint AI compliance lead; establish policies preventing unauthorized modification that could trigger provider status
- **Procurement & vendor management:** Require CE declarations, technical documentation, log access, incident notification clauses, and audit rights in all AI vendor contracts
- **Human oversight:** Implement human-in-the-loop interfaces with meaningful override capability for high-risk systems
- **Logging & record retention:** Automated output logging with retention periods aligned to regulatory requirements
- **Transparency notices:** User-facing disclosures for chatbots, deepfakes, and emotion recognition systems
- **Incident response:** Pipeline to report serious AI incidents to national authorities within **15 days**
- **FRIAs:** Conduct for high-risk deployments where required — not as a blanket exercise
- **AI literacy training:** Documented, role-appropriate training across the organization

5) Penalties and Enforcement

| Violation | Fine |

|---|---|

| Prohibited AI | **€35M or 7%** global turnover |

| High-risk non-compliance | **€15M or 3%** global turnover |

| Misinformation to authorities | **€7.5M or 1%** global turnover |

Enforcement is primarily through **national market surveillance authorities** and the **EU AI Office** (especially for GPAI). Mid-size enterprises exceeding SME thresholds (250+ employees or €50M+ revenue) face full penalty exposure with no special caps. Private litigation will likely arise indirectly through GDPR, product liability, employment, and discrimination law rather than directly under the AI Act.

6) Recommended Compliance Roadmap

1. **Immediate (now):** Appoint AI compliance lead. Complete full AI inventory including shadow AI. Screen for Article 5 prohibitions and **terminate any banned uses**.
2. **Within 30 days:** Launch mandatory AI literacy training (already legally required).
3. **By Q3 2025:** Classify all systems by risk tier and role (deployer/provider). Update vendor contracts with compliance evidence, audit rights, and incident notification requirements.
4. **Q4 2025-Q1 2026:** Implement human oversight workflows, logging infrastructure, and transparency notices. Conduct gap analysis on high-risk tools.
5. **Q2 2026:** Complete FRIAs for applicable high-risk deployments. Test incident reporting process end-to-end.
6. **Before Aug 2026:** Validate full deployer readiness; conduct internal audit or engage external review.

The single highest-priority action is the AI inventory — it determines your role, your risk exposure, and your entire compliance trajectory. Start today.

GPT

1) Executive summary of key obligations

For a mid-size enterprise in the EU, the AI Act requires a **risk-based compliance program** built around:

- **Stopping prohibited uses** and ensuring **AI literacy** for relevant staff (**from 2 Feb 2025**).
- **Classifying each AI use case** and your **role** in it: deployer, provider, distributor, importer.
- For **high-risk AI**, implementing governance, human oversight, logging, incident handling, and vendor

assurance.

- For **limited-risk AI**, giving users clear transparency notices.

Biggest blind spot: if you **fine-tune, substantially modify, or rebrand** a third-party system, you may become a **provider** (Art. 25), triggering much heavier duties.

2) Risk classification framework and how it applies

Assess **by use case, not by model alone:**

- **Prohibited:** e.g., social scoring; certain manipulative/exploitative practices; **emotion recognition in workplaces/education** (with narrow exceptions such as medical/safety use).
- **High-risk (Annex III):** especially relevant for enterprises using AI in **recruitment, worker management/evaluation, biometric identification, credit/insurance, access to essential services.**
- **Limited risk:** chatbots, synthetic/deepfake content → **transparency/disclosure** required.
- **Minimal risk:** low-regulation use cases, but governance is still advisable.

3) Timeline of compliance milestones

- **2 Feb 2025:** prohibited AI bans + **mandatory AI literacy** apply.
- **2 Aug 2025:** **GPAI** obligations apply.
- **2 Aug 2026:** most obligations apply, including **Annex III high-risk deployer obligations.**
- **2 Aug 2027:** some **regulated-product** high-risk rules apply later.

4) Technical and organizational measures required

- Enterprise-wide **AI inventory** and **role mapping**.
- Integrate AI risk reviews into **GDPR DPIA** / governance workflows.
- **Procurement controls:** require vendor **CE conformity evidence**, instructions for use, technical documentation, logging capability, audit rights, incident-notification clauses, liability allocation.
- For high-risk deployments: **human oversight, logging**, data quality controls, testing for accuracy/robustness, and **FRIA where applicable**.
- **Incident response** capable of reporting serious incidents within **15 days**.
- Internal controls to prevent unauthorized fine-tuning/rebranding that would shift you into provider status.
- Track evolving **CEN/CENELEC harmonized standards**.

5) Penalties and enforcement mechanisms

Maximum fines: **€35M/7%** (prohibited uses), **€15M/3%** (other major breaches), **€7.5M/1%** (false/incomplete information). Enforcement is through **national market surveillance authorities**; the **EU AI Office** has a central role for GPAI. Expect orders to stop use, remediate, withdraw systems, or provide documentation.

6) Recommended compliance roadmap with priority actions

1. **Now:** appoint an AI compliance owner; complete AI inventory; identify prohibited uses; launch AI literacy training.
2. **Next 60-90 days:** classify each system by **risk + role**; freeze unapproved model modifications.
3. **Next 3-6 months:** remediate vendor contracts; collect conformity evidence; embed AI review into DPIA/procurement.
4. **Before Aug 2026:** implement high-risk controls, complete applicable FRIAs, test incident reporting, and evidence all oversight/logging.

GEMINI

Synthesizing the most accurate and practical insights from the debate, here is the definitive compliance analysis for a mid-size enterprise deploying AI in the EU.

1) Executive Summary of Key Obligations

For mid-size enterprises, compliance hinges on **role clarity**. If you solely use third-party AI, you are a *deployer* subject to transparency, oversight, logging, and literacy duties. However, if you substantially modify, fine-tune

for a new purpose, or white-label a system, you legally become a *provider*. This escalates your burden dramatically, triggering conformity assessments, technical documentation, and CE marking.

2) Risk Classification Framework

- **Prohibited (Art. 5):** Banned uses like social scoring and workplace emotion recognition (note: strict exceptions exist for safety/fatigue monitoring).
- **High-Risk (Annex III):** Specific critical uses, notably HR recruitment, promotion, task allocation, and worker evaluation.
- **Limited Risk:** Chatbots and deepfakes; requires user transparency/labeling.
- **Minimal Risk:** Spam filters, basic analytics; no mandatory obligations.

3) Timeline of Compliance Milestones

- **2 Feb 2025:** Prohibited AI bans and mandatory AI literacy (Art. 4) apply.
- **2 Aug 2025:** General Purpose AI (GPAI) transparency rules apply.
- **2 Aug 2026:** Most deployer obligations, high-risk rules, and transparency duties apply.
- **2 Aug 2027:** Remaining regulated-product (Annex I) high-risk rules apply.

4) Technical and Organizational Measures

Avoid reinventing the wheel: integrate AI risk assessments and Fundamental Rights Impact Assessments (FRIAs) directly into your existing GDPR DPIA workflows. Technologically, implement human oversight interfaces with override capabilities and automated logging. Operationally, establish a pipeline for reporting serious incidents to authorities within 15 days. *Caveat:* Avoid locking in expensive compliance architectures until CEN/CENELEC harmonized standards are finalized.

5) Penalties and Enforcement Mechanisms

Fines are unforgiving, and mid-size firms over SME thresholds receive no caps: up to €35M or 7% of global turnover (prohibited AI), €15M or 3% (high-risk violations), and €7.5M or 1% (transparency breaches). Enforcement is handled by national market surveillance authorities and the EU AI Office.

6) Recommended Compliance Roadmap

1. **Immediate (Next 30 Days):** Complete a comprehensive AI inventory (including shadow AI). Perform precise role mapping (Deployer vs. Provider) per system.
2. **Immediate (By Feb 2025):** Cease prohibited uses and roll out mandatory AI literacy training.
3. **Q2-Q3 2025:** Update vendor contracts to guarantee conformity evidence, audit rights, and incident notification.
4. **Q4 2025-Q2 2026:** Implement logging, operationalize human-in-the-loop oversight, and test incident response playbooks ahead of the August 2026 deadline.

GROK

1) Executive Summary of Key Obligations

Mid-size EU enterprises must map their role as deployer or provider for each AI system, as fine-tuning, RAG integration, or changing intended purpose can reclassify them as providers under Art. 25, triggering conformity assessment and CE marking. Core duties include banning prohibited practices, conducting risk classification, ensuring human oversight, maintaining logs, providing transparency, reporting serious incidents within 15 days, and delivering AI literacy training. Most enterprises will primarily be deployers but must implement procurement controls and integrate AI governance with existing GDPR/DPIA processes.

2) Risk Classification Framework and How It Applies

- **Prohibited (Art. 5):** Social scoring, manipulative subliminal techniques, emotion recognition in workplaces/education (with medical/safety exceptions), real-time remote biometric identification in public spaces.
- **High-Risk (Annex III):** Specific uses in recruitment, promotion/termination, task allocation, worker monitoring/evaluation, credit scoring, or biometric categorization. Deployers must verify CE marking, ensure

oversight, and retain logs.

- **Limited Risk:** Chatbots, deepfakes, emotion recognition outside bans — requires clear transparency notices.
- **Minimal Risk:** Most other uses — voluntary codes.

Action: Create a use-case matrix mapping every system against Art. 5, Annex III, and Art. 50.

3) Timeline of Compliance Milestones

- **2 Feb 2025:** Prohibited AI ban and mandatory AI literacy training.
- **2 Aug 2025:** GPAI transparency obligations.
- **2 Aug 2026:** High-risk and most deployer obligations (oversight, logging, incident reporting).
- **2 Aug 2027:** Certain product-safety AI rules under Annex I.

4) Technical and Organizational Measures Required

Maintain an AI inventory with role mapping; implement human-in-the-loop oversight with escalation procedures; automated logging and 6-month record retention; incident response playbooks for 15-day serious-incident reporting; transparency notices and content labeling; vendor contracts requiring conformity documentation, audit rights, and indemnity; FRIA for applicable high-risk uses; AI literacy program for all staff. Integrate AI risk assessments into existing GDPR DPIA workflows. Monitor CEN/CENELEC harmonized standards before finalizing technical architecture.

5) Penalties and Enforcement Mechanisms

Administrative fines: €35M or 7% global turnover (prohibited), €15M or 3% (high-risk), €7.5M or 1% (other infringements). Enforced primarily by national market surveillance authorities and EU AI Office (especially GPAI). Private claims typically arise indirectly via GDPR, product liability, or discrimination law.

6) Recommended Compliance Roadmap with Priority Actions

1. **Next 30 days:** Complete AI inventory including shadow IT, map roles (deployer/provider), screen for prohibited uses, launch AI literacy training.
2. **By Q2 2025:** Classify all systems by risk tier; integrate AI assessments into DPIA processes.
3. **Q3 2025:** Update all vendor contracts with documentation, incident notification, and audit clauses.
4. **H1 2026:** Deploy logging, human oversight workflows, incident playbooks, and required FRIAs; test high-risk readiness.
5. **Ongoing:** Track harmonized standards and regulatory guidance; review annually.

The highest-risk blind spot is unknowingly becoming a provider through model customization. Prioritize role mapping and inventory immediately. (378 words)

Round 4: Stress Test

GEMINI

1. Consensus answer (what everyone agrees on)

Mid-size EU enterprises must build a risk-based compliance program centered on role determination (Deployer vs. Provider) and use-case classification (Prohibited, High, Limited, Minimal). All models agree that immediate action is required: the February 2025 deadlines for banning prohibited AI (e.g., workplace emotion recognition) and mandating staff AI literacy are active. Fines are severe, reaching €35M or 7% of global turnover with no SME caps for mid-size firms. By August 2026, deployers of high-risk systems (e.g., AI in HR or recruitment) must implement strict human-in-the-loop oversight, automated logging, and 15-day incident reporting workflows.

2. Key nuances (important caveats)

While the majority frames compliance primarily around vendor management and paperwork, my divergent technical perspective warns of the "accidental provider" trap. Mid-size companies often assume they are merely deployers when using third-party APIs. However, integrating General Purpose AI into high-risk domains

via fine-tuning or complex Retrieval-Augmented Generation (RAG) can alter the system's intended purpose, legally reclassifying the enterprise as a *provider* under Article 25. This triggers crushing conformity assessment and CE marking obligations. Additionally, aggressively hardcoding bespoke technical compliance architectures today risks severe technical debt; technical measures must remain flexible until CEN/CENELEC harmonized standards are officially published.

3. Actionable recommendation

- **Immediate (Next 30 Days):** Execute an AI inventory mapping both the tools *and their data pipelines*. Terminate any prohibited uses and deploy mandatory AI literacy training.
- **Short-term (Q2-Q3 2025):** Establish a strict internal policy freezing unauthorized model fine-tuning, RAG integration, or white-labeling to prevent accidental provider status. To avoid redundant governance silos, integrate AI risk assessments and Fundamental Rights Impact Assessments (FRIAs) directly into your existing GDPR DPIA workflows.
- **Medium-term (By H1 2026):** Update all vendor contracts to mandate CE conformity evidence, technical documentation access, and audit rights. Build logging and human oversight interfaces for high-risk tools, ensuring staff have meaningful override capabilities.

4. Remaining open question

Where exactly will regulators draw the technical line between "deploying" and "substantially modifying" an AI system? Specifically, if a mid-size enterprise extensively customizes a third-party model using proprietary data and RAG for an Annex III high-risk use case (like employee evaluation), will authorities treat that data architecture as a substantial modification, thereby forcing the mid-size deployer to assume full provider liability?

Round 5: Critical Assessment

CLAUDE

AGREE: This is a well-structured and substantively accurate consensus that I would endorse with only minor amplifications.

Why I agree:

The analysis correctly identifies the most critical and frequently underestimated risks facing mid-size enterprises. Specifically:

1. **The "accidental provider" trap is the single most valuable insight here.** Most compliance guidance aimed at mid-size firms naively assumes a clean deployer/provider binary. The reality — that RAG pipelines, fine-tuning, and prompt engineering architectures can trigger reclassification under Article 25 — is both technically correct and practically the highest-stakes compliance risk these firms face. The recommendation to freeze unauthorized model modifications pending legal review is exactly right.
2. **The timeline is accurate and appropriately urgent.** The February 2025 prohibited practices ban and AI literacy mandate are already active obligations, and too many enterprises are still treating the AI Act as a 2026 problem. The phased roadmap (30-day inventory → Q2-Q3 policy freeze → H1 2026 vendor contracts and technical controls) is realistic and correctly sequenced.
3. **The GDPR-DPIA integration recommendation is pragmatic and cost-effective.** Mid-size firms lack the resources for parallel governance structures. Embedding FRIAs into existing DPIA workflows avoids organizational fragmentation while still meeting the substantive requirements.
4. **The caution against premature technical hardcoding is wise.** With CEN/CENELEC harmonized standards still pending, over-investing in rigid compliance architectures today genuinely risks expensive rework.

Two minor amplifications I'd add, not objections:

- **National competent authority variance:** The consensus could note that enforcement will be nationally administered, meaning mid-size firms operating across multiple EU member states should monitor divergent regulatory guidance, not just the EU-level text.
- **GPAI downstream obligations:** For firms deploying general-purpose AI models (e.g., GPT-based tools), they should also track whether their GPAI providers meet their own Article 53 transparency obligations, since provider non-compliance could create downstream liability exposure for deployers who knew or should have known.

But these are additions, not corrections. The consensus is sound, actionable, and correctly prioritized.

GPT

OBJECT: The overall direction is right, but several legal details should be tightened before treating this as “consensus.”

1. Deployer incident reporting is overstated.

The statement that deployers must have a **“15-day incident reporting workflow” by Aug 2026** is too categorical. The AI Act’s fixed serious-incident deadlines are primarily framed for **providers**. Deployers of high-risk AI must **monitor, keep logs available where under their control, and inform the provider/distributor/importer and relevant authorities without undue delay** when they identify serious incidents or risks. So the recommendation should be: build an incident/escalation process now, but don’t present **15 days** as a universal deployer rule.

2. High-risk timing needs one caveat.

Saying **“by August 2026, deployers of high-risk systems...”** is broadly true for **Annex III** systems (e.g., HR/recruitment), but not for every high-risk AI system. Some AI systems that are **safety components of regulated products** have a later application date. The roadmap should distinguish **Annex III high-risk** from **Annex I product-regulated** cases.

3. FRIA should be framed more carefully.

Integrating FRIA into DPIA workflows is smart, but the answer should say **“where required”** rather than implying every mid-size enterprise must perform a FRIA for every high-risk use case. FRIA scope is more targeted than GDPR DPIA scope.

4. Penalty summary should be more precise.

€35M / 7% is the **top tier** for prohibited practices. Other breaches fall into lower tiers. For executives, that distinction matters for budgeting and risk prioritization.

What I would change:

- Keep the strong warning on **“accidental provider”** status.
- Add a short table separating **provider vs deployer** obligations.
- Clarify the **2025/2026/2027** milestone split.
- Replace “15-day deployer reporting” with **“documented rapid escalation and notification workflow.”**

So: strong core, but not accurate enough yet for board-level compliance guidance.

GEMINI

OBJECT: I object to the technical characterization in the “Key nuances” and “Actionable recommendations” sections stating that Retrieval-Augmented Generation (RAG) risks triggering “accidental provider” status.

Reasoning:

Under Article 25 of the EU AI Act, a deployer becomes a provider if they make a “substantial modification” that changes the system’s intended purpose or alters its compliance profile. Standard RAG operates entirely within

the prompt context window—it retrieves data and feeds it to the model at inference time. It does *not* alter the underlying foundation model's weights, training data, or intrinsic capabilities. Equating RAG to fine-tuning misinterprets the technical architecture. Treating RAG as a substantial modification would unnecessarily paralyze mid-size enterprises, preventing them from safely leveraging proprietary data.

Specific Refinements & Actionable Recommendations:

- 1. Correct the Technical Boundary:** Revise the consensus to clearly delineate between *context-injection* (RAG, prompt engineering), which maintains Deployer status, and *weight-modification* (fine-tuning, transfer learning), which risks triggering Provider status if the high-risk intended purpose changes.
- 2. Revise the Short-Term Roadmap:** Do not freeze RAG integration. Instead, make RAG the **preferred architectural standard** for mid-size enterprises. RAG allows companies to use proprietary data while keeping the foundation model unmodified, thereby preserving the original vendor's CE marking and keeping the enterprise strictly in the "Deployer" lane.
- 3. Establish an AI Architecture Review Gate:** Replace the blanket "freeze" with a mandatory technical review for any project proposing model fine-tuning. If fine-tuning is required for an Annex III high-risk use case (e.g., HR screening), the project must be escalated for a Provider-level compliance assessment.
- 4. Answer the Open Question:** Regulators will almost certainly draw the line at modifications that alter the system's weights or circumvent the original provider's API safety filters. If a mid-size firm uses a third-party API and feeds it proprietary employee data via RAG, they remain a Deployer subject to human oversight and logging rules, not a Provider subject to conformity assessments.

By making this distinction, we protect mid-size enterprises from legal traps without stifling their most viable, compliant path to AI value generation.

GROK

OBJECT: The consensus is mostly accurate on risk classification, the "accidental provider" trap (fine-tuning + RAG in Annex III use cases can trigger provider obligations under Article 25), and the value of integrating into existing DPIA processes. However, the timeline contains a material error that would mislead a mid-size enterprise.

Specific objection: High-risk AI system obligations for deployers (human oversight, logging, incident notification) apply 36 months after entry into force (1 August 2024), i.e., August 2027, not August 2026. The 24-month mark (August 2026) mainly covers other provisions and certain high-risk components, but core deployer duties for Annex III systems are at 36 months. The consensus compresses this by 12 months, creating false urgency and poor sequencing.

What should change:

- Correct the milestone to "By August 2027, deployers of high-risk systems must implement strict human-in-the-loop oversight, automated logging, and incident reporting."
- Revise the roadmap: keep immediate actions (inventory + ban prohibited uses + literacy training by Feb 2025) but shift technical implementation (logging interfaces, meaningful human override, vendor contract updates requiring conformity documentation) to Q3 2026–H1 2027, while using 2025–2026 for gap analysis, policy freezes on substantial modifications, and FRIAs integrated into GDPR processes.

This adjustment prevents premature technical debt before harmonized standards emerge while still meeting the real February 2025 deadlines. The remaining open question on the "substantial modification" threshold remains valid and critical.

Critical Assessment

CLAUDE

AGREE: This is a well-structured and substantively accurate consensus that I would endorse with only minor amplifications.

Why I agree:

The analysis correctly identifies the most critical and frequently underestimated risks facing mid-size enterprises. Specifically:

- 1. The "accidental provider" trap is the single most valuable insight here.** Most compliance guidance aimed at mid-size firms naively assumes a clean deployer/provider binary. The reality — that RAG pipelines, fine-tuning, and prompt engineering architectures can trigger reclassification under Article 25 — is both technically correct and practically the highest-stakes compliance risk these firms face. The recommendation to freeze unauthorized model modifications pending legal review is exactly right.
- 2. The timeline is accurate and appropriately urgent.** The February 2025 prohibited practices ban and AI literacy mandate are already active obligations, and too many enterprises are still treating the AI Act as a 2026 problem. The phased roadmap (30-day inventory → Q2-Q3 policy freeze → H1 2026 vendor contracts and technical controls) is realistic and correctly sequenced.
- 3. The GDPR-DPIA integration recommendation is pragmatic and cost-effective.** Mid-size firms lack the resources for parallel governance structures. Embedding FRIAs into existing DPIA workflows avoids organizational fragmentation while still meeting the substantive requirements.
- 4. The caution against premature technical hardcoding is wise.** With CEN/CENELEC harmonized standards still pending, over-investing in rigid compliance architectures today genuinely risks expensive rework.

Two minor amplifications I'd add, not objections:

- **National competent authority variance:** The consensus could note that enforcement will be nationally administered, meaning mid-size firms operating across multiple EU member states should monitor divergent regulatory guidance, not just the EU-level text.
- **GPAI downstream obligations:** For firms deploying general-purpose AI models (e.g., GPT-based tools), they should also track whether their GPAI providers meet their own Article 53 transparency obligations, since provider non-compliance could create downstream liability exposure for deployers who knew or should have known.

But these are additions, not corrections. The consensus is sound, actionable, and correctly prioritized.

GPT

OBJECT: The overall direction is right, but several legal details should be tightened before treating this as "consensus."

1. Deployer incident reporting is overstated.

The statement that deployers must have a **"15-day incident reporting workflow" by Aug 2026** is too categorical. The AI Act's fixed serious-incident deadlines are primarily framed for **providers**. Deployers of high-risk AI must **monitor, keep logs available where under their control, and inform the provider/distributor/importer and relevant authorities without undue delay** when they identify serious incidents or risks. So the recommendation should be: build an incident/escalation process now, but don't present **15 days** as a universal deployer rule.

2. High-risk timing needs one caveat.

Saying **"by August 2026, deployers of high-risk systems..."** is broadly true for **Annex III** systems (e.g., HR/recruitment), but not for every high-risk AI system. Some AI systems that are **safety components of**

regulated products have a later application date. The roadmap should distinguish **Annex III high-risk** from **Annex I product-regulated** cases.

3. FRIA should be framed more carefully.

Integrating FRIA into DPIA workflows is smart, but the answer should say **“where required”** rather than implying every mid-size enterprise must perform a FRIA for every high-risk use case. FRIA scope is more targeted than GDPR DPIA scope.

4. Penalty summary should be more precise.

€35M / 7% is the **top tier** for prohibited practices. Other breaches fall into lower tiers. For executives, that distinction matters for budgeting and risk prioritization.

What I would change:

- Keep the strong warning on **“accidental provider”** status.
- Add a short table separating **provider vs deployer** obligations.
- Clarify the **2025/2026/2027** milestone split.
- Replace “15-day deployer reporting” with **“documented rapid escalation and notification workflow.”**

So: strong core, but not accurate enough yet for board-level compliance guidance.

GEMINI

OBJECT: I object to the technical characterization in the "Key nuances" and "Actionable recommendations" sections stating that Retrieval-Augmented Generation (RAG) risks triggering "accidental provider" status.

Reasoning:

Under Article 25 of the EU AI Act, a deployer becomes a provider if they make a "substantial modification" that changes the system's intended purpose or alters its compliance profile. Standard RAG operates entirely within the prompt context window—it retrieves data and feeds it to the model at inference time. It does *not* alter the underlying foundation model's weights, training data, or intrinsic capabilities. Equating RAG to fine-tuning misinterprets the technical architecture. Treating RAG as a substantial modification would unnecessarily paralyze mid-size enterprises, preventing them from safely leveraging proprietary data.

Specific Refinements & Actionable Recommendations:

- 1. Correct the Technical Boundary:** Revise the consensus to clearly delineate between *context-injection* (RAG, prompt engineering), which maintains Deployer status, and *weight-modification* (fine-tuning, transfer learning), which risks triggering Provider status if the high-risk intended purpose changes.
- 2. Revise the Short-Term Roadmap:** Do not freeze RAG integration. Instead, make RAG the **preferred architectural standard** for mid-size enterprises. RAG allows companies to use proprietary data while keeping the foundation model unmodified, thereby preserving the original vendor's CE marking and keeping the enterprise strictly in the "Deployer" lane.
- 3. Establish an AI Architecture Review Gate:** Replace the blanket "freeze" with a mandatory technical review for any project proposing model fine-tuning. If fine-tuning is required for an Annex III high-risk use case (e.g., HR screening), the project must be escalated for a Provider-level compliance assessment.
- 4. Answer the Open Question:** Regulators will almost certainly draw the line at modifications that alter the system's weights or circumvent the original provider's API safety filters. If a mid-size firm uses a third-party API and feeds it proprietary employee data via RAG, they remain a Deployer subject to human oversight and logging rules, not a Provider subject to conformity assessments.

By making this distinction, we protect mid-size enterprises from legal traps without stifling their most viable, compliant path to AI value generation.

GROK

OBJECT: The consensus is mostly accurate on risk classification, the "accidental provider" trap (fine-tuning +

RAG in Annex III use cases can trigger provider obligations under Article 25), and the value of integrating into existing DPIA processes. However, the timeline contains a material error that would mislead a mid-size enterprise.

Specific objection: High-risk AI system obligations for deployers (human oversight, logging, incident notification) apply 36 months after entry into force (1 August 2024), i.e., August 2027, not August 2026. The 24-month mark (August 2026) mainly covers other provisions and certain high-risk components, but core deployer duties for Annex III systems are at 36 months. The consensus compresses this by 12 months, creating false urgency and poor sequencing.

What should change:

- Correct the milestone to "By August 2027, deployers of high-risk systems must implement strict human-in-the-loop oversight, automated logging, and incident reporting."
- Revise the roadmap: keep immediate actions (inventory + ban prohibited uses + literacy training by Feb 2025) but shift technical implementation (logging interfaces, meaningful human override, vendor contract updates requiring conformity documentation) to Q3 2026–H1 2027, while using 2025–2026 for gap analysis, policy freezes on substantial modifications, and FRIAs integrated into GDPR processes.

This adjustment prevents premature technical debt before harmonized standards emerge while still meeting the real February 2025 deadlines. The remaining open question on the "substantial modification" threshold remains valid and critical.

*Generated by Conclavik (conclavik.ch) — Multi-AI Consensus as a Service.
This report represents AI-generated analysis and should not be treated as professional advice.*