LLMs Without Leaks: Keep Your Code, Data & IP Where They Belong

What are LLMs?

AI models trained on vast text data Examples: GPT (175B parameters)

Applications: Text generation, translation, customer support

The Problem: Data Leaks in LLMs

Types of leaks: Training data exposure, prompt leaks, model inversion

Consequences: Privacy breaches, IP theft, competitive loss

Data Leaks and Their Impact

Real-world examples: Samsung (2023), Healthcare breach

Economic impact: Avg. data breach cost \$4.45M (IBM, 2023)

Incident	Impact	Lesson Learned
Samsung (ChatGPT)	Exposed source code	Enforce data input policies
Healthcare Provider	Patient data leaked	Sanitize data before use
Tech Startup	IP stolen via prompt	Monitor prompt interactions

Technologies for Securing LLMs

Differential Privacy: Adds noise to protect privacy

Federated Learning: Trains on decentralized data

Homomorphic Encryption: Processes encrypted data (emerging)

Data Privacy Solutions for LLMs

Data Privacy Vaults: Secure repositories (e.g., Skyflow)

Tools: LLM Shield (Patented.ai) filters sensitive data

Best Practices for Using LLMs Securely

Sanitize training data

Restrict access with controls

Audit regularly

Train employees

Regulatory Compliance and LLMs

Key regulations: GDPR, HIPAA, CCPA

Compliance needs: Data protection, erasure rights

Regulation	Requirement	Compliance Strategy
GDPR	Right to erasure	Use differential privacy
HIPAA	Protect health info	Deploy data privacy vaults
ССРА	Consumer data rights	Filter inputs with tools

Implementing Secure LLMs

Steps: Assess data, select tech, set controls, train staff, audit

Future Trends and Conclusion

Future: Privacy-preserving AI, new protocols Summary: Secure LLMs to protect assets Call to action: Prioritize AI security now