Legal Prediction

PILOT: Legal Case Outcome Prediction with Case Law

Team 4 : The Golden Lord Ring

Table of Content

PILOT Model Summary

Objective of PILOT

- Predict outcomes of legal cases in case law systems.
- Find similar past cases (precedents).
- Adapt to changes in legal decisions over time.

Data Source

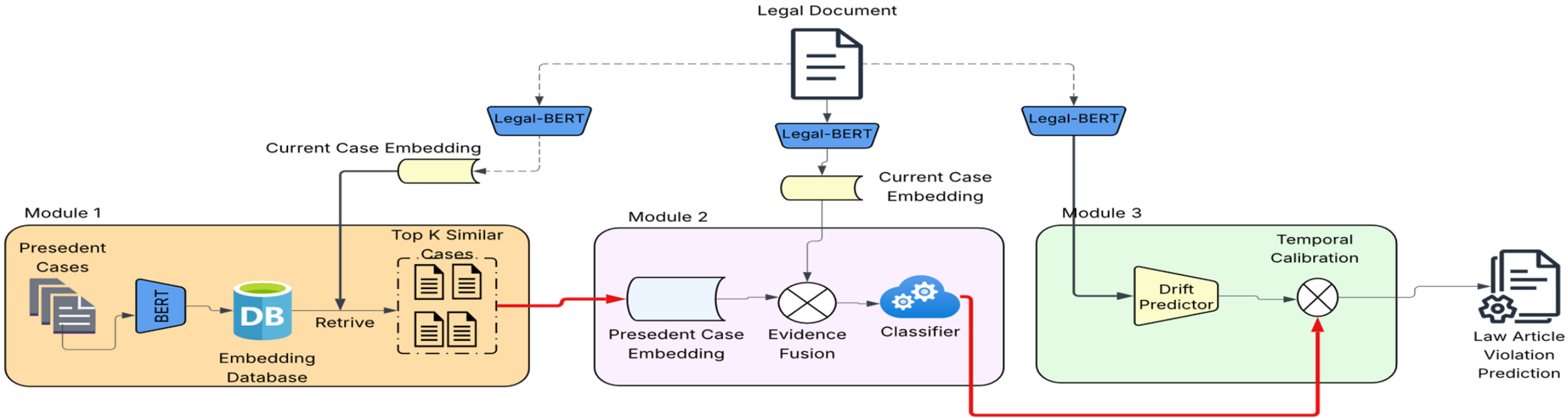
- European Court of Human Rights (ECHR).
- Dataset: ECHR2023 with thousands of real cases.

Data Preparation

- Original documents were long, complex, and multilingual.
- Used GPT-3.5-turbo to summarize the FACT section.
- Each case includes summary, date, legal articles, and outcome.
- Dataset split by time: train (8,138), validation (3,000), test (3,000).

Methodology

Module 1 Relevant Case Retrieval	Module 2 Case Encoder with Evidence Fusion	Module 3 Temporal Shift Mining
▪ Train BERT using contrastive learning ▪ Generate and store embeddings using Legal-BERT ▪ Retrieve top-k similar cases using compute similarity between current case and precedent cases	▪ Encode target case with Legal-BERT ▪ Compute weighted evidence embedding ▪ Concatenate and classify combined vector	▪ Predict drift vector using MLP. ▪ Add drift to classifier output for temporal calibration
▪ Top-k most relevant precedent case embeddings	▪ Preliminary prediction vector (y_orig)	▪ Final calibrated prediction (y_final) indicating law article violations



PILOT: Experiment Summary

Evaluation Strategy

3 Instances
- Training (8138 Cases)
- Validation (3000 Cases)
- Testing (3000 Cases)

4 Metrics
- micro-F1
- micro- PR-AUC
- micro-Jaccard
- micro-ROC-AUC

5 Runs
- Different random seeds

8 Baselines
- BERT
- BERT+CL+kNN
- HIER-BERT
- BERT+TemporalAttention
- BERT-LWAN
- LWDROV2
- EPM-base
- ChatGPT 5-shots

Experiment Outcome

Top-Ranked Performer: **PILOT** ▲
Lowest Performer: **ChatGPT** ▼

Legal Case Outcome Prediction
- Outperformed all models.
- MICRO-F1 ↑ 2.74% over LWDROV2.

Ablation Study
- Performance dropped with legal semantics.

Case Retrieval (Qualitative)
- Retrieved precedents + comprehensive coverage.

Hyperparameter Analysis
- Balanced settings successfully identified.

PILOT Limitations

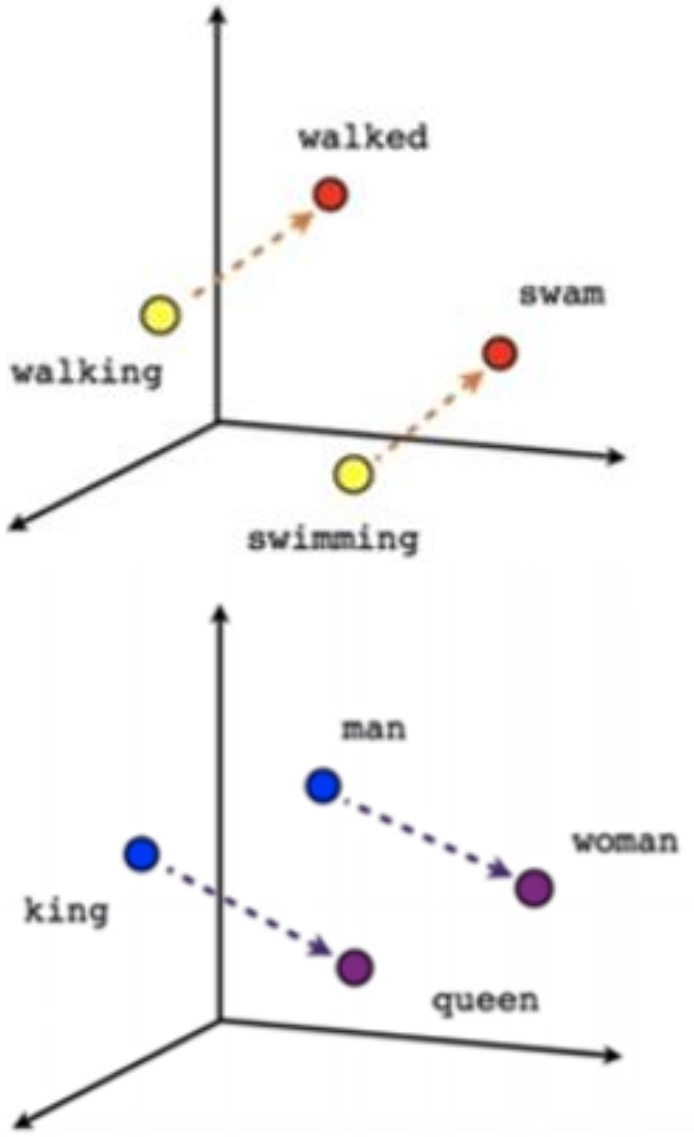
- The paper's final section opens by recognizing that real world **legal cases are a very complex process**.
    - Testimony, intent, evidence, human reasoning etc.
    - Law is not just a pattern, it’s an argument

PILOT Limitations

- The paper's final section opens by recognizing that real world **legal cases are a very complex process**.
    - Testimony, intent, evidence, human reasoning etc.
    - Law is not just a pattern, it’s an argument
- We are remined that this is a **research paper**, and to facilitate this research, many settings were simplified.
- As such, the authors openly acknowledge that PILOT, as it stands, **cannot be applied** in the real world.

PILOT Limitations

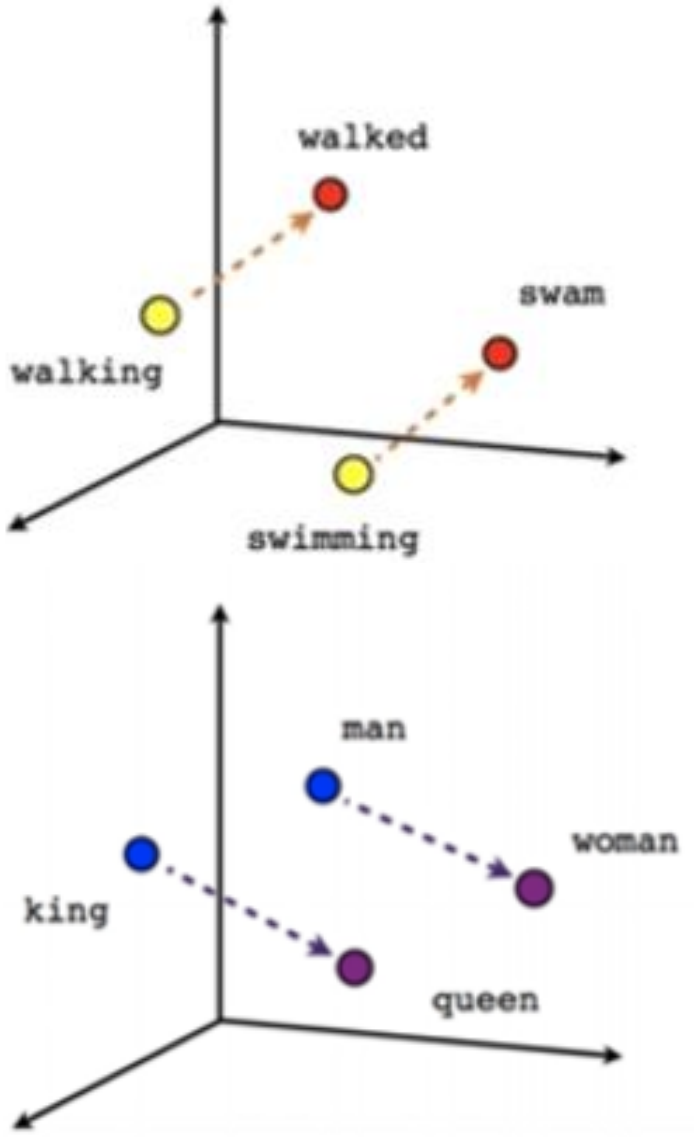
- The selection of precedent cases which the paper largely attributes to its success over previous models are **determined on semantic similarity alone**.



PILOT Limitations

- The selection of precedent cases which the paper largely attributes to its success over previous models are **determined on semantic similarity alone**.

- Addressing temporal shift is another prominent feature. However, a **simple linear decay function is used.**

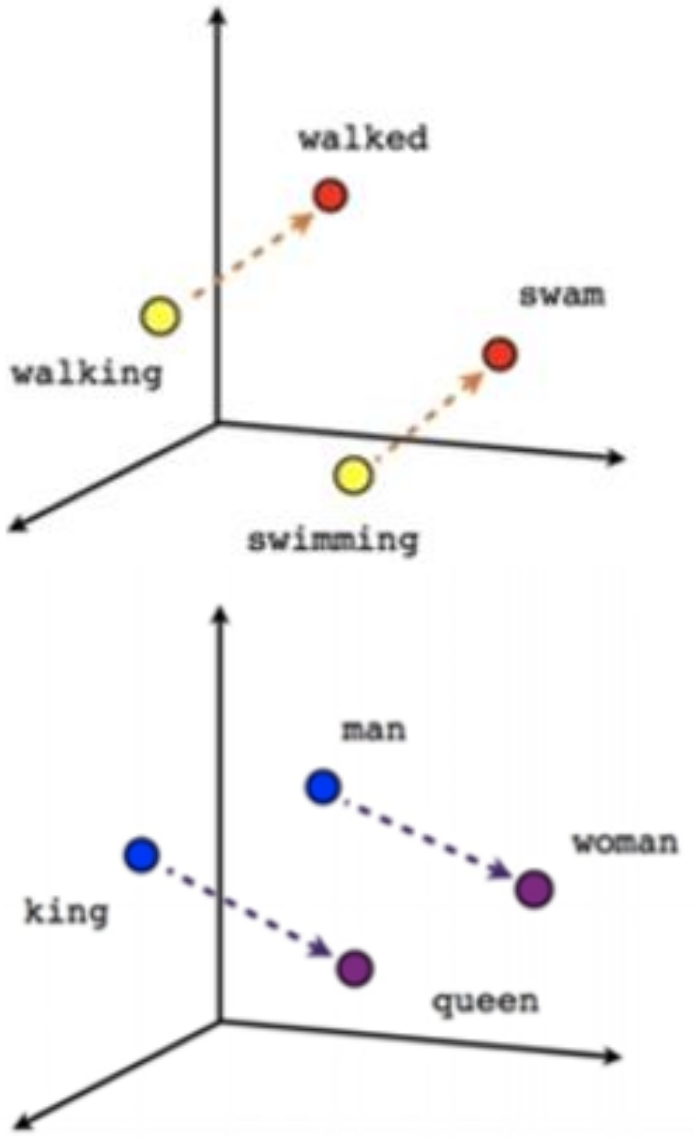


PILOT Limitations

- The selection of precedent cases which the paper largely attributes to its success over previous models are **determined on semantic similarity alone**.

- Addressing temporal shift is another prominent feature. However, a **simple linear decay function is used.**

- PILOT **only uses a legal case's factual section** to make predictions, a fraction of what is available.



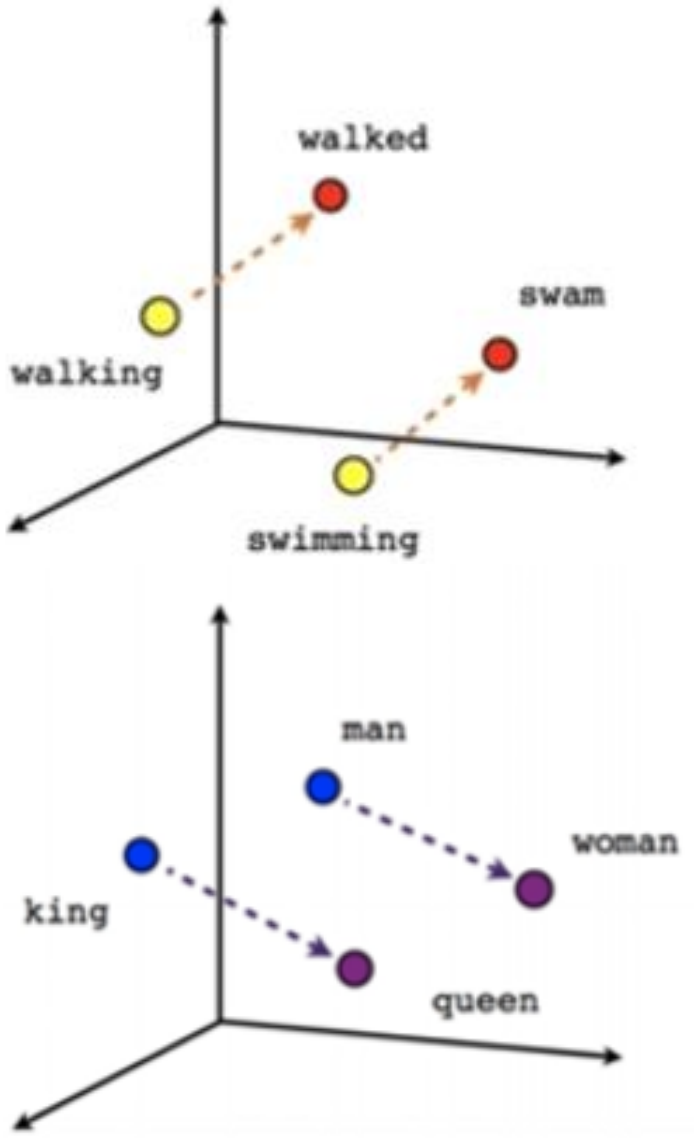
PILOT Limitations

- The selection of precedent cases which the paper largely attributes to its success over previous models are **determined on semantic similarity alone**.

- Addressing temporal shift is another prominent feature. However, a **simple linear decay function is used.**

- PILOT **only uses a legal case's factual section** to make predictions, a fraction of what is available.

- Historical bias in data.
  Model explainability.
  Ethical consequences



Key Takeaways

The complexity and sparsity of legal cases make AI unsuitable, as it struggles to capture the intricate nuances which is require for judgment.

AI is unsuitable for law due to its **non-human-like** nature, lacking the ability to sense human **emotions** and **ethical**

- Fahad

Judges or lawyers may rely on PILOT as a crutch, **reducing genuine human legal deliberation**. PILOT should not replace legal professionals.

Every law case should be treated with a level of **equity rather than a pattern to be mimicked**.

- Noel

The PILOT model shows that AI in law must be used **carefully**. It should be **fair**, protect people's **privacy**, and **help** judges, not replace them. The model also needs to be clear about how it makes decisions, so people can trust it.

- Sekar

Legal judgement goes beyond technical analysis. It is shaped by **power, place** and **space**. The law should be **tempered with mercy**.

The AI models **lack the capacity to replicate** human emotion or understand social context.

- Timmy

Thank you!

Q&A Time