Illustration by: MidJourney

Dan Fernandez

- Data Nerd
- Product Manager in Cybersecurity
- Instructional Associate @ GaTech

@danielfernandez

/in/dafdz

danielfernandez@infosec.exchange

danielfernandez.medium.com

hey@dafnz.com

Disclaimer

- Opinions are my own and not my employers.
- Suggestions and considerations around data privacy are not legal advice.

Presentation Resources

- The slides will be made available in Github immediately after the presentation.
- Slides contain references to content with additional information.

💡 Resource: Will look like this

Agenda

1. Intro to LLMs
2. Data Governance
3. Data Security & Privacy
4. ML Governance & Security
5. Securing ML Apps

Illustration by: MidJourney

Intro To LLMs

Illustration by: MidJourney

What are Large Language Models?

- A type of artificial intelligence (AI) that can generate text content by predicting word sequences.
- Trained on massive datasets of text and code.
- Becoming increasingly powerful, and are being used in a wide range of applications.

Illustration by: MidJourney

💡Resource: [Why ChatGPT Works - Stephen Wolfram Writings]

Brief History

Photo Credit IEEE Spectrum: Joseph Weizenbaum

- 1950s - 1960s: First rule-based LLMs developed.
- 1980s: Statistical LLMs introduced.
- 2010s: Deep learning LLMs introduced.
- 2020s: LLMs trained on massive datasets and capable of performing a wide range of tasks.

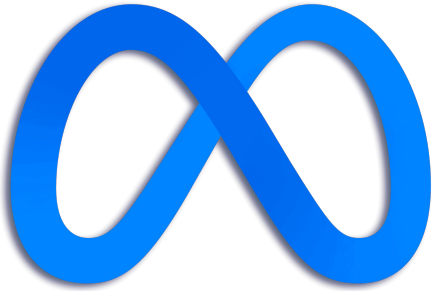
💡 Resource: [Comprehensive Timeline - Voicebot.AI]


LARGE LANGUAGE MODELS
LIGHTYEAR
LARGE LANGUAGE MODELS EVERYWHERE
imgflip.com

Examples of Large Language Models

- GPT
- BERT
- Llama

OpenAI



Meta

Google

💡Resource: [A comprehensive view of Large Language Models]

LLM Use Cases

- **Text generation**: News articles, blog posts, and product descriptions.
- **Question answering**: Natural Language Answers
- **Summarization**: Index and summarize data
- **Code generation**: Software development
- **Translation**: Large Scale Language Translations

💡 Resource: What are large language models used for? Nvidia

Illustration by: MidJourney

Business Examples and Benefits

Illustration by: MidJourney

- **Security**: Query, analyze and understand security data
- **Customer service**: Chatbots that can answer customer questions and provide support.
- **Marketing**: Personalized marketing campaigns and generate content that resonates with customers.
- Finance: Analyze financial data.
- **Healthcare**: Personalize patient care.
- **Education**: Personalized learning experiences and help students learn more effectively.
- **Law**: Analyze legal documents and help lawyers with their research.

💡 Resource: 10 LLM Project Ideas - Towards Data Science

Data Governance

Illustration by: MidJourney

The need for Training LLMs

Why do it?

- ***Maximize Value*** of Company's Data
- ***Improve Accuracy*** and Performance for Custom Use Cases
- ***Enhance Security*** Features to the model

Training Options

- Custom LLM
- Fine Tuned LLM
- Prompt General Purpose LLM

Illustration by: MidJourney

💡Resource: [Using LLMs on Private Data]

The Process

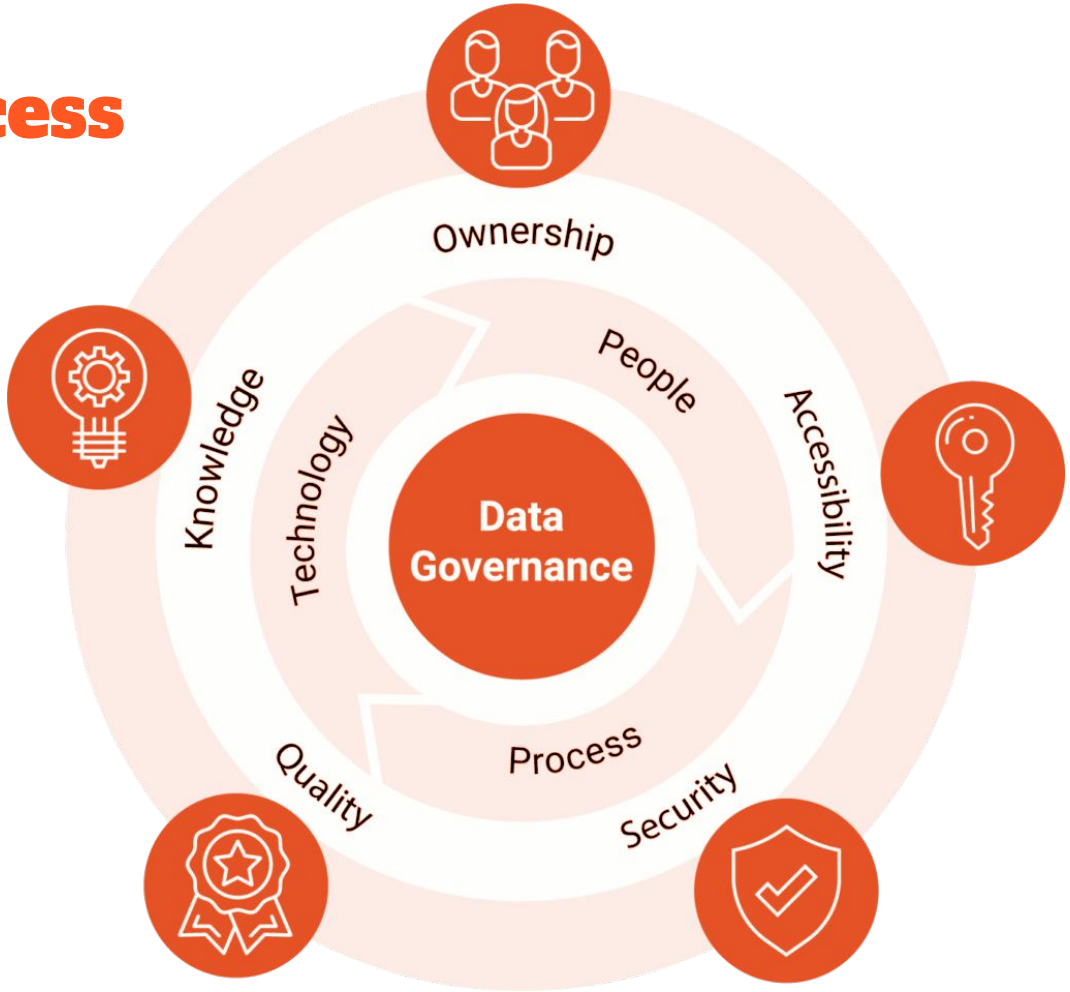


Image Credit: IT Concepts

Data Governance Components

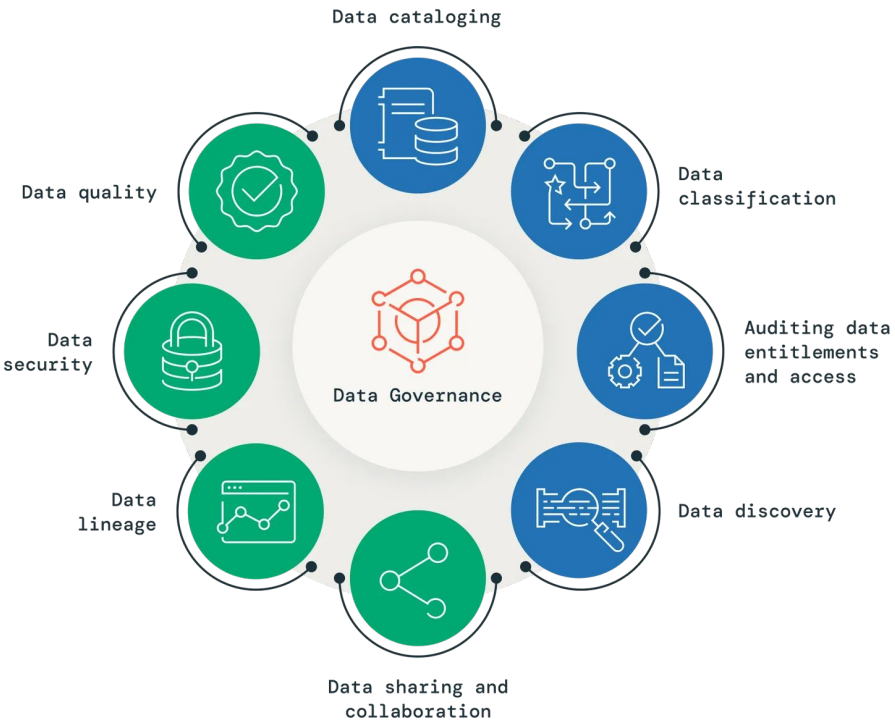


Illustration by: MidJourney & Data Bricks

💡Resource: [Data Governance_ Comprehensive Guide - DataBricks]

Data Governance Challenges

Illustration by: MidJourney

More Data = More Problems

- Complex & Siloed Data
- Ethical Concerns
- Regulatory Compliance

Data Security & Privacy

Illustration by: MidJourney

Data (Store) Security

Fine Tuning

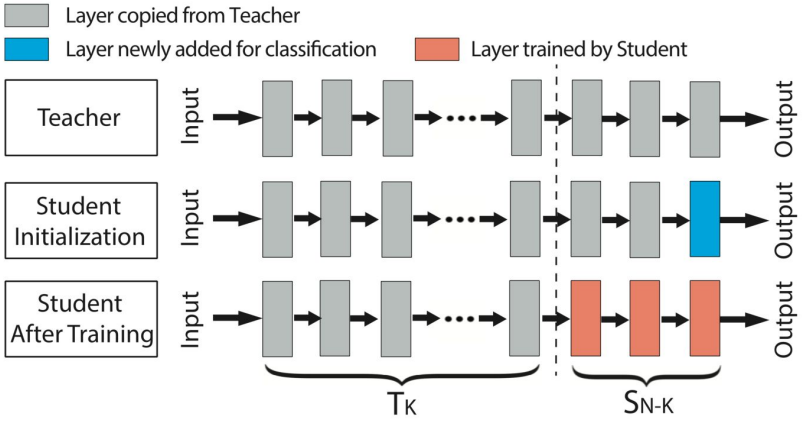


Image Source: University of Chicago

LangChain Style Integrations

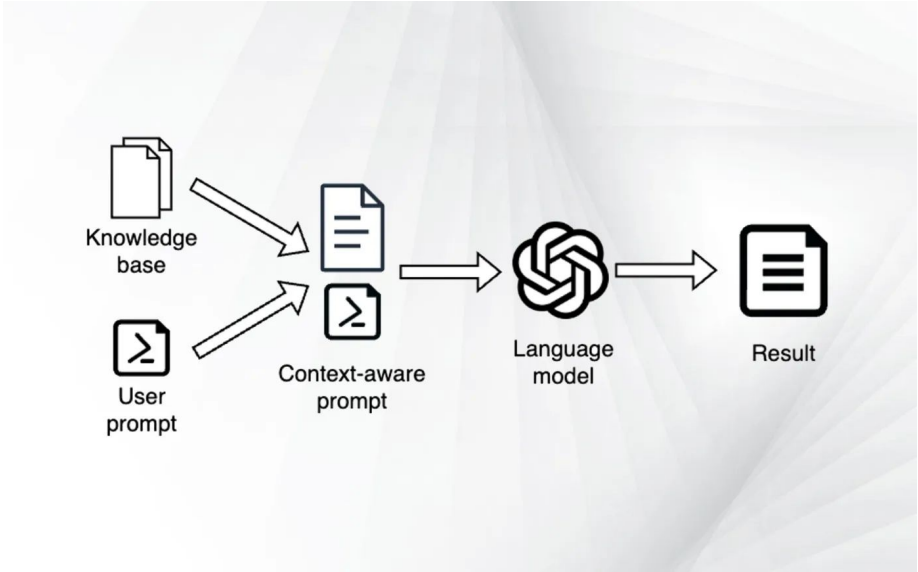


Image Source: BD Tech Talks

Data (Store) Security: Fine Tuning

Areas for Concern

- Access Controls
- Data Poisoning Attacks
- Dataset Tampering

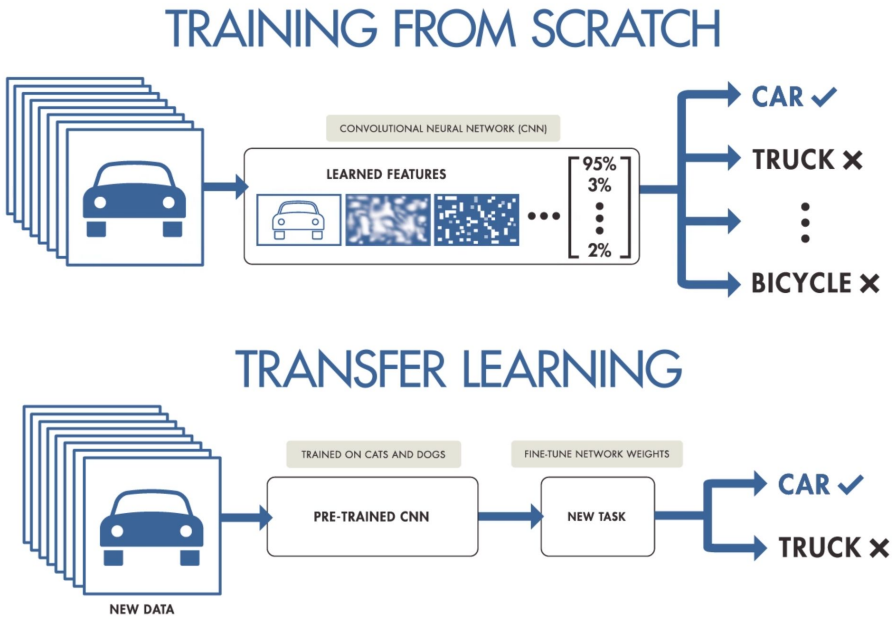
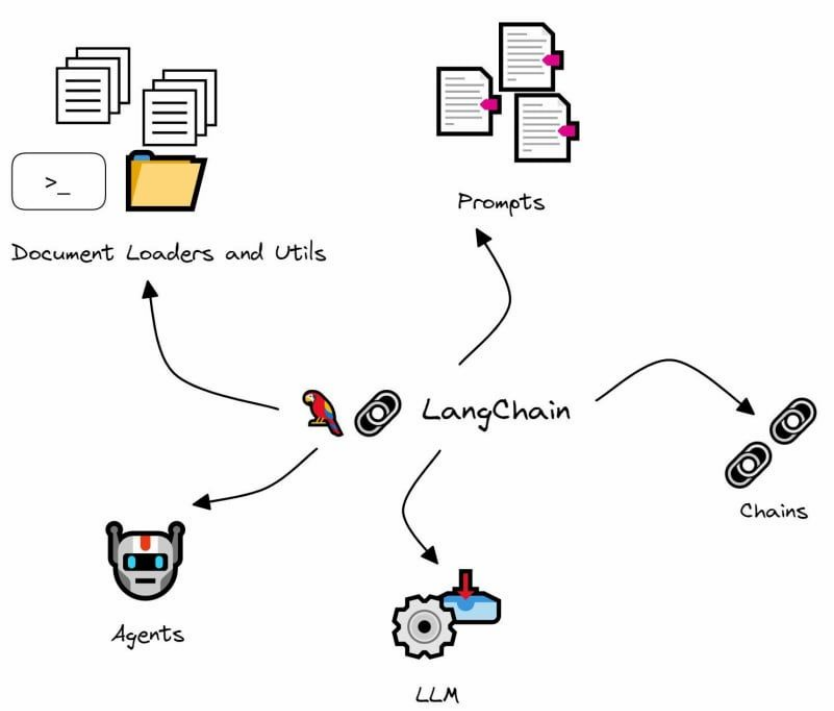


Image Source: Pinterest

💡Resource: Dataset Security for Machine Learning: Data Poisoning, Backdoor Attacks, and Defenses

Data (Store) Security: LangChain Integration



Areas for Concern

- Access Controls
- Database Permissions
- API Security
- Code Vulnerabilities

Image Credit: KD Nuggests

LLM in the News

The Verge / Tech / Reviews / Science / Entertainment / More +

SECURITY / POLICY / TECH

ChatGPT's history bug may have also exposed payment info, says OpenAI

Photo by Amelia Holowaty Krales / The Verge

/ The company has published additional details and a technical breakdown of what lead it to take down the service on Monday.

By Mitchell Clark

Mar 24, 2023, 7:16 PM EDT | 4 Comments / 4 New

If you buy something from a Verge link, Vox Media may earn a commission. ***See our ethics statement.***

Data Privacy & Regulatory Considerations

Considerations

- Sensitive Information Exposure
- Data Access Rights

Recommendations

- Data Anonymization
- Data Minimization

💡 Resource: Privately Fine-Tuning Large Language Models with Differential Privacy

💡 Resource: ProPILE: Probing Privacy Leakage in Large Language Models

Right to Be Forgetten in LLMs

Training of LLMs

- Training Data Memorizations
- Hallucinations

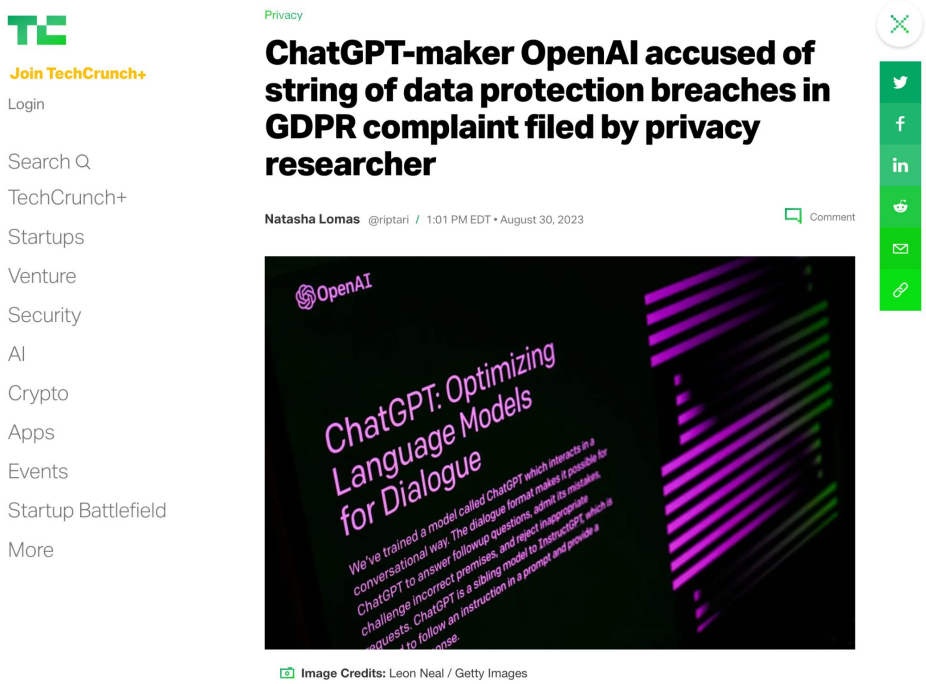
App Usage

- User Chat History
- In-Model Data

💡Resource: [Right to be Forgotten in the Era of Large Language Models: Implications, Challenges, and Solutions]

Illustration by: MidJourney

LLM in the News



💡Resource: [ChatGPT-maker OpenAI accused of string of data protection breaches in GDPR complaint filed by privacy researcher]

ML Governance & Security

Illustration by: MidJourney

ML Pipeline Overview

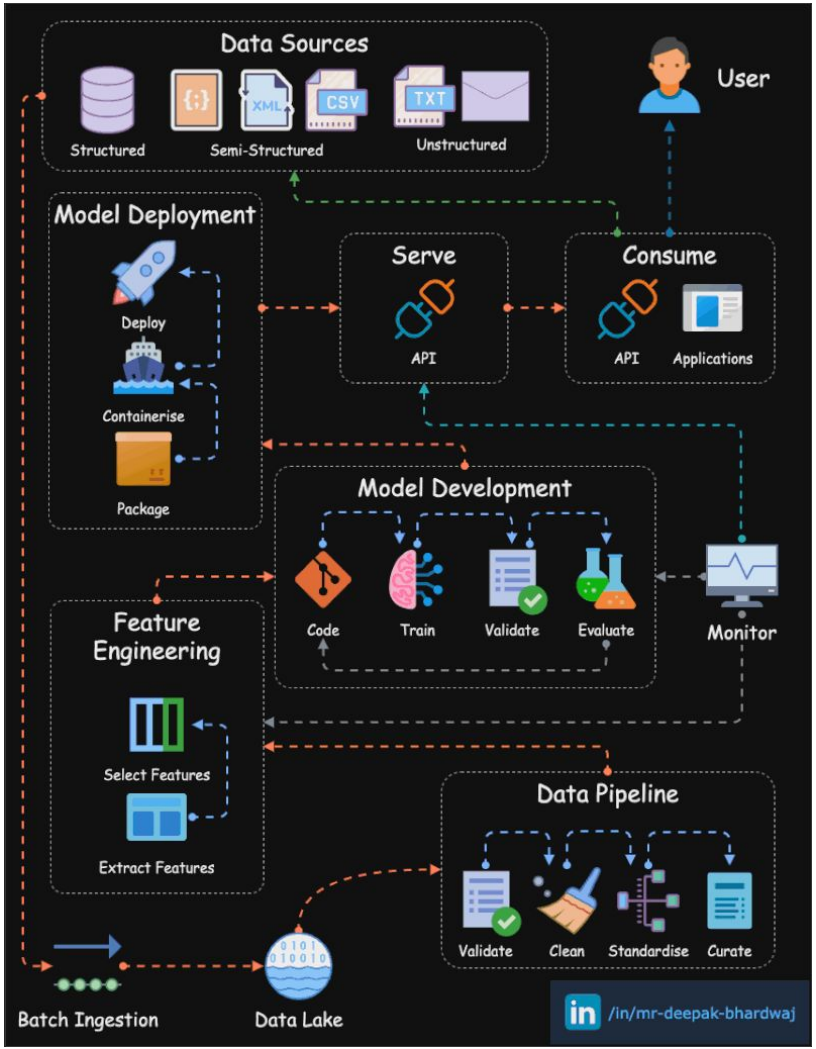


Image Credit: Deepak Bhardwaj

ML Pipeline Overview

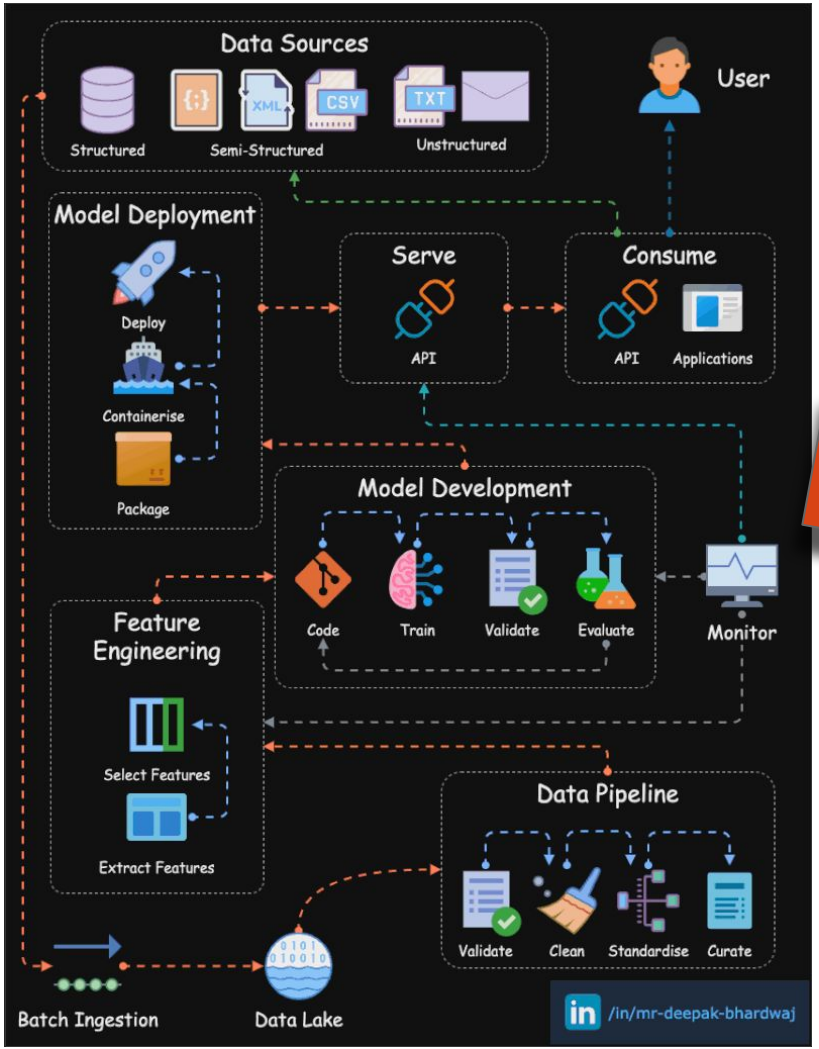


Image Credit: Deepak Bhardwaj

ML Model Governance Basics

Model Quality Control

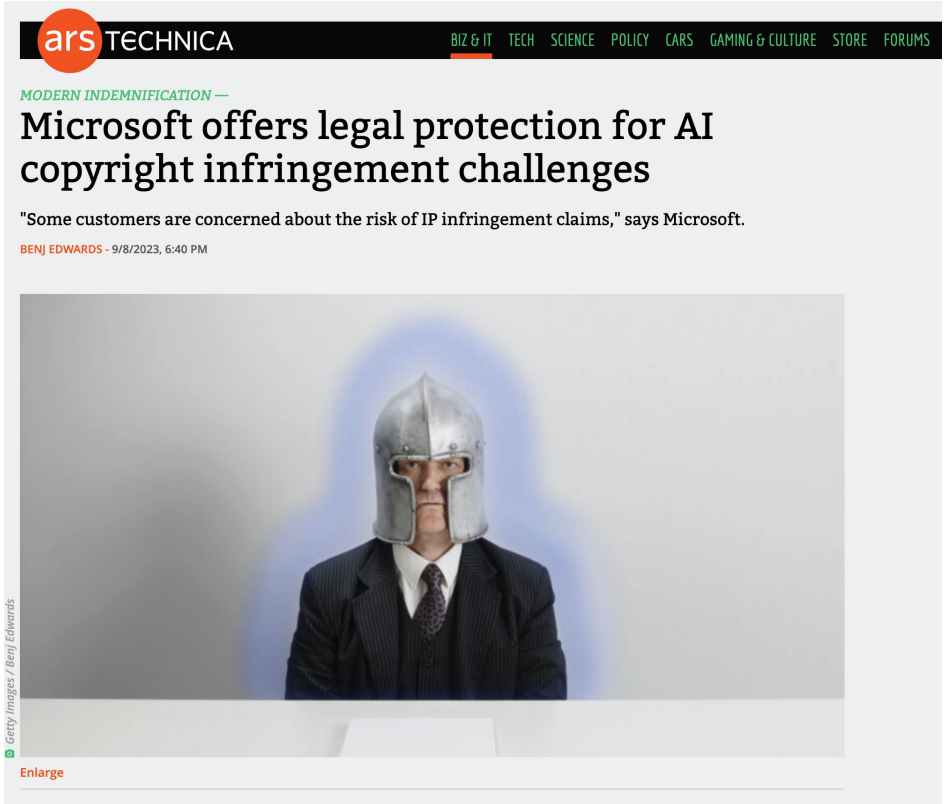
- Measuring Performance Degradation

- Model Update Monitoring & Metrics

💡Resource: Large Language Model Evaluation

Illustration by: MidJourney

LLM in the News



ars TECHNICA

BIZ & IT TECH SCIENCE POLICY CARS GAMING & CULTURE STORE FORUMS

MODERN INDEMNIFICATION —

Microsoft offers legal protection for AI copyright infringement challenges

"Some customers are concerned about the risk of IP infringement claims," says Microsoft.

BENJ EDWARDS - 9/8/2023, 6:40 PM

Getty Images / Benj Edwards

Enlarge

💡Resource: [Microsoft offers legal protection for AI copyright infringement challenges]

ML Pipeline Overview

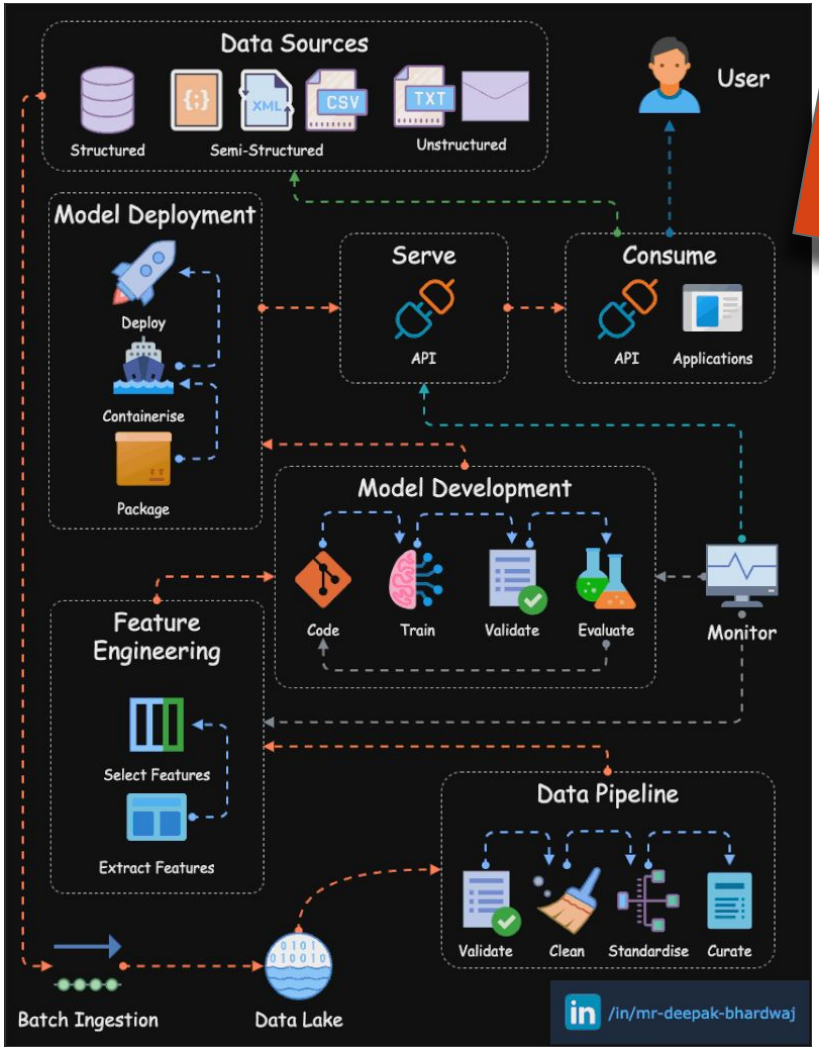


Image Credit: Deepak Bhardwaj

ML Pipeline Security: The Last Mile

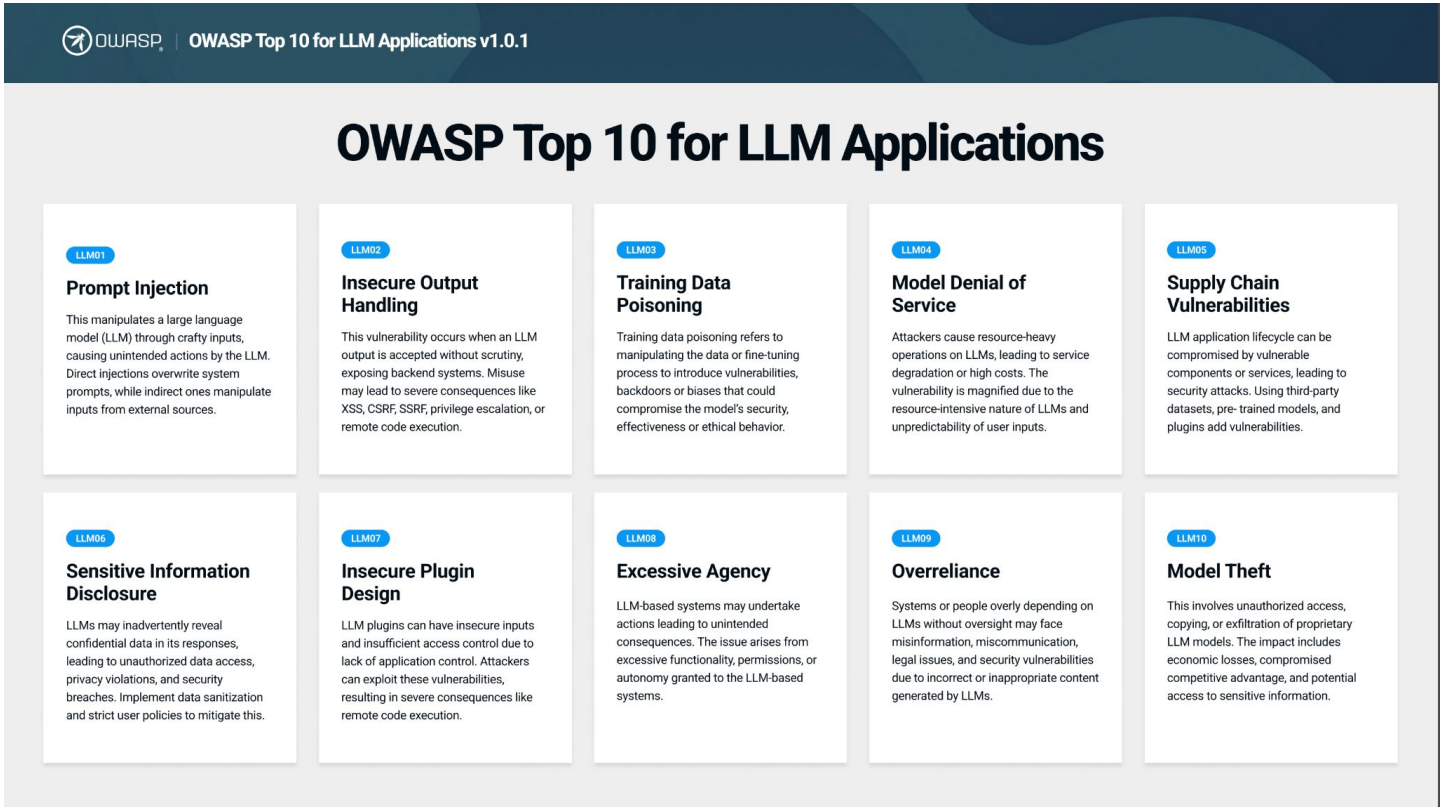
- Prediction Service
  - Code Injection
  - API Vulnerability
  - DDoS
- General Application Security
  - Broken Access Controls
  - Security Misconfiguration
  - Authentication Failures

Illustration by: MidJourney

Securing ML Applications

Illustration by: MidJourney

ML Security Knowledge Bases



💡 Resource: [OWASP Top 10 for LLM Applications]

ML Security Knowledge Bases

Mitre ATLAS

💡 Resource: [MITRE Atlas Matrix]

Parting Thoughts

- Data Security & Privacy
  - Data Strategy & Data Governance are needed
  - Same rules of data store security apply
  - Data Privacy can be more challenging with LLMs

- ML Pipeline Security Areas of Focus
  - Model Development & Deployment
  - Model Governance
  - Model Serving
  - General Application Security

Illustration by: MidJourney

Questions

hey@dafnz.com

@danielfernandez

/in/dafdz

danielfernandez@infosec.exchange

danielfernandez.medium.com

Presentation Slides: https://github.com/dnlfdz/talks