



Massachusetts
Institute of
Technology

Slide Hustle

Creating a Search Engine for Slides



Michel Leroy



Maxime Basse

McKinsey&Co Representatives: Suzana Iacob, Omer Keinan, Rodrigo Olivares, Julia Cheung
Faculty Advisor: Alexandre Jacquillat

Problem Statement

Context

McKinsey is a consulting firm that provides clients with expertise and valuable insights to tackle their challenges. The firm's consultants leverage an extensive internal slide library to learn from past studies and design presentations grounded in existing data and analysis.

Problem

There's no search tool that indexes both slide content (text, data...) and visuals (charts, diagram...), forcing manual, time-consuming searches.



A consultant making a slide needs inspiration



They query for the slide they want



The tool returns a panel of relevant slides

Objective

- **Integrate** a new slide search engine to McKinsey's PowerPoint plugin
- **Retrieve** 20 most relevant slides based on a content or layout query, in less than two seconds

Datasets



PowerPoint Slides
Diverse slides featuring charts, images and consulting layouts



User queries
collected through workshops and synthetic data generation

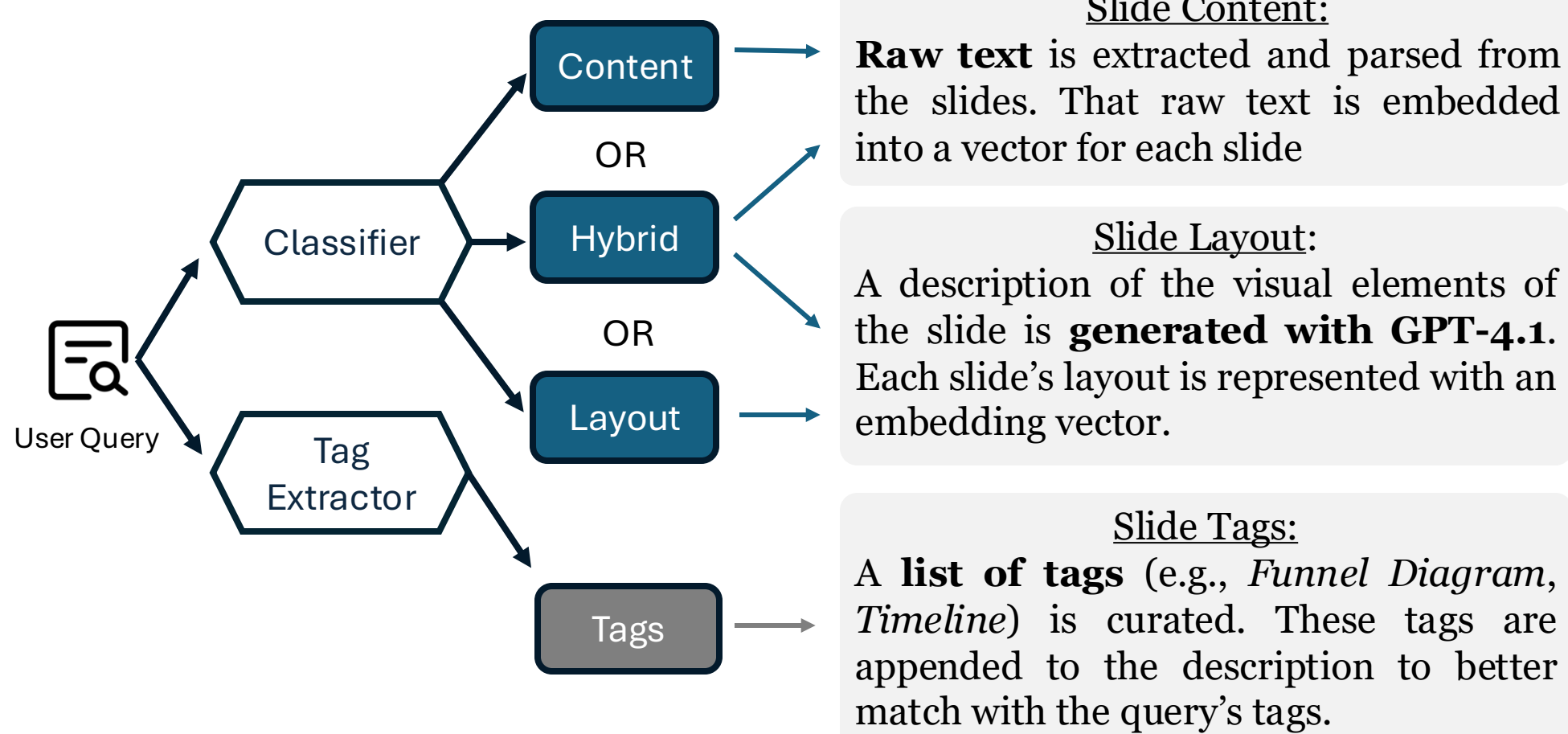
Methodology

User Query Processing

Slide Processing

Benchmarking System

Full Processing Pipeline

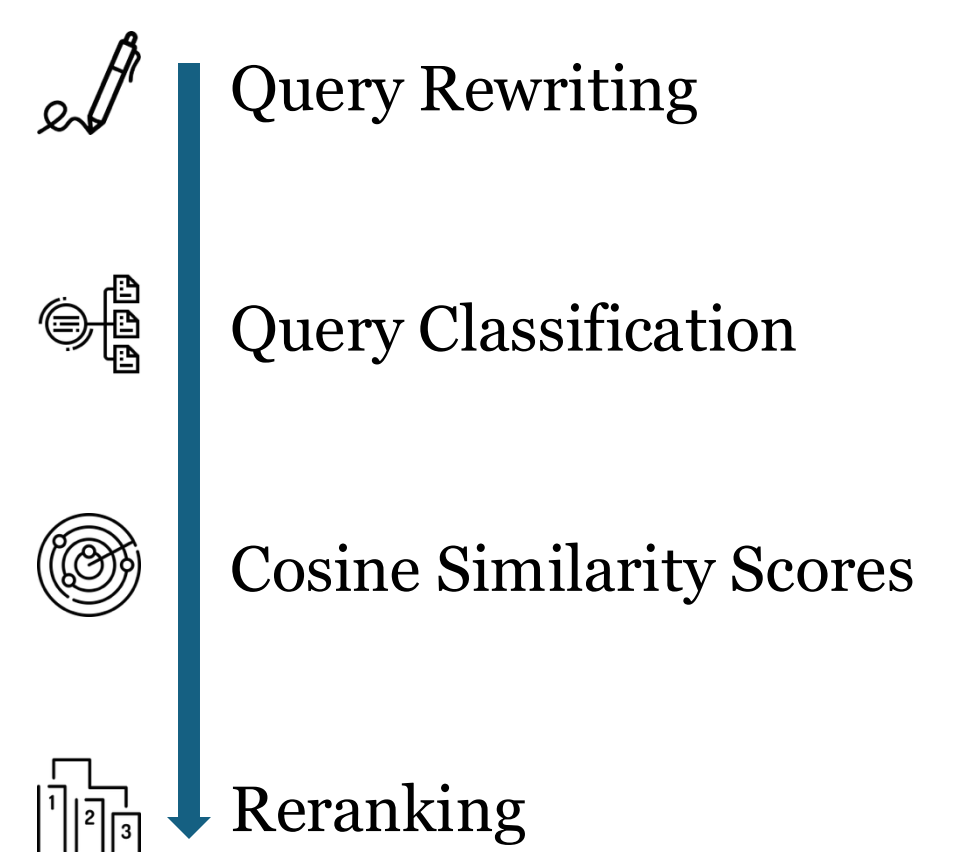


4,000 synthetic queries (each generated from a target slide) are tested against different versions of the pipeline.

Parameters tuned with the benchmark:

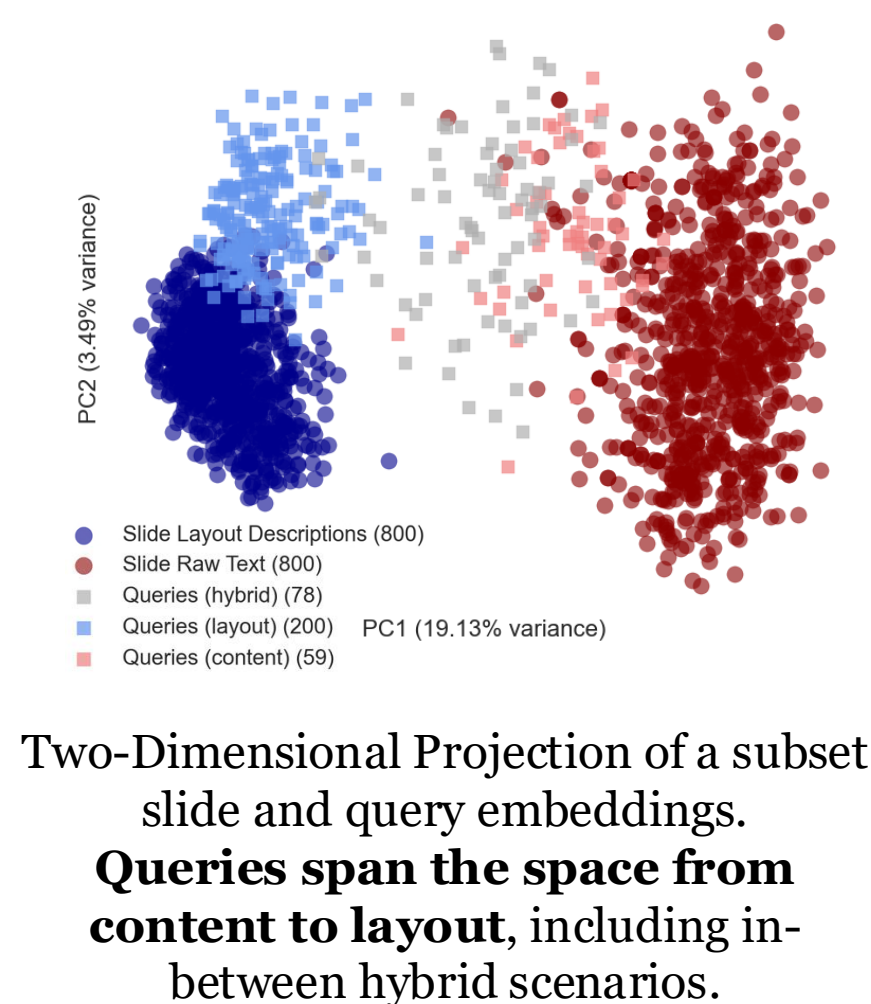
- Rewriting Prompt
- Description generator prompt
- Classifier prompt
- LLM used for each API call
- Reranker model
- Weights adjustments

Benchmarking Metrics		
Recall@K "Out of the relevant slides, how many are retrieved in the top K?"	Precision@K "What proportion of the top K results are actually relevant?"	NDCG@K "Considers the position of relevant items in the top K to derive a score."



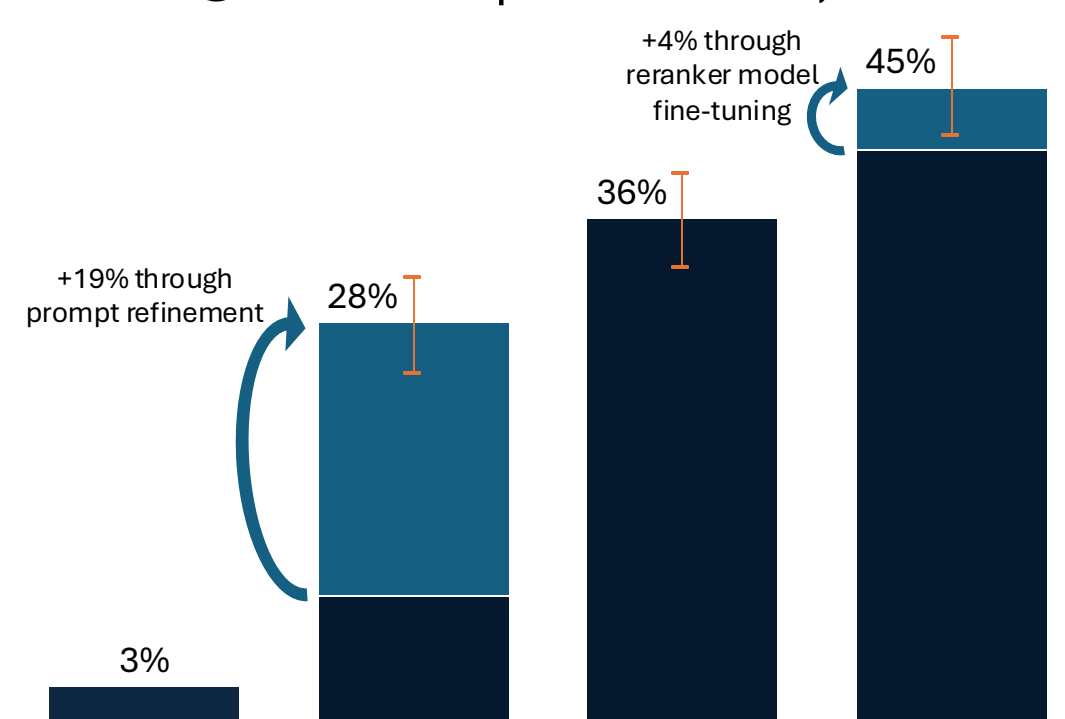
Results

Embeddings Analysis



Recall performance

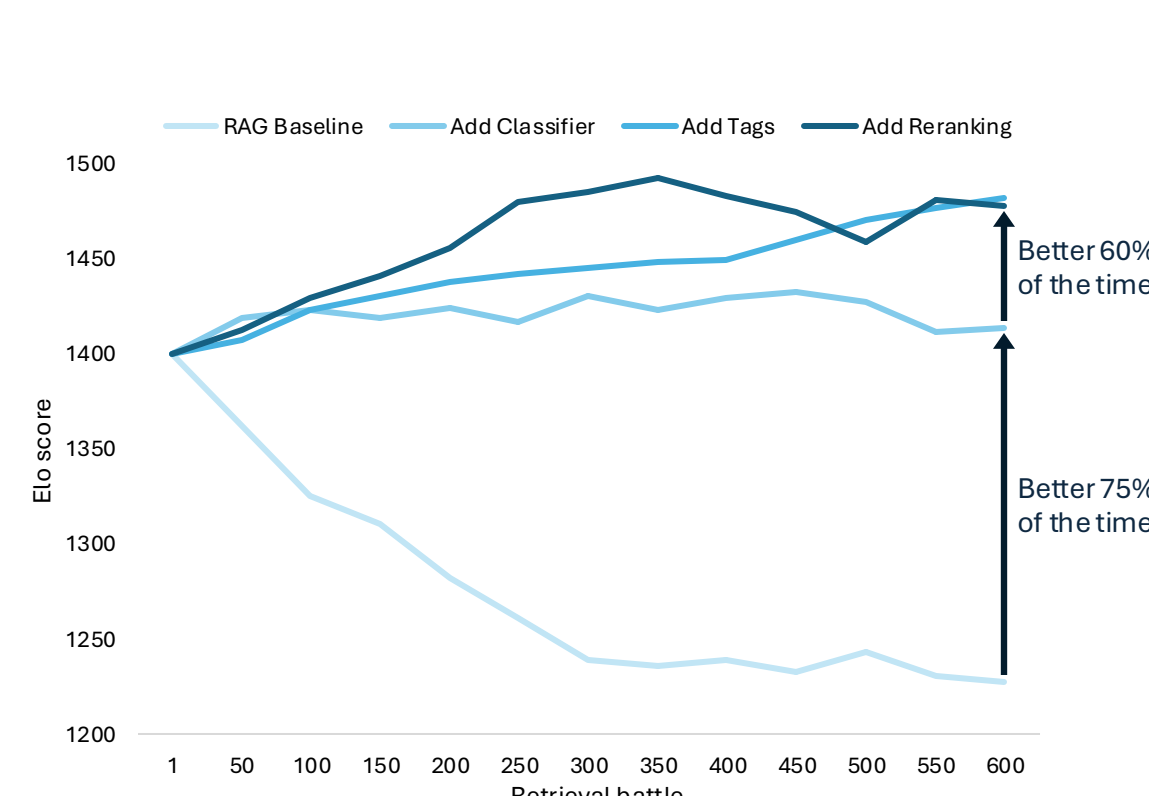
Recall@20 for 800 queries and 10,000 slides



We compared the performance of several pipelines on queries that target the **visual components** of a slide

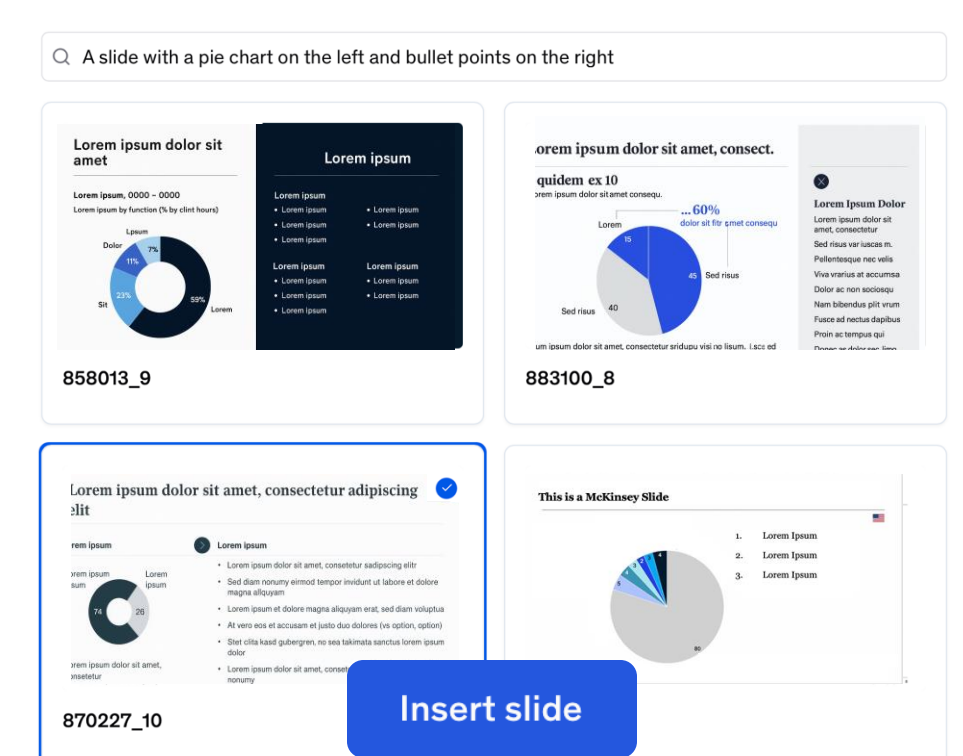
Precision Performance

ELO score evolution between 4 models



We **validated** those results with a precision benchmark system consisting of a series of duels

Plugin Embedded Output UI



The user accesses the search engine directly in a tab **within PowerPoint**. Slides can be inserted at the current position in the deck.

Business Impact

- ✓ Higher quality slides retrieved in less than 2 seconds
- 🚀 Currently deployed for test audience with roadmap in place to deploy firmwide
- 🕒 ~325k yearly consultant hours saved via the PowerPoint Plugin (already 15k users)

Future Work

- ❑ Enable continuous ingestion of new documents into the retrieval pool via integration with McKinsey's knowledge base
- ❑ Tags adjustments based on consultant experience feedback and tool monitoring