LLM Security & Regulation Essentials

OWASP LLM Top-10 & EU AI Act Quick-Reference

OWASP LLM Top-10 (2024)

- 01 – Prompt Injection
- 02 – Insecure Output Handling
- 03 – Training Data Poisoning
- 04 – Model Denial of Service
- 05 – Supply-Chain Vulnerabilities
- 06 – Sensitive Info Disclosure
- 07 – Over-reliance / Over-trust
- 08 – Excessive Agency
- 09 – Privacy Violations
- 10 – Policy Bypass / Jailbreak

EU AI Act – Key Points for Prompt Engineers

- Risk-based tiers: Unacceptable, High, Limited, Minimal
- High-risk systems → strict data-gov, transparency, logging, human oversight
- User-facing gen-AI → disclosure obligations (deepfake labelling)
- Prohibited: social scoring, real-time biometric ID in public (few exceptions)
- Fines: up to €35 M or 7 % global turnover

Prompt-Level Compliance Tips

- Map each requirement to an explicit prompt clause or policy layer.
- Log user consent & PII handling steps inside the system prompt.
- Maintain prompt version IDs and link them to audit trails.
- Combine automated moderation with human-in-the-loop for high-risk tiers.