Prompt engineering and provision of context in domain specific use of GPT

Marton Ribary (Royal Holloway, University of London)
Paul Krause (University of Surrey)
Miklós Orbán (OPL gunnercooke / University of Surrey)
Eugenio Vaccari (Royal Holloway, University of London)
Thomas Wood (Fast Data Science / Ulster University)

JURIX 2023 - 36th International Conference on Legal Knowledge and Information Systems
Maastricht, the Netherlands
18-20 December 2023

Ulster University | Fast Data Science | University of Surrey | Royal Holloway University of London

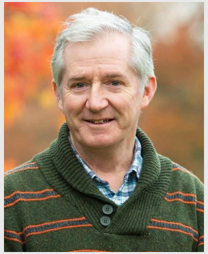
slides: https://bit.ly/insolvency-llm-jurix-2023

paper: https://doi.org/10.3233/FAIA230979

The team

Marton Ribary
(Royal Holloway, University of London)

legal reasoning, private law, computational methods



Paul Krause
(University of Surrey)

complex systems, theoretical and practical aspects of AI

Miklós Orbán
(OPL gunnercooke / University of Surrey)

legal tech, industry perspective, data-driven legal analysis

Eugenio Vaccari
(Royal Holloway, University of London)

qualified Italian lawyer, corporate insolvency and bankruptcy law

Thomas Wood
(Fast Data Science / Ulster University)

data science, machine learning models, information extraction

Headlines

Hypothesis: query responses from an LLM will be improved if the model is enhanced with a trusted domain specific knowledge base.

Area: we tested LLMs' performance in the procedural and technical area of **insolvency law** in which our team has relevant expertise.

Results: on the "unseen test set":

- the Insolvency Bot based on gpt-3.5-turbo outperformed gpt-3.5-turbo ($p = 1.8\%$)
- the Insolvency Bot based on gpt-4 outperformed gpt-4 ($p = 0.05\%$)

The English insolvency context

Micro-, Small- and Medium-Sized Enterprises (MSMEs):

- the backbone of modern economies
- increased insolvency risk (Covid, rising costs of dept etc.)

> “company insolvencies in Q2 2023 was the highest since Q2 2009, 9% higher than in Q1 2023, and 13% higher than in Q2 2022”

England and Wales:

- no simplified MSME-specific rules (unlike many common law jurisdictions)
- but sufficiently flexible and modular approach to corporate restructuring

⇒ potential to assist: MSMEs in distress, solo legal practitioners and smaller law firms lacking sufficient legal expertise to assess situations related to insolvency

The knowledge base

Statute

custom database of relevant statutes (5), one row for each section

source: legislation.gov.uk

structure: chapter, subchapter, preamble, paragraph, title, text

2898 rows in total

HMRC forms

custom database of relevant HMRC forms

source: forms for insolvency rules and forms for limited companies

structure: form name, form instruction, cited legislation

190 rows in total

Case law

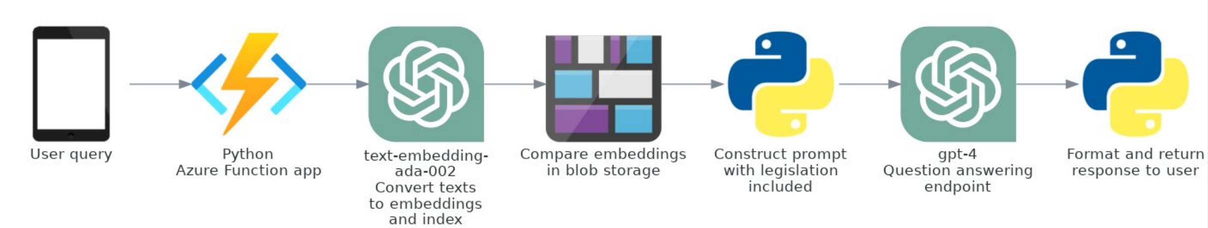
custom database of related cases

source: Find Case Law (The National Archives) and the Insolvency Lawyers' Association

structure: summary (some custom written), manually assigned keywords (plain English), cited legislation

198 cases in total

Design of the Insolvency Bot



Triaging the user query and identifying relevant law:

- rule-based keyword matching algorithm
- zero-shot classification - sentence embeddings (text-embedding-ada-002)
- matching with (vectorised) items from the domain specific knowledge base (cosine distance)

Prompt engineering:

- structured query enhanced with relevant law passed to LLM via API
- response returned to user

[Insolvency Bot (gpt-4)]

Methodology

Test and fine tuning:

- 12 queries related to insolvency (from a batch of 60) from the "Legal, Employment and Insolvency" section of the UK Business Forum platform

Final testing:

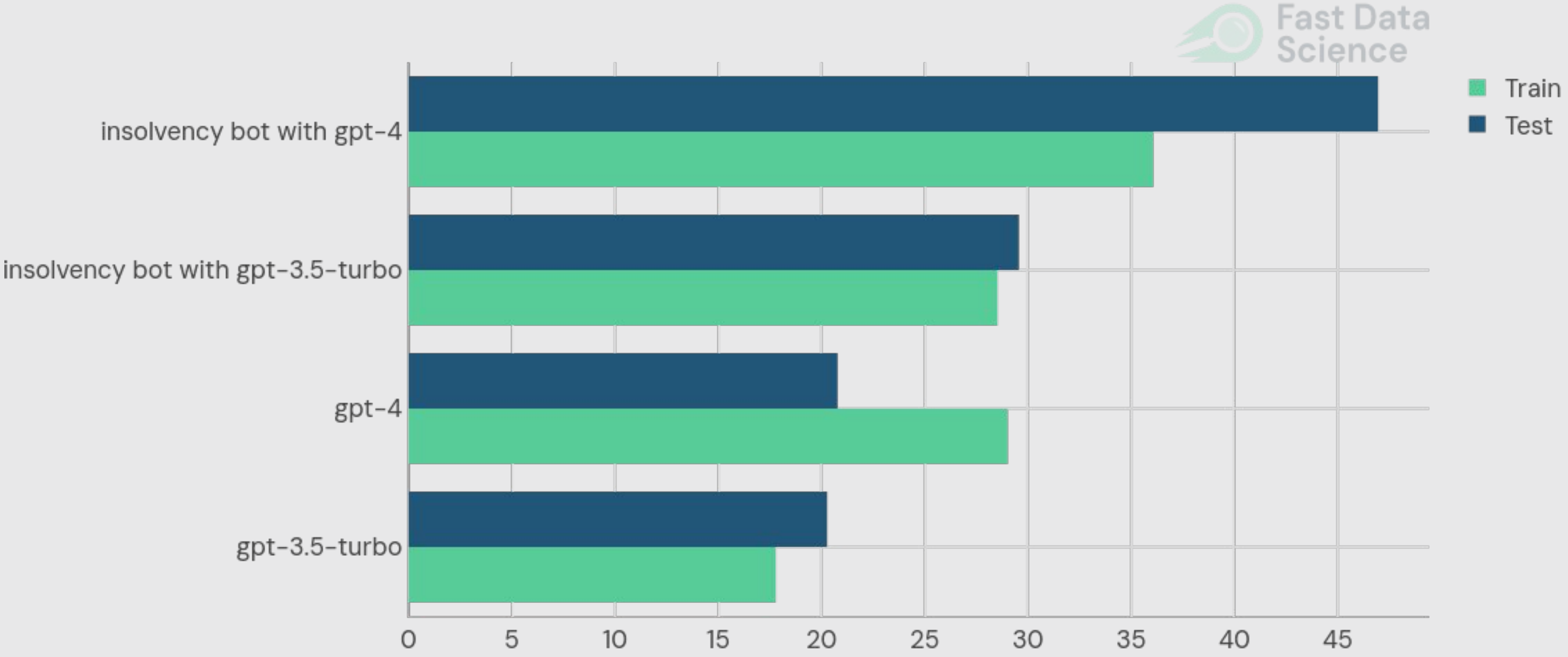
- additional 12 queries created by a domain expert
- matching mark schemes created by a domain expert (7-10 weighted question, total of 25 points) commensurate to a Level 6-7 (=1st-2nd year undergraduate) law student
- automated evaluation by parsing answers to simple yes-no questions

Results (by test queries)

Table 1. Scores of unmodified GPT bots and those enhanced by the Insolvency Bot (IB) according to our marking scheme

Question no.	Points available	gpt-3.5-turbo	gpt-4	IB (gpt-3.5-turbo)	IB (gpt-4)
Q1	25	6	12	12.5	15.5
Q2	24	3	3	3	4.5
Q3	25	3	3	3	10
Q4	25	3	3	9	5
Q5	22	3	3	3	12
Q6	25	6	6	9	14
Q7	25	3	6.5	1.5	9.5
Q8	25	11	3	16.5	19.5
Q9	25	3	3	6	15
Q10	25	11	11	16	19.5
Q11	25	3	3	3	4.5
Q12	25	5	5	5	10
Total	**296**	**60**	**61.5**	**87.5**	**139**
Percent	-	**20%**	**21%**	**30%**	**47%**

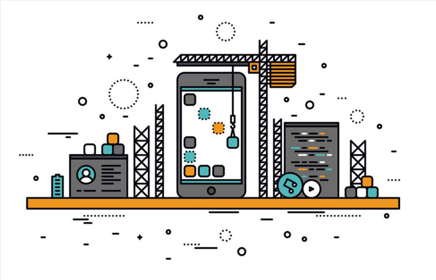
Results (train/test by bots)



Conclusion…

Our hypothesis confirmed: query responses from an LLM will be improved if the model is enhanced with a trusted domain specific knowledge base - modest results confirm that we are on the right track.

...next steps



System development

- more sophisticated techniques for matching queries with relevant law
- testing other LLMs, e.g. IBM foundation models, Mistral 7B
- training a custom language model

User experience

- testing the model online with real users (on [Prolific])
- organising a workshop with representatives of MSMEs, small law firms and public service providers

Expanding the insolvency scope

- presenting our work to insolvency experts, e.g. 35th Insolvency Symposium (London, Apr 2024) and INSOL Europe (Sorrento, Oct 2024)
- expanding the system to cover other jurisdictions as well as cross-jurisdictional queries

Thank you!

slides: https://bit.ly/insolvency-llm-jurix-2023

paper: https://doi.org/10.3233/FAIA230979