

One-Click-to-Publish

Automating Knowledge Curation with GenAI

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



Samantha
Tsang

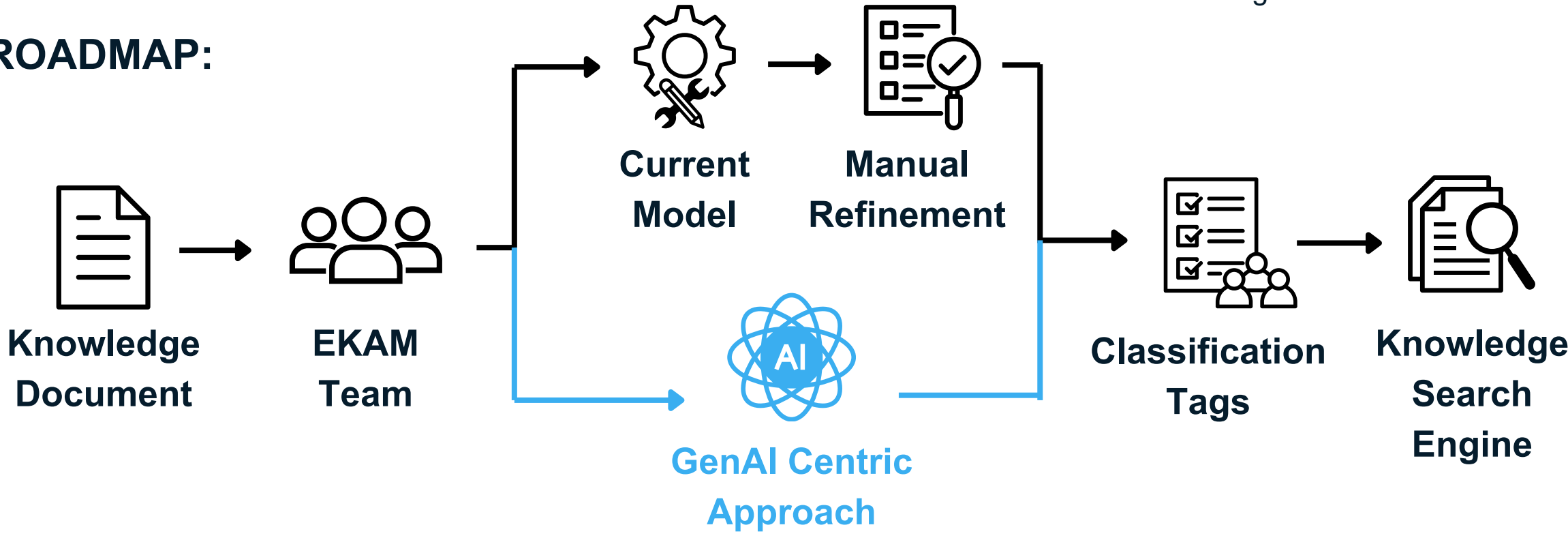


Vojta
Machytka

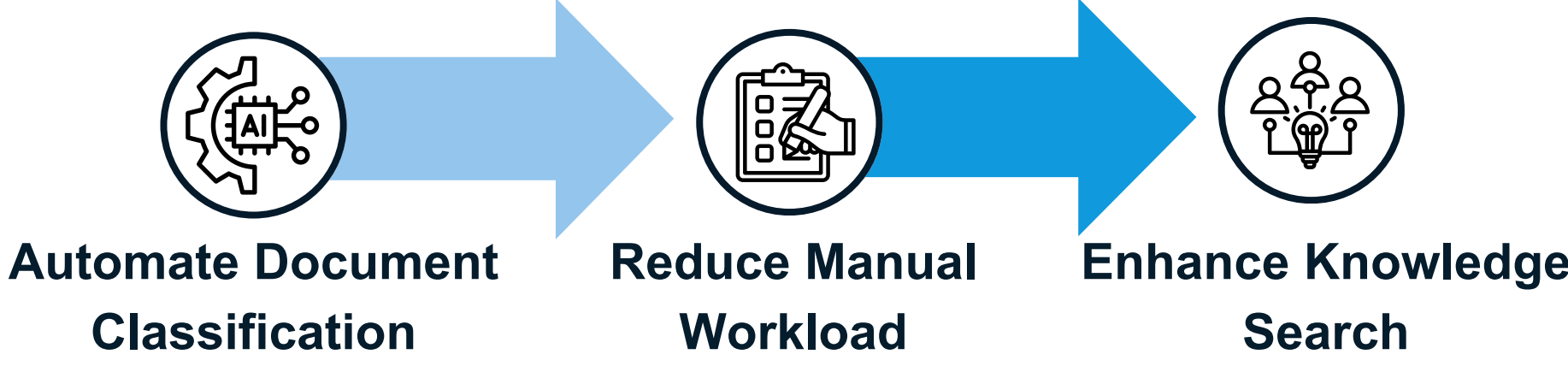
Problem Statement

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**.

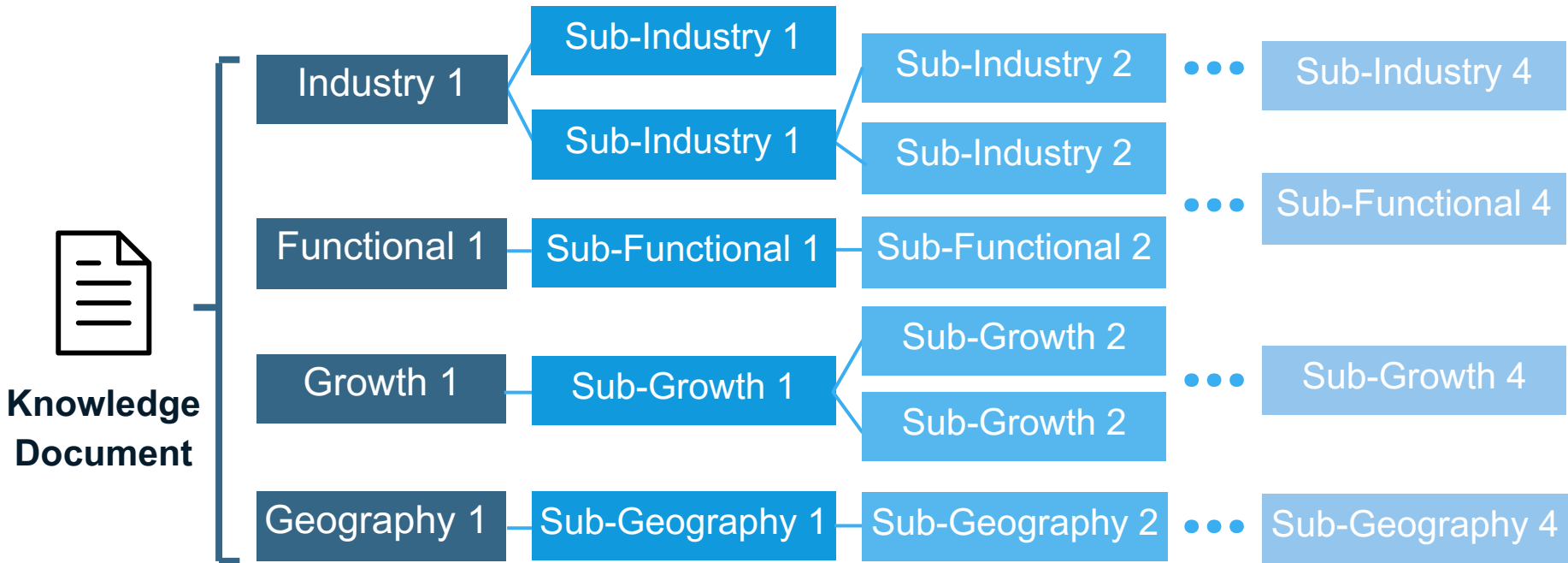
ROADMAP:



Objective



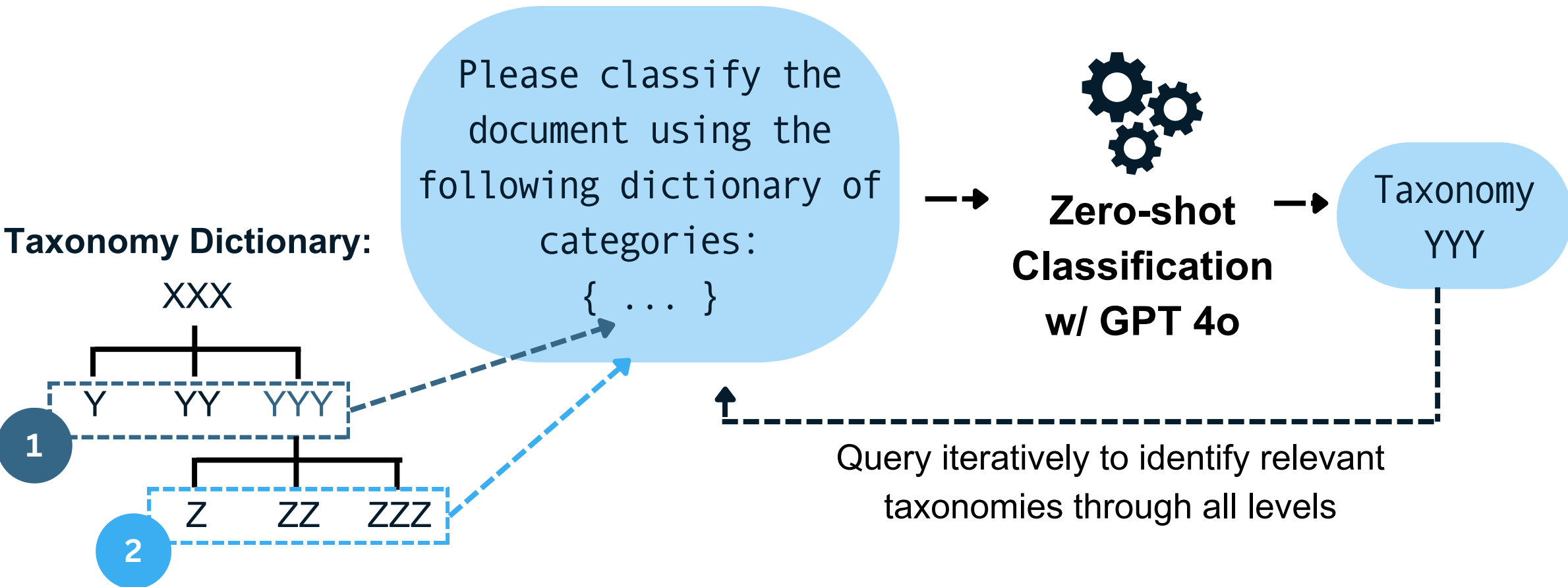
Dataset



Methodology

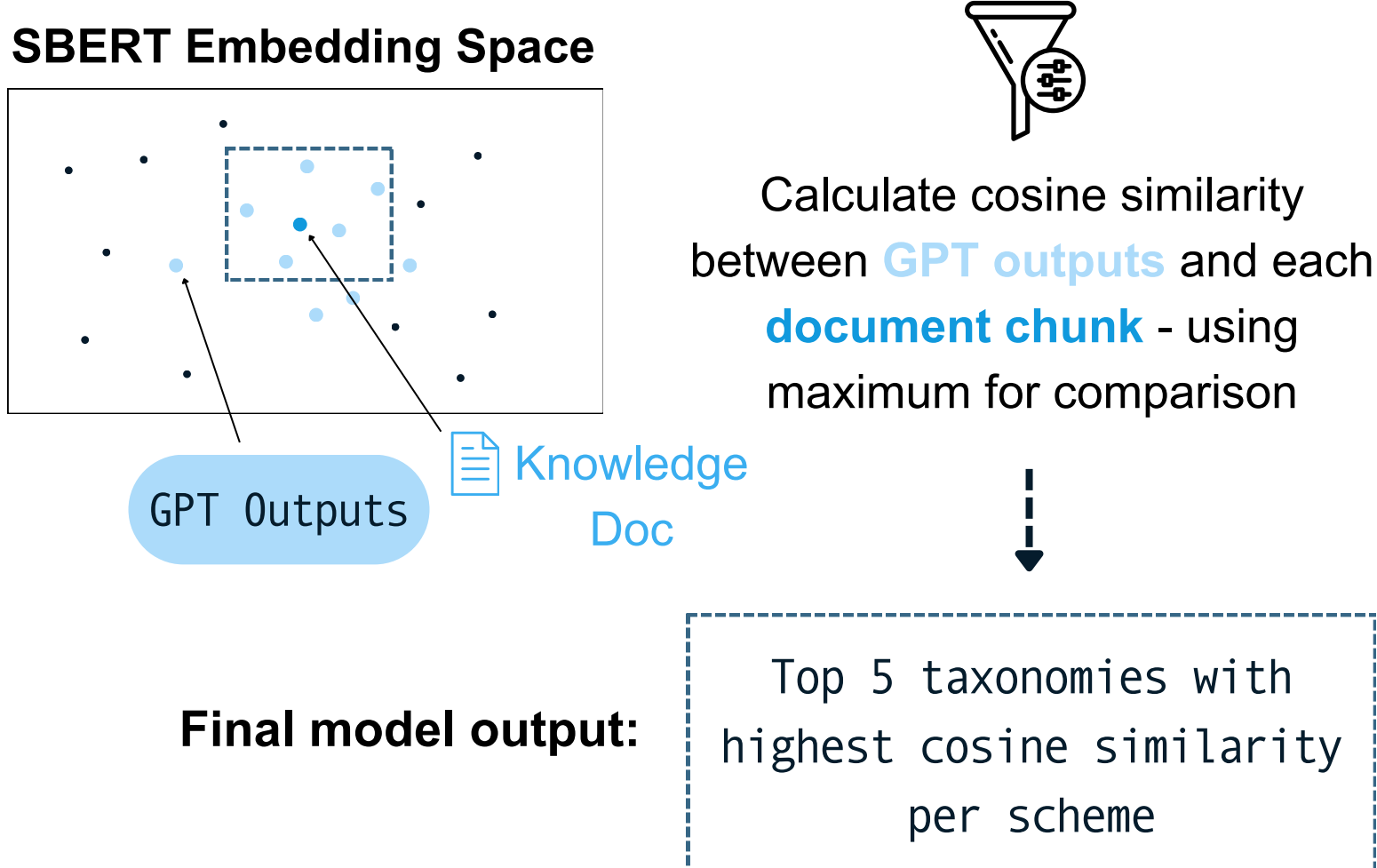
HIERARCHICAL CLASSIFICATION

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

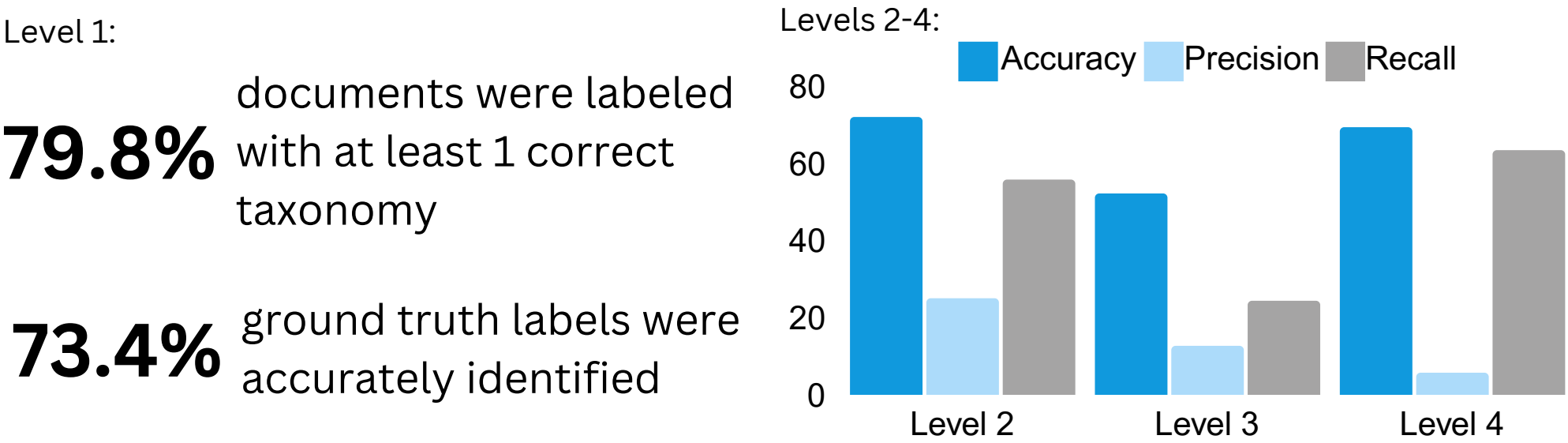
More is not always better: We filter GPT's output using **cosine similarity**, providing only the **most relevant** results for users



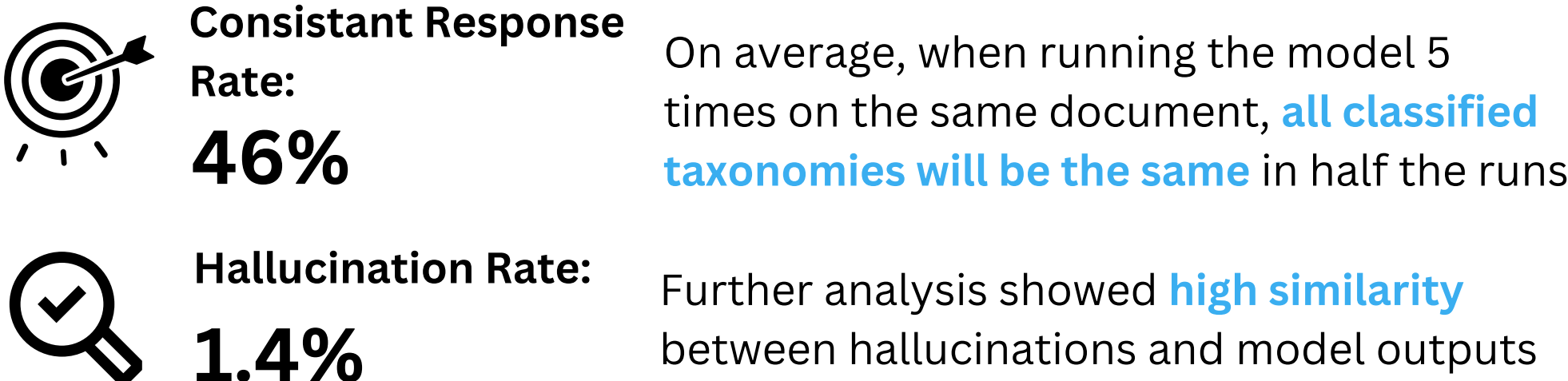
Results

(Metrics obtained from sample of 150 manually labeled documents)

MODEL PERFORMANCE



RESULT CONSISTENCY



TIME AND COST SCALABILITY



Business Impact

MODEL PRODUCTIONALIZED FIRM-WIDE

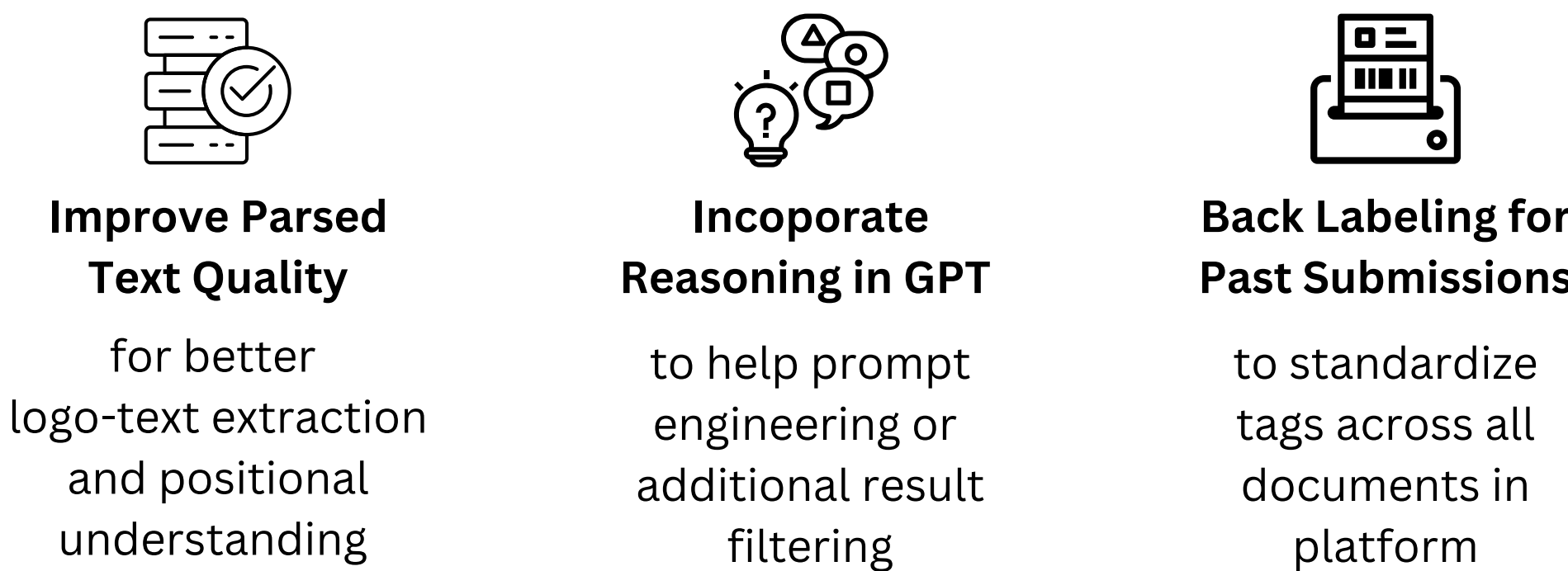
Implementation in **knowledge search engine** for **firm-wide** usage.

Estimated to label: **26K documents annually** Saving up to: **676 hours per analyst 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



“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)

“Collaborating with the exceptional MIT team has been a pleasure [...] they have been able to **address a long-standing problem** of knowledge curation at McKinsey and **reduced the time from 20s to 3s** for each document [...] this MIT capability will allow us to **maintain our preeminent position** in the consulting industry

- Suraj Sharma
(Director, Digital Client Capabilities)