Data Privacy in AI

PII/PHI in the age of cloud based AI

Presented by: Kevin Perry & Chris Miller

BRIO

Overview

- Popularity of AIaaS
- Rise of the LLM
- Data risks
- Azure PII demo

AI as a Service (AIaaS)

Lots of Offerings

- Image Recognition
- Recommendation Systems
- Predictive Analysis
- Anomaly Detection
- Speech Recognition
- NLP

BRIO

Rise of the LLM

Larger Parameter Set

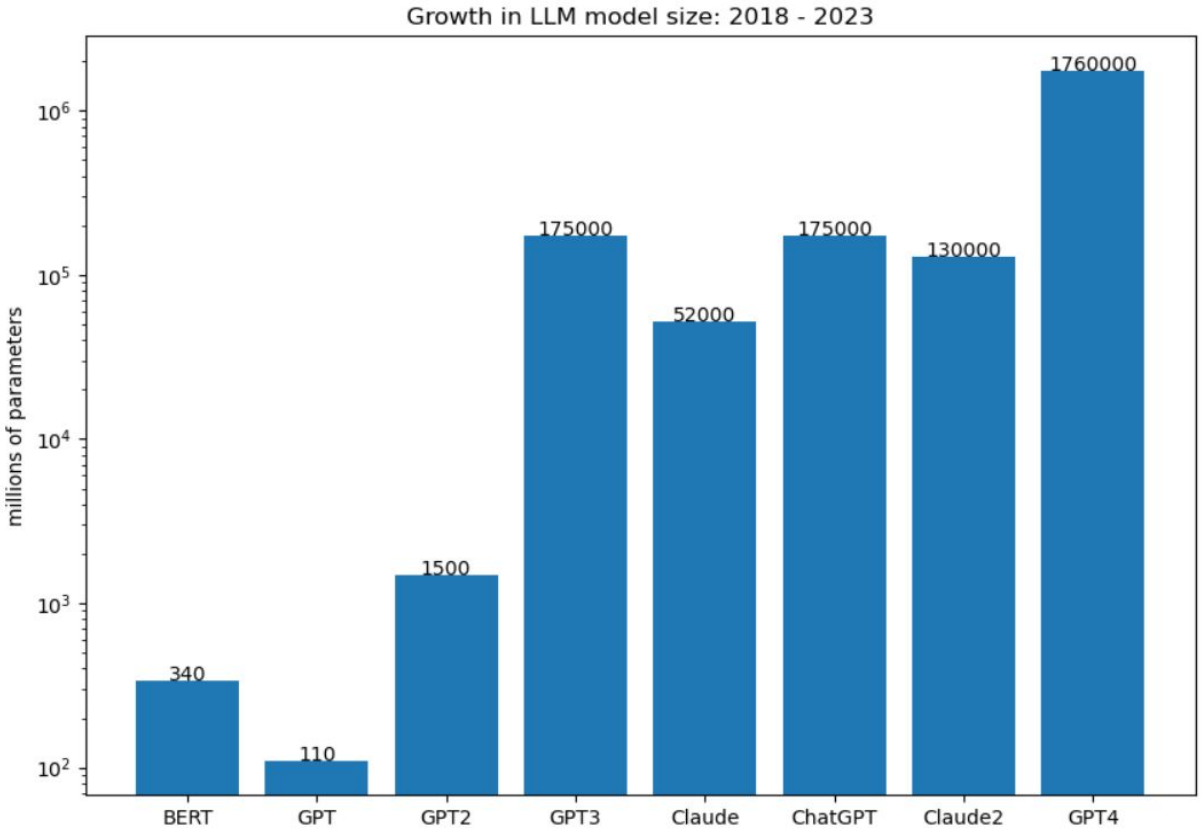
- More powerful models
- Difficult to host on your own
- Specialized hardware and lots of memory
- Training time

Explosive Growth

- 2018 - BERT - 340 million parameters
- 2018 - GPT - 110 million parameters
- 2019 - GPT2 - 1.5 billion parameters
- 2020 - GPT3 - 175 billion parameters
- 2021 - Claude - 52 billion parameters
- 2022 (November) - ChatGPT - 175 billion parameters
- 2023 (July) - Claude2 - 130 billion parameters
- 2023 (July) - GPT4 - 1.76 trillion parameters

BRIO

Logarithmic Growth In LLM Model Size

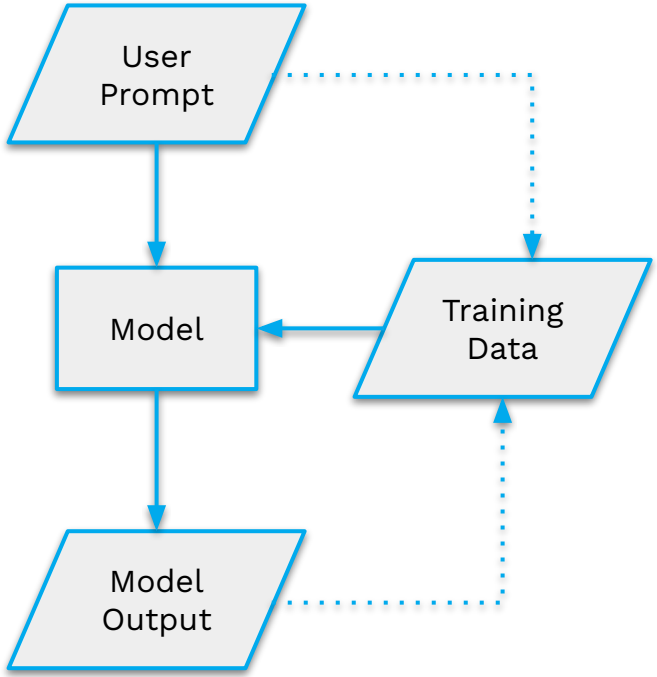


BRIO

Hosted LLMs - What are the risks?

Sensitive data exposure:

- To other users of the LLM through response outputs
- To employees of the 3rd party hosting the LLM
- During a data breach



BRIO

Compliance Concerns

Data privacy laws:

- HIPAA (Health Insurance Portability and Accountability Act)
- GDPR (General Data Protection Regulation)
- CCPA/CPRA (California Consumer Protection Act/California Privacy Rights Act)
- Other regional regulations

BRIO

Know Your Provider

Various services have different considerations

- Shared vs private models
- Is content moderation in place?
- Can it be disabled?
- Compliance measures in place?

Within the Azure domain:

- OpenAI models are private instances
- OpenAI model Inputs are monitored but can be disabled (by request)
- Cognitive services - data retention and monitoring is turned off by default for the PII and health feature endpoints

BRIO

Who’s Naughty? Who’s Nice?

Child behavioral Report (CBT):

```
Timmy Johnson | 7 years | 123 Main St, Pleasant PA 19305 | 345-555-4433
Notes: Timmy's had an ok year. He's doing better in school and mostly
stopped pulling the dog’s tail. However, he's fallen in with a bit of a bad
crowd and started pushing crypto schemes on his classmates.  He convinced
them to invest their lunch money in his new 'TimmyCoin2.0' and then told
them it lost all its value when they tried to cash out. He's also begun
selling NFTs of his drawings, insisting they will only go up in value.
```

Desired output:

```
Timmy Johnson | 7 years | 123 Main St, Pleasant PA 19305| 345-555-4433 |
NAUGHTY
```

BRIO

Data De-identification

- De-identifying data before using it with a generative AI mode can reduce risk.
- De-identified data is generally unregulated by privacy laws
- Methods of De-Identification include deletion, generalization, encryption, data masking, pseudonymization
- Example: GDPR Requirements - Direct identifiers must be removed. Examples are name, address, phone number, image, other unique identifiers

```
XXXXX XXXXXXX | 7 years | XXX XXXX XX, XXXXXXXX XX XXXXX | XXX-XXX-XXXX
Notes: XXXXX's had an ok year….
```

BRIO

Questions?

Thanks for attending!

Source code and presentation
https://github.com/BrioGit/DataPrivacyInAI

BRIO