McKinsey & Company | OPERATIONS RESEARCH CENTER | MIT MANAGEMENT SLOAN SCHOOL

One-Click-to-Publish

Automating Knowledge Curation with GenAI

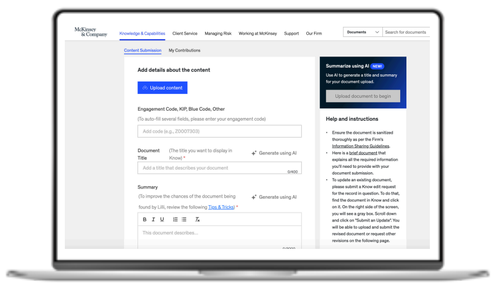
McKinsey Team: Suzana Iacob, Neha Mendiratta
MIT Advisor: Professor Chara Podimata

Samantha Tsang

Vojta Machytka

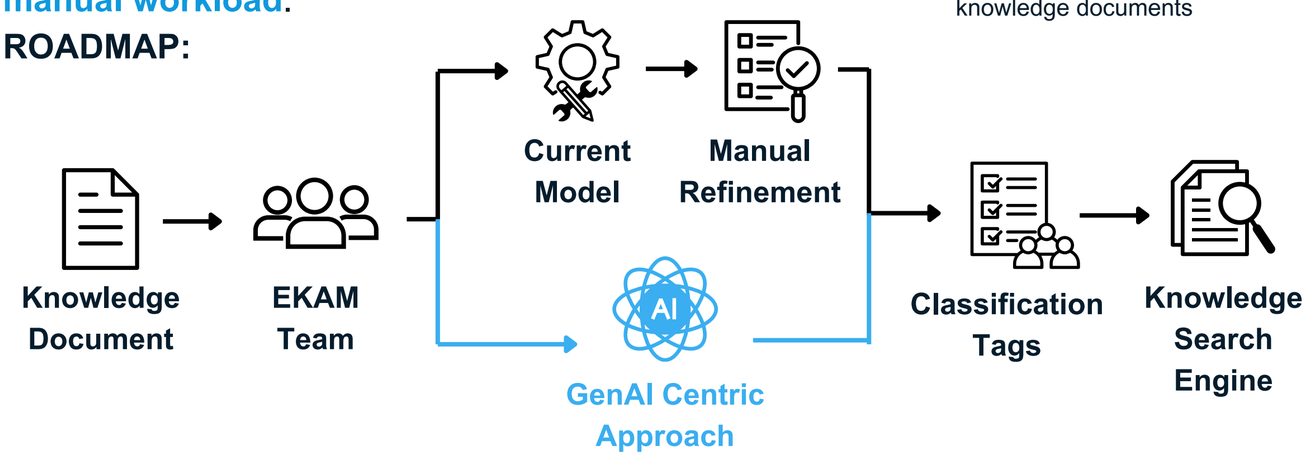
Problem Statement

CONTEXT: McKinsey faces a challenge of manually curating and tagging documents in its internal knowledge repository, with the current model operating at ~50% accuracy. A **GenAI-centric** classification process can **streamline the tagging process**, **enhance searchability**, and **reduce the manual workload**.



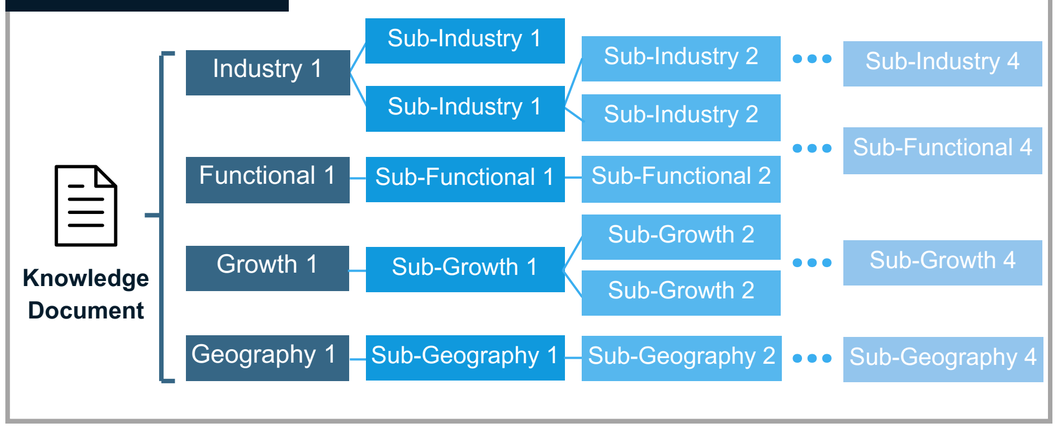
PD Submission Page for new knowledge documents

ROADMAP:



Objective

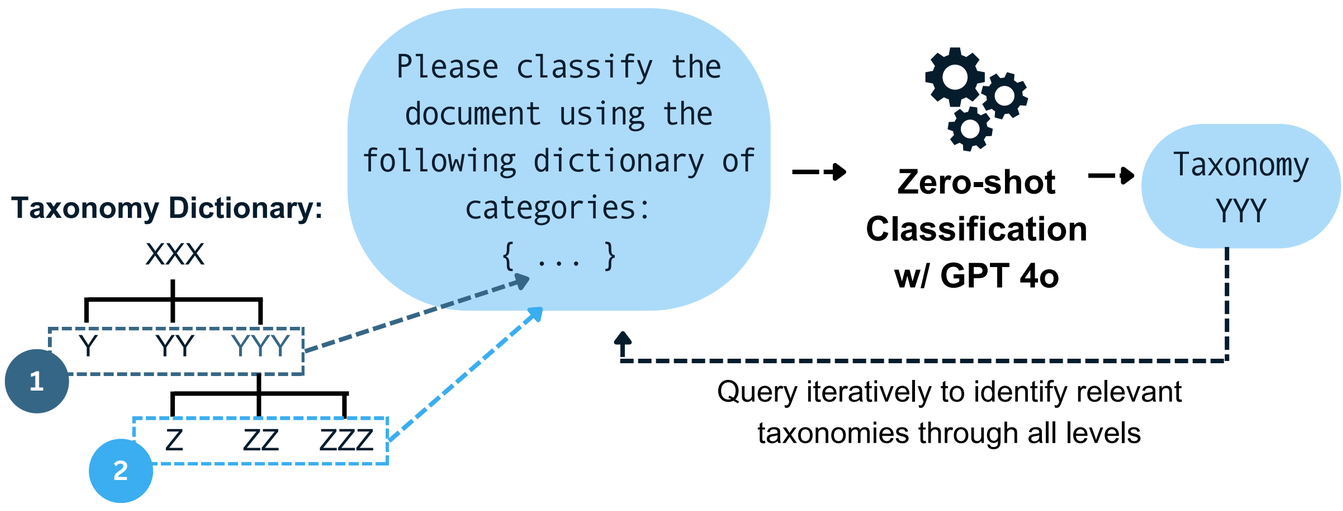
Dataset



Methodology

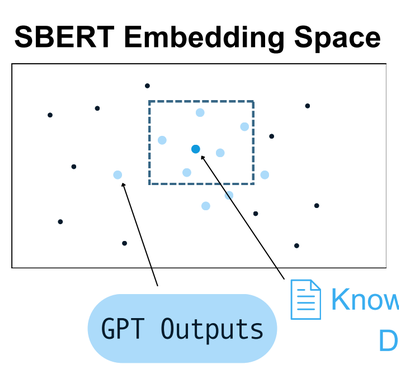
HIERARCHICAL CLASSIFICATION

Learning from context: We utilize **zero-shot capabilities** of GPT to classify documents in lieu of ground truth data, tagging documents to taxonomies **as granular as possible**



RELEVANCE FILTERING

More is not always better: We filter GPT's output using **cosine similarity**, providing only the **most relevant** results for users



Calculate cosine similarity between **GPT outputs** and each **document chunk** - using maximum for comparison

Final model output:

Top 5 taxonomies with highest cosine similarity per scheme

Results

(Metrics obtained from sample of 150 manually labeled documents)

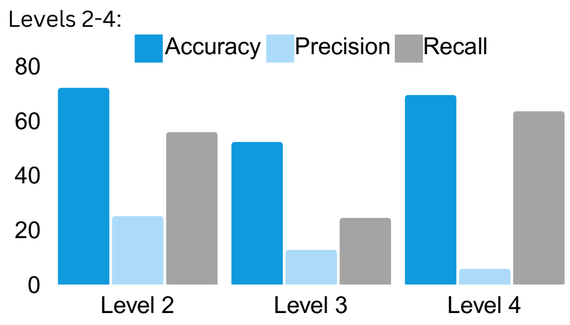
MODEL PERFORMANCE

Level 1:

79.8% documents were labeled with at least 1 correct taxonomy

73.4% ground truth labels were accurately identified

Levels 2-4:



RESULT CONSISTENCY

Consistant Response Rate: **46%**

On average, when running the model 5 times on the same document, **all classified taxonomies will be the same** in half the runs

Hallucination Rate: **1.4%**

Further analysis showed **high similarity** between hallucinations and model outputs

TIME AND COST SCALABILITY

3.6 seconds average per document

$ 0.35 average per document

> The new model for the auto classifier holds **great promise** for the EKAM team and for **McKinsey's search results as a whole** [...] it would spare the EKAM team **much time and effort** that currently goes into researching which terms to add for each document.
>
> - Kimberly Perman
> (Senior Manager of Knowledge Operations EKAM)

Business Impact

MODEL PRODUCTIONALIZED FIRM-WIDE

Implementation in **knowledge search engine** for **firm-wide** usage for: **180K** **document searches annually**

Estimated to label: **26K** **documents annually**

EXPANDING MODEL USABILITY TO LILLI

Our model's output will be used as **source of ground truth** for training new models to classify and enhance metadata of other documents in Lilli, **boosting search algorithm** for: **200K** **documents** **140K** **user queries weekly**

Future Work

Improve Parsed Text Quality

for better logo-text extraction and positional understanding

Incoporate Reasoning in GPT

to help prompt engineering or additional result filtering

Back Labeling for Past Submissions

to standardize tags across all documents in platform

> **[[placeholder for testimonial 2]]**