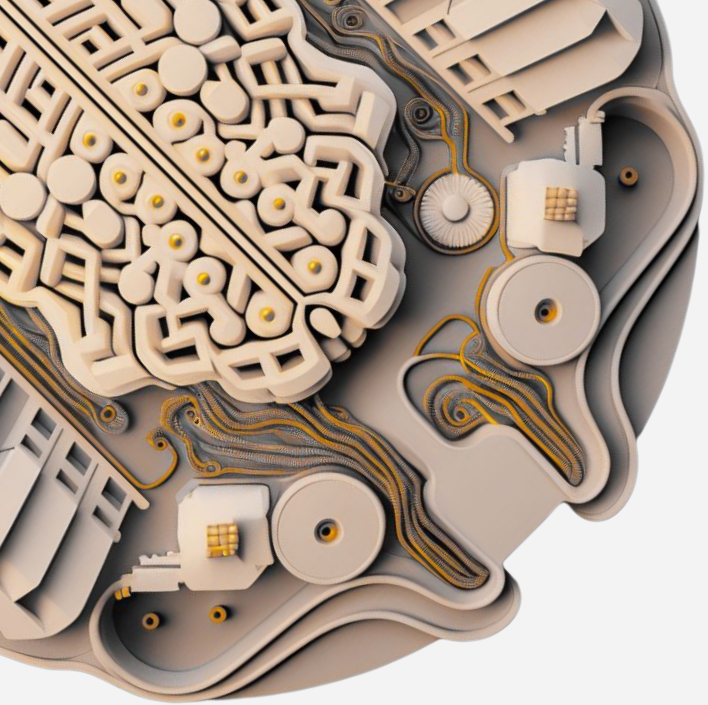


PTEC: Prompt Tuned Embedding Classification for Industry Sector Allocation

EQT MOTHERBRAIN

VU UNIVERSITY AMSTERDAM

UNIVERSITY OF AMSTERDAM

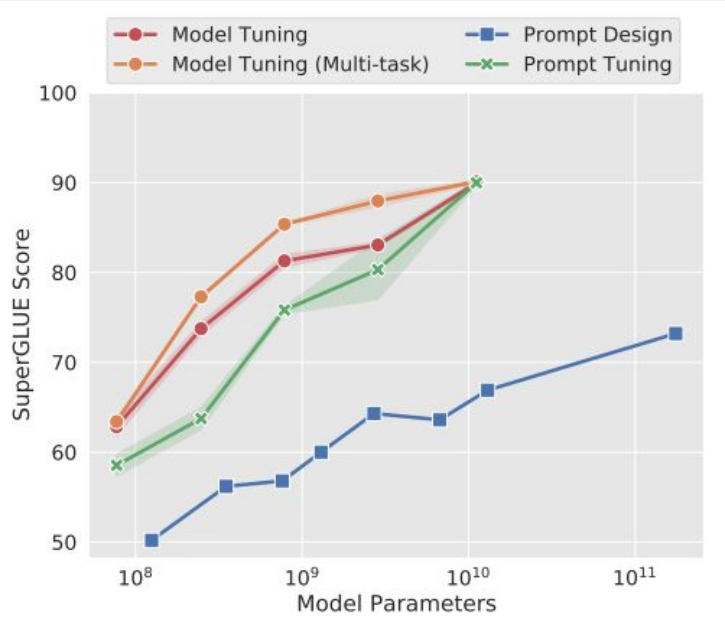
Valentin Buchner
EQT Motherbrain & VU Amsterdam

Lele Cao
EQT Motherbarin

Jan-Christoph Kalo
University of Amsterdam & VU Amsterdam

Vilhelm von Ehrenheim
EQT Motherbarin

Parameter Efficient Fine Tuning (PEFT)



Lester et al., 2021

🌱 Cheaper

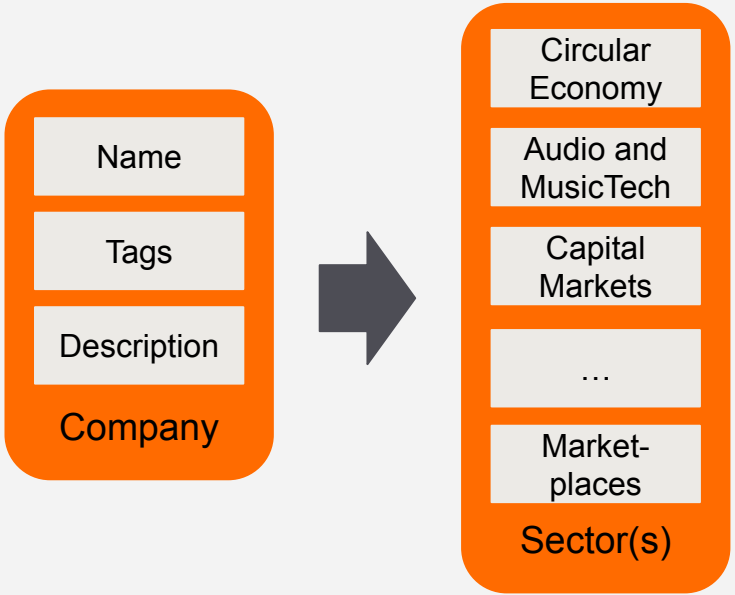
- Less computation during training
- Only fine-tuned parameters need additional storage

🤓 No catastrophic forgetting

- Maintains performance on OOD data

Thematic Investment

- Identifying promising macro-trends
  - E.g. renewable energy, circular economy
- Finding investments within these macro-trends → Multi-label text classification task



Prompt Tuning

- LLM parameters remain frozen
- “Soft prompt” is optimized using gradient descent

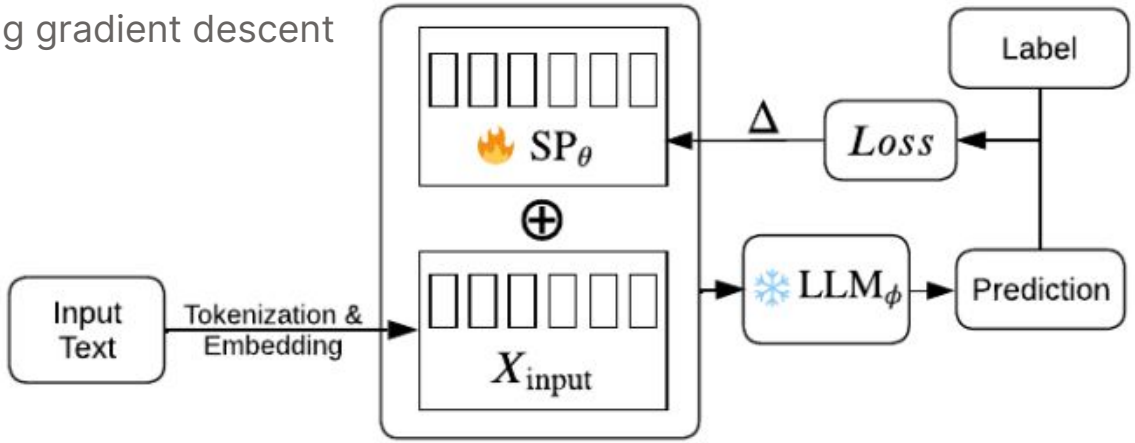


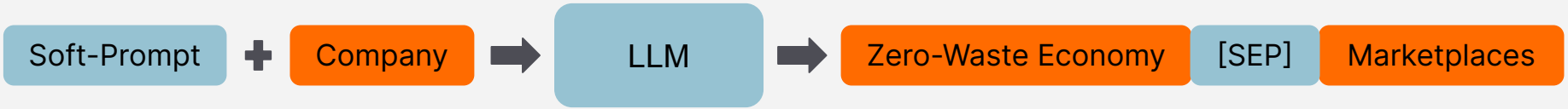
Figure 1: Schematic overview of Prompt Tuning, showing the trainable *soft prompt* (matrix SP_θ), the tokenized and embedded input text (X_{input}), and the LLM with frozen parameters (LLM_ϕ).

MLC as a Text-to-Text Problem

Multi-label text-to-text - sequential label generation: Sector1; Sector2; Sector3...

Challenges:

a. Model will often produce a different label of similar meaning
b. There is no logical order to the labels
c. Model returns binary decision rather than confidence scores/probabilities



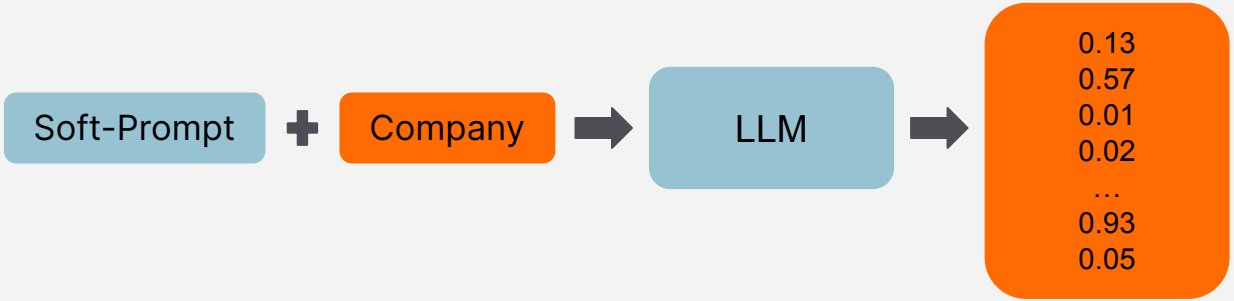
MLC as a Text-to-Text Problem

Multi-label text-to-text - sequential label generation: Sector1; Sector2; Sector3...

Challenges:

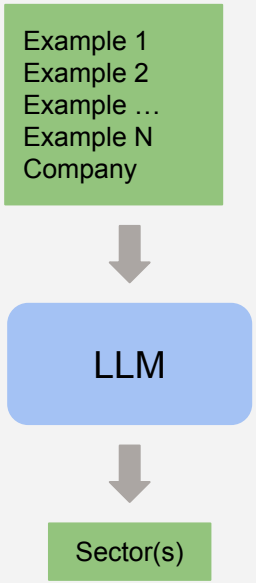
a. Model will often produce a different label of similar meaning
b. There is no logical order to the labels
c. Model returns binary decision rather than confidence scores/probabilities

→ Replacing the language head with a classification head

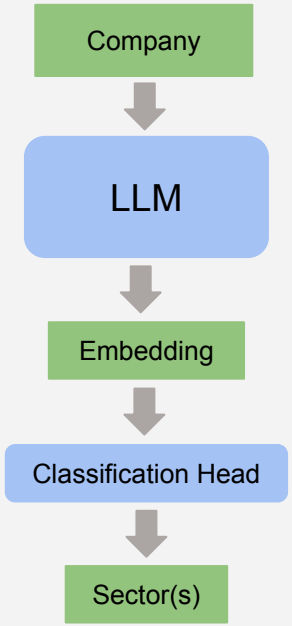


Experiments

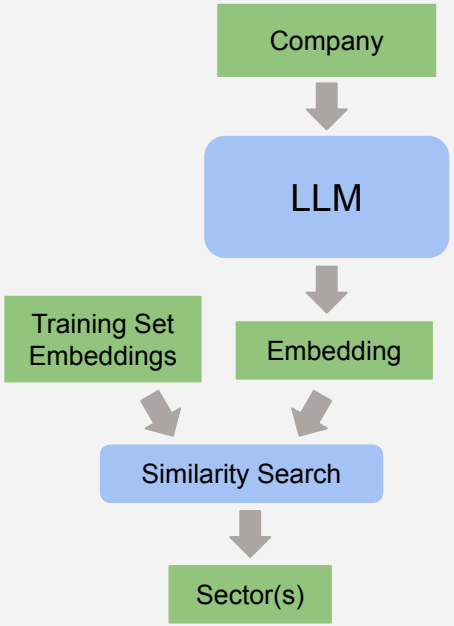
N-Shot Prompting



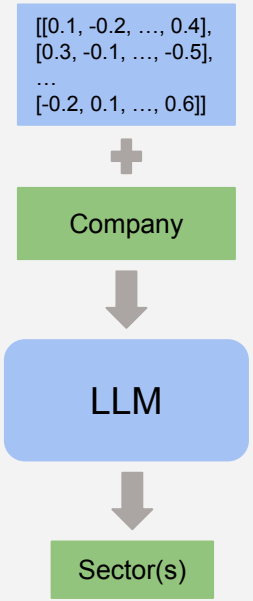
Classification Head



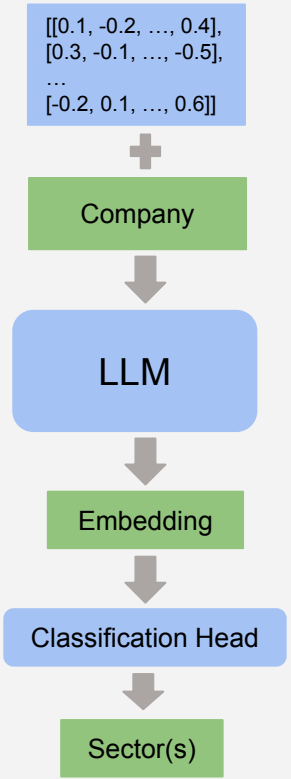
Vector Similarity



Prompt Tuning

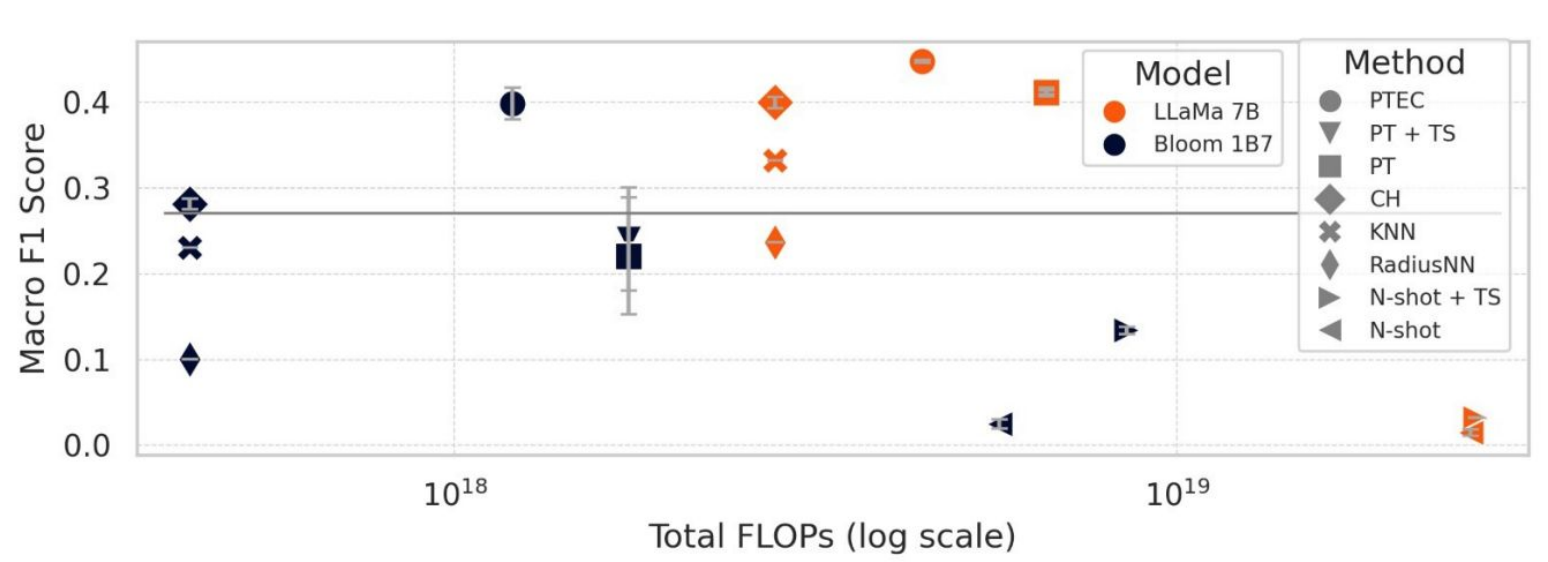


PTEC



Results

- PTEC outperforms all other methods
- PTEC is more efficient than text-to-text prompt tuning
- PTEC allow for adjusting between precision and recall



Results

- PTEC outperforms all other methods
- PTEC is more efficient than text-to-text prompt tuning
- PTEC allow for adjusting between precision and recall

- LLMs lack long-tail knowledge

  → Task specific domain adaptation