Introduction to Generative AI
Back
Generative AI
11. Generative AI Opportunities: Benefits of Generative AI
12. Case Study: Grab's Use of LLMs for Data Classification
13. Generative AI Opportunities: Industry Opportunities
14. Developing a Perspective on AI and AI Strategy
15. Case Study: How Airbnb Leveraged GenAI to Reshape Customer Support
16. Developing a Perspective on Risks of AI
17. Lesson Review
Generative AI
11. Generative AI Opportunities: Benefits of Generative AI
12. Case Study: Grab's Use of LLMs for Data Classification
13. Generative AI Opportunities: Industry Opportunities
Generative AI
11. Generative AI Opportunities: Benefits of Generative AI
12. Case Study: Grab's Use of LLMs for Data Classification
13. Generative AI Opportunities: Industry Opportunities
14. Developing a Perspective on AI and AI Strategy
15. Case Study: How Airbnb Leveraged GenAI to Reshape Customer Support
16. Developing a Perspective on Risks of AI
17. Lesson Review

Case Study: Grab's Use of LLMs for Data Classification

Grab's Use of LLMs for Data Classification

Grab, Southeast Asia's leading "superapp", positions itself as a user's "everything app"; it provides everyday services such as delivery, mobility, financial services, and much more. Grab's data operations are substantially complex, spanning these many areas of user service. Grab uses Large Language Models (LLMs) to automate their data classification at scale which provides tremendous efficiency impacts to their data operations. (While this is a more narrow, more technical use case than other case studies here, it nevertheless offers some lessons for those companies that plan to implement LLMs.)

Grab uses GenAI to address three primary business challenges:

- **Manual Tagging Bottleneck**: Grab historically relied on manual data classification, requiring significant time investment and inconsistencies. LLMs offered a solution that eliminated manual tagging and enabled adaptation to custom requirements.
- **Data Volume and Scalability**: Manual classification proved progressively impractical as data volume surged. The LLM-powered system scaled to handle large datasets, improving both efficiency and accuracy.
- **Inefficient Schema-Level Classification**: Classifying entire schemas based on single sensitive tables resulted in overly restrictive access control and hindered data accessibility.
- **Inconsistency and Bias**: Varying interpretations of data governance policies by app developers led to inconsistencies and potential bias. LLMs, guided by clear prompts and user feedback, mitigated such issues to provide more consistent classification.

Strategic LLM Application:

- **Column-Level Tagging**: Grab utilizes LLMs to classify individual data fields within schemas, achieving granular data governance and enabling personalized access control based on sensitivity tiers.
- **Prompt Engineering**: To ensure accurate LLM outputs, Grab employs techniques like few-shot learning and schema enforcement, guiding the model with relevant examples and explicit formatting guidelines.
- **Orchestration Service (Gemini)**: Grab built Gemini, an internal service (named before Google gave its GenAI suite the same name), to manage data classification requests, schedule jobs, and leverage multiple classification engines, including GPT-3.5, concurrently.

Benefits and Implications:

- **Significant Time Savings**: Automating classification through LLMs reduces manual effort, saving Grab an estimated 360 person-days per year. This frees up resources for core tasks and improves overall data management efficiency.
- **Enhanced Data Governance**: Accurate data classification allows for granular access controls and dynamic data masking, improving data security and privacy. This fosters trust and compliance with data regulations.
- **Unlocking Downstream Applications**: Classified data tags pave the way for various applications in security, data discovery, and analytics. This enables further data insights and optimization.

Implications for Other Organizations:

- **LLMs as Automation Potential**: Grab's success demonstrates the potential of LLMs to automate complex data governance tasks, improving efficiency and accuracy. Organizations can explore similar approaches to streamline data management.
- **Prioritizing Prompt Engineering**: Carefully crafted prompts are crucial for guiding LLMs and ensuring desired outputs. Organizations should invest in expertise and techniques for effective prompt engineering.
- **Continuous Improvement and User Feedback**: Iterative improvement through user feedback and data analysis is crucial for optimizing LLM performance and maximizing its benefits.

Grab's strategic and innovative use of LLMs for data classification offers a compelling case study for other organizations seeking to automate data governance, improve efficiency, and unlock further data insights. I especially appreciate Grab's focus on prioritizing prompt engineering; I can't emphasize enough the role of clever prompt writing to not only wrangle an LLM into giving desired responses, but doing so at a token cost that does not break your budget.