

AI as a Black Box: What It Means for High-Stakes Deployments

MANAGEMENT SUMMARY & CORE TAKEAWAY

AI models are black boxes because no one can trace how a language model retrieves its compressed knowledge to produce a specific output. In high-stakes settings like diagnosis or credit decisions, that's a real liability and governance risk. The EU AI Act (Art. 13) has required sufficient transparency for high-risk systems since 2 August 2026—Switzerland is taking a sector-specific path with the Council of Europe AI Convention ratification instead of blanket EU AI Act adoption.

CONTEXT & EXECUTIVE RELEVANCE

AI models are black boxes because no one—not even their developers—can fully trace how a language model retrieves its compressed knowledge to produce a specific output. In practice, this means that when you deploy AI for high-stakes tasks like medical diagnosis, credit decisions, or legal risk assessment, you're relying on a system that has processed more data than any human ever could—yet cannot show its work.

Geoffrey Hinton, one of the founding figures of deep learning, has described the fundamental difference between human and machine learning with unusual precision. The human brain learns over roughly two billion seconds of lived experience using a comparatively small amount of data. An LLM learns in a handful of training runs on a vastly larger volume of data—despite having far fewer connections than the brain.

CORE ACTION ITEMS

- 01 Medical diagnostic and triage systems
- 02 Judicial risk scoring and recidivism assessments
- 03 Credit scoring and insurance underwriting decisions
- 04 Autonomous control of critical infrastructure

SWISS COMPLIANCE & ARCHITECTURE GUIDELINES

- ☑ Full compliance with revised Swiss Data Protection Act (revDSG).
- ☑ Zero vendor lock-in on closed proprietary models.
- ☑ Seamless integration into Abacus, Bexio and M365 ecosystems.

Non-Binding Process Analysis

Let us review in 20 minutes where AI agents create immediate measurable efficiency in your workflows.

ki-outsourcing.ch