Automatic Linking of Judgements to UK Supreme Court Hearings

Hadeel Saadany[1], Constantin Orăsan[2] , Mikolaj Barczentewicz[3], Catherine Breslin[4], and Sophie Walker[5]

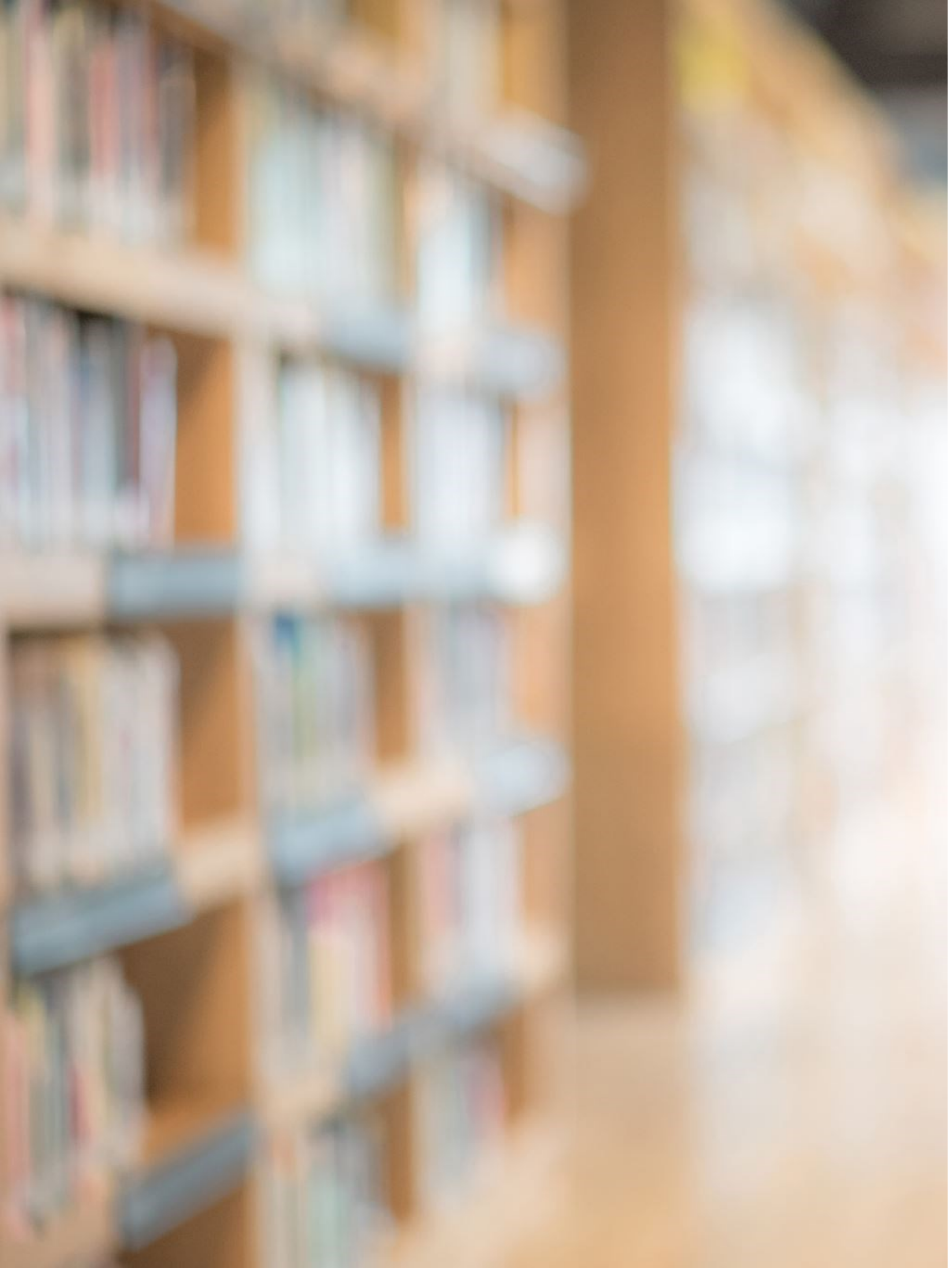
1,2 Centre for Translation Studies, University of Surrey, United Kingdom

3 School of Law, University of Surrey, UK

4 Kingfisher Labs Ltd, United Kingdom

5 Just Access, United Kingdom

LREC-COLING 2024



Context of this research

- Collaboration with industry to improve access to legal information
- Focus on making sessions of UK Supreme Court hearings more accessible

- Project funded by InnovateUK

https://dinel.org.uk/research/projects/harnessingNLP4court/

Challenges

- Audio material for a case typically spans over several hours on several days.
- Time, effort and money consuming to extract important information.
- There are over 449,000 cases each year in the UK across all court tribunals which are largely transcribed by humans.

Our Tool: UI for automatic linking of SC judgement and video timespans

Agbaje v Agbaje

Judgment

1. Part III of the Matrimonial and Family Proceedings Act 1984 was enacted to give the English court the power to grant financial relief after a marriage had been dissolved (or annulled) in a foreign country. This appeal raises for the first time at this appellate level the proper approach to the operation of Part III of the 1984 Act.

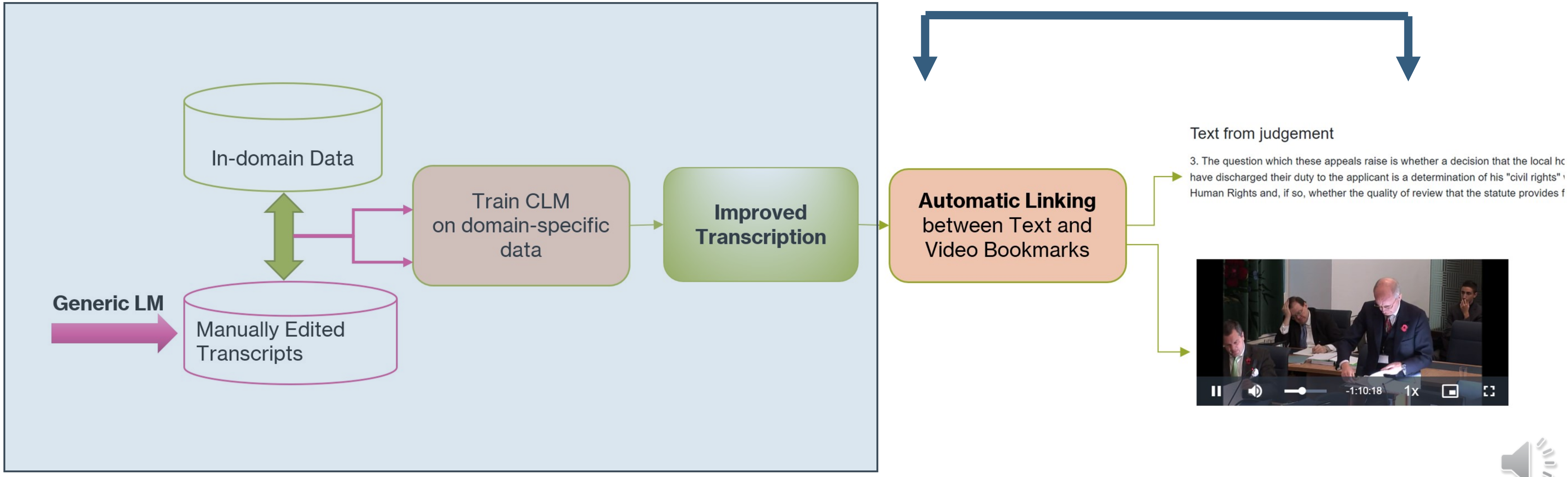
2. Mr and Mrs Agbaje ("the husband" and "the wife") were married for 38 years prior to their divorce in 2005 on the husband's petition in Nigeria. They were born in Nigeria, but both have British and Nigerian citizenship. All five children of the family were born in England. The wife has been living in England continuously since 1999, when the marriage broke down. The assets are about £700,000, of which £530,000 represents two houses in London in the husband's name, and the balance represents properties in Nigeria. The Nigerian court awarded the wife a life interest in a property in Lagos (which, as found by the Nigerian court, had a capital value of about £86,000) and a lump sum which was the equivalent of about £21,000.

Day 1 Session 1
02:28:37

00:31:50.440 - 00:32:39.870 - Day 1 Session 1

I'm going to come onto this later, when you look at the judgement of Lord Justice Ward, he has made the point, well, if you don't grant an anti-suit injunction, but you stay the English petition here on the basis that Nigeria is the appropriate forum. How can it be right to say that Nigeria shouldn't deal with the case? Well, the irony is that the judge who dealt with the application postulated the possibility of the wife making a Part 3 application in his judgement. I was dealing with the the forum shopping acquittal.

A Pipeline for Automatic Judgement-Hearing Linking



Hadeel Saadany, Catherine Breslin, Constantin Orăsan, and Sophie Walker (2023) Better Transcription of UK Supreme Court Hearings . In *Workshop on Artificial Intelligence for Access to Justice (AI4AJ 2023)* , Braga, Portugalt

Our Research Contributions

We introduce an application of Doc2Doc IR from two distinct linguistic modes, written and spoken, with legal-specific jargon and vocabulary.

We compile and release a publicly available dataset which contains links between segments of UK Supreme Court hearings to paragraphs from court judgements[1].

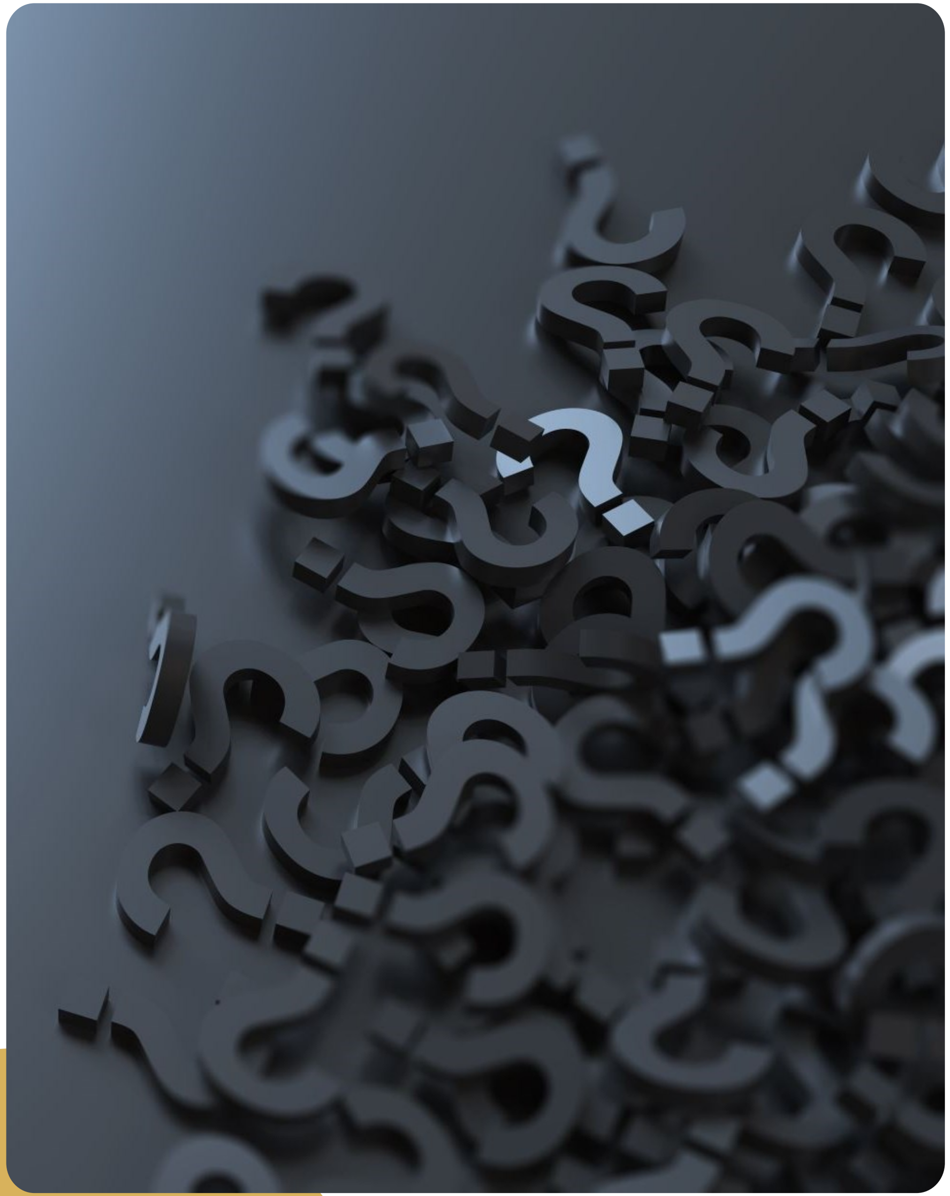
We show that the GPT 3 text embeddings customisation produce the best results with respect to the IR document representations

1. https://github.com/surrey-nlp/Linking-Judgements

Data Compilation

1. **Transcribe** the video sessions using a custom speech-to-text language model we developed in stage one of the project (Saadany et al., 2022)
2. **Preprocess and segment** the judgement into paragraphs.
3. **Treat** paragraph(s) as a ***query*** and the transcript of the case as the **corpus** in which we search for an answer to that query.

7 UKSC case judgements = 1.4M tokens and over 53 hours of video material



Retrieval methods

- **Frequency based method**
  - BM25
- **Embedding-based methods**
  - Document similarity with pooling
  - Entailment search
  - LegalBERT
  - Asymmetric Semantic Search
  - GPT Question-answer linking

Zero-shot Pre-fetching Stage

Embed all judgement and transcript segments for **one case** into the same vector space.

1

Use the cosine similarity as our semantic distance metric to extract the top closest 20 transcript timespans per judgement segment.

2

Evaluate manually first 20 links produced by each model.

3

Assess performance, choose **best model** for annotation. Annotate the rest of dataset to create a gold standard.

Zero-shot Information Retrieval

Model	MAP@5	Recall@5	MAP@10	Recall@10	MAP@15	Recall@15
GPT	**0.691**	**0.391**	**0.622**	0.657	**0.711**	**0.914**
Entailment	0.615	0.348	0.568	0.611	0.66	0.885
Glove	0.526	0.316	0.506	0.602	0.607	0.884
BM25	0.655	0.377	0.612	**0.659**	0.698	0.902
Asymmetric	0.602	0.347	0.553	0.619	0.664	0.908
LegalBert	0.557	0.326	0.531	0.613	0.632	0.896

Results of zero-shot IR for linking judgements to video transcripts for all cases

Judgement-hearing relevancy model

- Train a model that classifies a judgement and hearing segments as relevant or not
- The data annotation is very expensive
- Experimented with data augmentation

Data augmentation

- Used Generative AI technology to produce a larger gold-standard

- **Positive instances generation:**
  - InstructGPT API set role prompt strategy "***I want you to act like a British lawyer. Paraphrase the following text***"

- **Negative instances generation:**
  - Random shuffling of judgement-hearing segments from different cases. Reduce randomness by choosing judgement-hearing segment pairs with the highest cosine similarity scores.
  - In-batch negative sampling during training.

- The augmented dataset has 7,248 judgement-hearing links and over 42m tokens

- A sanity check was conducted on a sample of the AI-generated paraphrases by a legal expert

Experiments

- **Baseline Model**: logistic regression with GPT3 embeddings
- **Cross-encoder** built on top of distilled version of RoBERTa-base
- **Cross Tension with In-batch Negative Sampling**
- **OpenAI GPT3 Embedding Customisation**

Results

Model	Accuracy	Precision	Recall	F1
GPT 3(-)	0.69	0.84	0.64	0.73
GPT 3(+)	0.78	**0.85**	0.75	0.80
GPT 3(+) + cos_sim	0.83	0.91	0.79	**0.85**
GPT 3 Customised(+)	0.83	0.84	0.83	0.83
GPT 3 Customised(+) + cos_sim	**0.85**	**0.85**	0.84	**0.85**
Cross-encoder(-)	0.69	0.61	**0.93**	0.74
Cross-encoder(+)	0.81	0.79	0.84	0.81
CT with in-batch negatives	0.69	0.63	0.90	0.74

Results of Relevancy Models on Augmented (+) and non-Augmented (-) Dataset

Thank you

Hadeel Saadany, hadeel.saadany@surrey.ac.uk
University of Surrey, United Kingdom

Constantin Orăsan, c.orasan@surrey.ac.uk
University of Surrey, United Kingdom

Mikolaj Barczentewicz
University of Surrey, United Kingdom

Catherine Breslin
Kingfisher Labs Ltd, United Kingdom

Sophie Walker
Just Access, United Kingdom