Responsible AI: Maintaining your Standards with Emerging Tech

MAY 24, 2024

JON NEIDITZ — KILPATRICK TOWNSEND & STOCKTON LLP

DEREK VENTLING — KIRKLAND & ELLIS LLP

Agenda

1. What is Generative AI and How is it Most Useful?
2. Consumer's GenAI's Limitations and Current Self-Help
3. Enterprise Platform Opportunities and Limitations
4. Particular Considerations: AI Risks
5. Key Principles of Responsible AI
6. Building an AI Program
7. Legal Ethics and AI
8. Guardrails

What is Generative AI & How is it Most Useful?

Generative AI uses machine learning to create- not just extract, sort, find or make decisions about - content

Predictive, Extractive AI	Generative AI
▶ Compare data on strategies, plans and frameworks	▶ Produce drafts of alternative strategies, plans and frameworks
▶ Detect patterns across a massive number of documents ("big data")	▶ Generate alternative new drafts of any type of document
▶ Improve and streamline many business and operational functions	▶ Replace business/operational functions with ordinary conversations
▶ Talking to your data	▶ Talking to something new

3

3

Consumer GenAI's Limitations & Current Self-Help
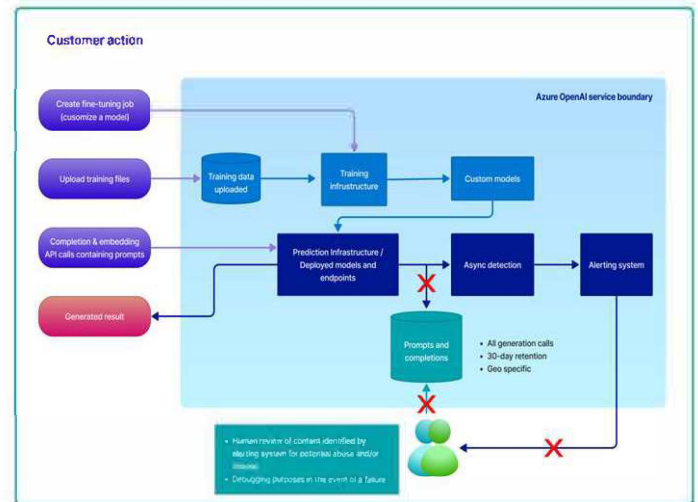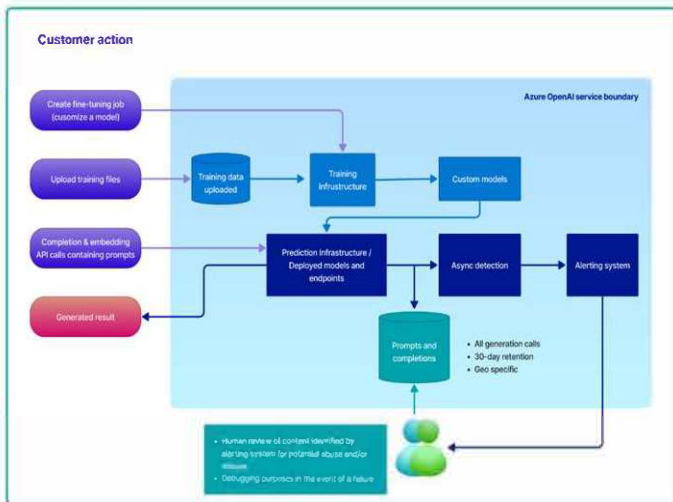
Overview

▶ **Hallucinations** and **Persuasion**
- ▶ Plausibility, no reference to databases
- ▶ "Superhuman persuasion"
- ▶ Fix: Interrogate & verify outputs; all outputs are *drafts.*

▶ **Token Length**
- ▶ Places limitations on how much generative AI can handle in a single sequence, including both input and output.
- ▶ Fixes: Outlines, sections, dialogues, GPTs.

▶ **Memory Limitations**
- ▶ ChatGPT and similar chatbots are only gradually acquiring memory to recall earlier interactions.
- ▶ Fixes: Instructions, fine-tuning, GPTs.

▶ Difficult to **ascertain the source of the data** underlying the content
- ▶ Fix: Web-connected versions can show sites visited and footnotes for verification.

▶ Without appropriate protections, **potential for exposing trade secrets/confidential information**
- ▶ Fix: Use platforms with terms and protections appropriate to the confidentiality need.

▶ Will produce **biased and discriminatory content** from underlying data sources
- ▶ Fixes: Audit, test, retrain.

▶ **Instability of GPTs**
- ▶ Jon's Fix: Different teams of diverse, dispensable, specialized GPTs at each stage.

4

4

Enterprise Platforms & Confidentiality Risks

Protecting Trade Secrets, Personal Data, Confidential Information & Patentability

Key Principles of Responsible AI

Title search: Responsible AI: Maintaining Your Standards with Emerging Tech

Also available as part of the eCourse
[Responsible AI: Current Regulations (2024) and Maintaining Your Standards](#)

First appeared as part of the conference materials for the
37th Annual Technology Law Conference session
"Responsible AI: Maintaining Your Standards with Emerging Tech"