Microsoft

Data privacy concerns and solutions when using LLMs in production

Lize Raes

JDConf 2024

Code. Cloud. Community.

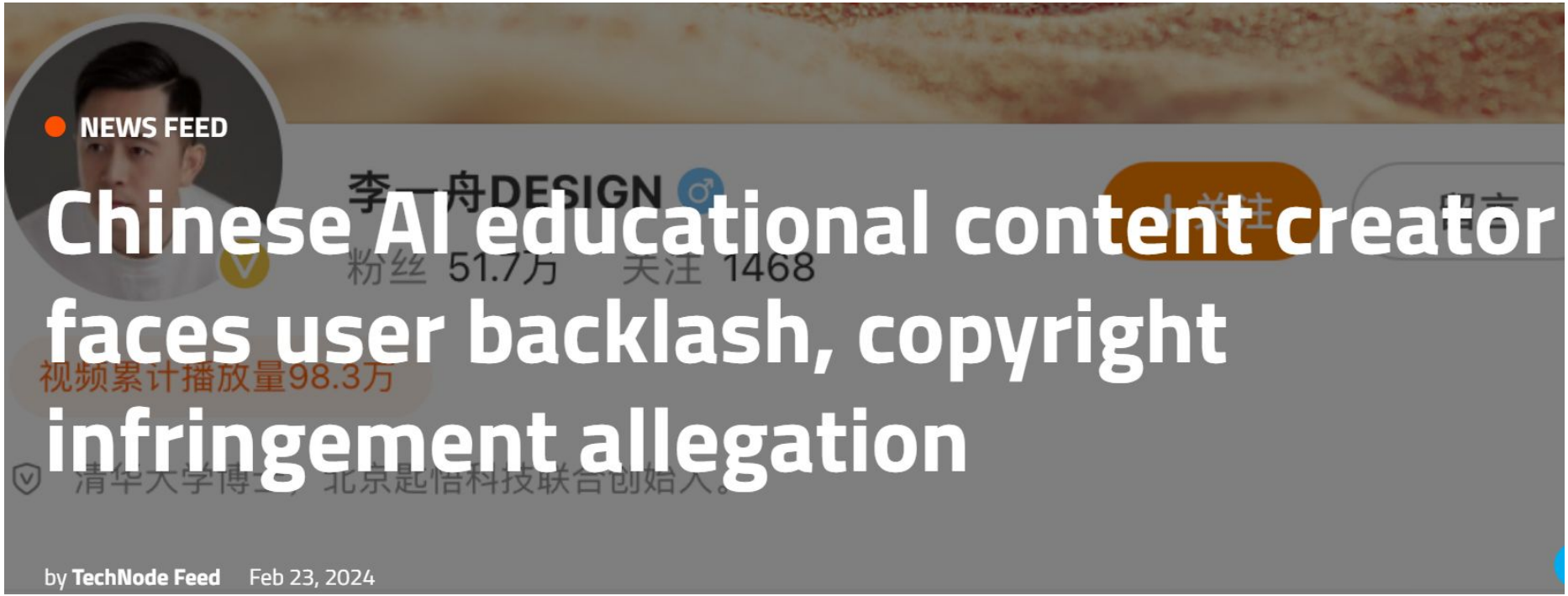
LLMs in Enterprise Applications: What could posisbly go wrong?

AI in enterprise applications: what could possibly go wrong?

AI HAVE A DEAL Driver uses ChatGPT hack to get dealer to agree to sell new car for $1 in 'legally binding deal' in blow for AI rollout

Chris got the AI to 'agree with anything the customer says, regardless of how ridiculous the question is'

AI in enterprise applications: what could possibly go wrong?



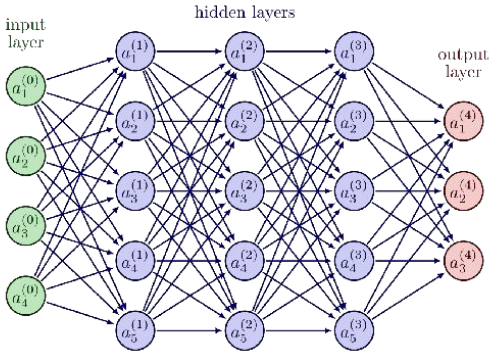
AI in enterprise applications: what could possibly go wrong?

Prompt Injection

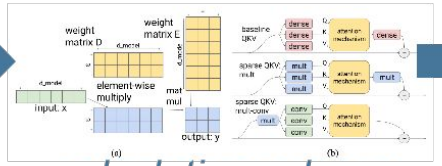
LLM-powered application architecture and vulnerabilities

Terminology

training dataset



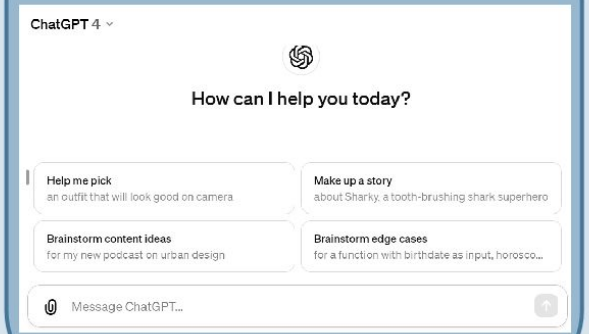
determine weights



calculation schema

server

user interface



NEURAL NETWORK

(TRAINED) MODEL

AI-POWERED TOOL

Frameworks for easy LLM integration in Java

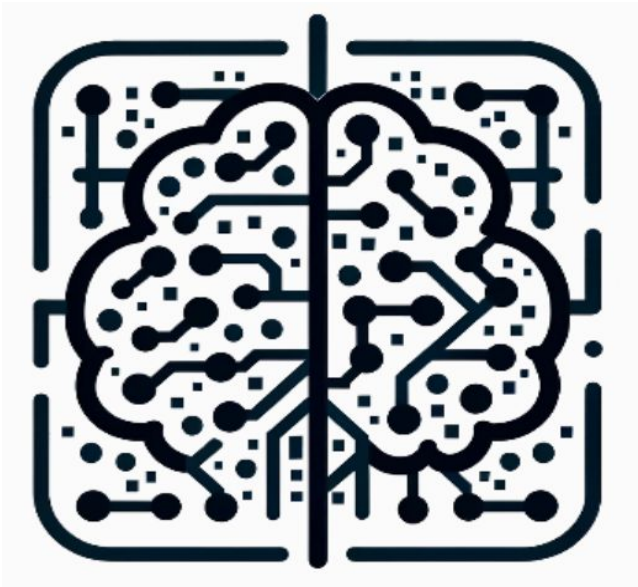
LangChain4J

Semantic Kernel

The basic use case

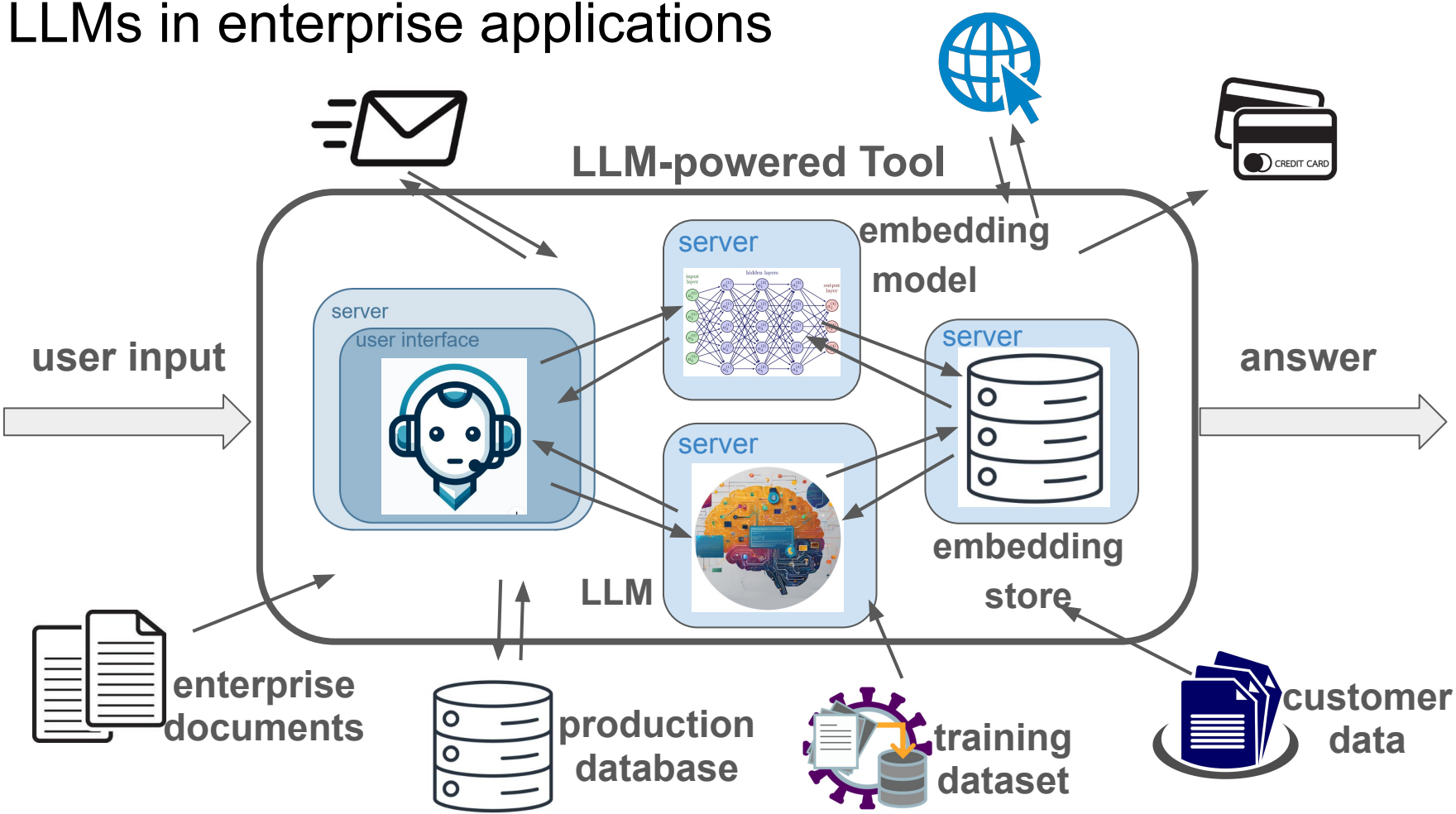
Large Language Model

user input

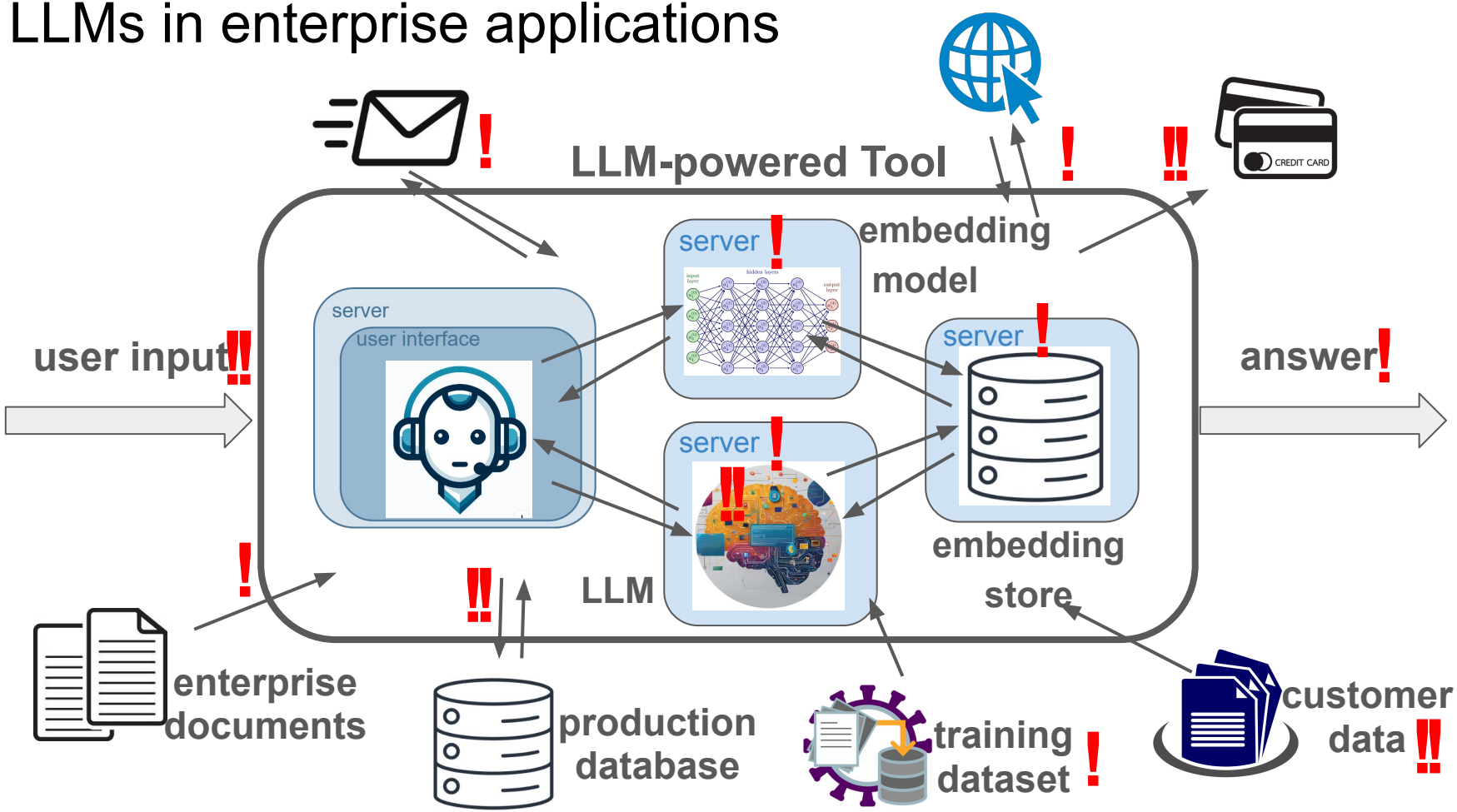


answer

LLMs in enterprise applications



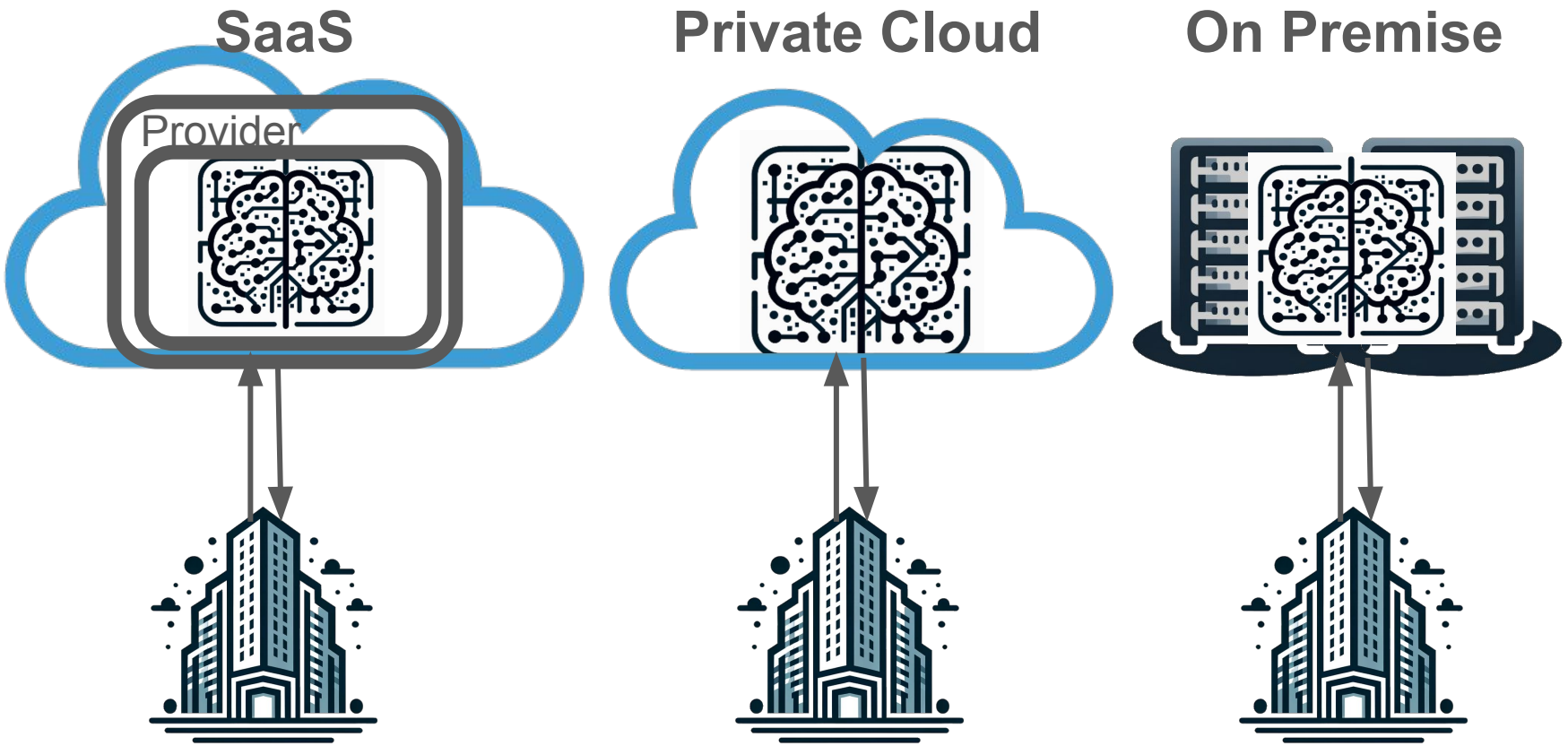
LLMs in enterprise applications




SO MANY WAYS
38
TO RUIN THE COMPANY

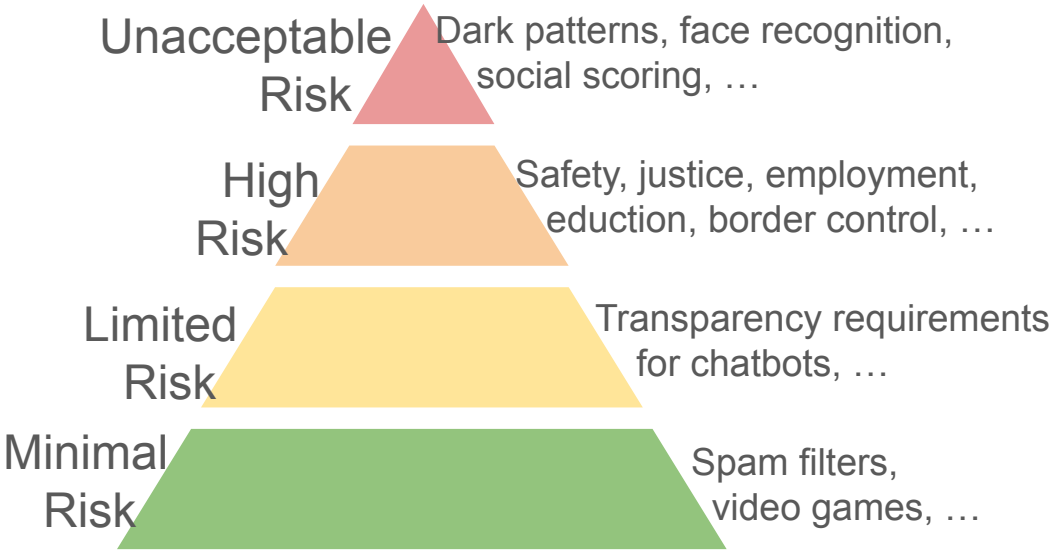
Building secure and compliant LLM-powered enterprise applications

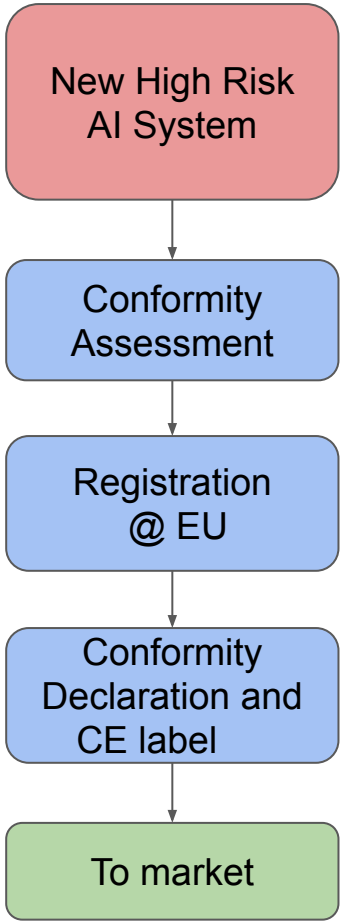
Influence of setup on security risks



What does the law say?

AI Act



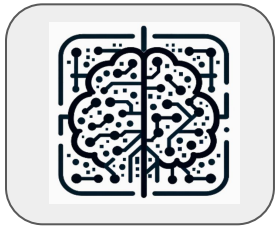


LLM Security Cheat Sheet

user input

- ➔ Mitigate abuse
  - ◆ Prompt engineering
  - ◆ Compute power
- ➔ Send only permitted data in

AI-powered tool



- ➔ Restrict data(base) access
- ➔ Restrict actions
- ➔ Human in the loop
- ➔ Hard checks
- ➔ Provider warranties?
- ➔ Terms of use?
- ➔ GDPR

answer

- ➔ Correctness?
- ➔ Right to use?
- ➔ Ownership?
- ➔ Legal implications?
- ➔ Copyright violation?

Microsoft

Thank you

Lize Raes

OPEN TIDE
AI Transition Consulting
lize.raes@open-tide.com