



UNIVERSITY OF
CAMBRIDGE

AI-based Court Outcome Prediction: Insights and Challenges

NAIL Conference 2026

Felix Steffek

Hamburg, 03/07/2026



Access to justice

What is the best thing we can do with AI in law?

Anticipate legal consequences = judgement prediction

Better dispute resolution

Better contracts

Increase in individual and societal welfare

Papers and book

The CLC-UKET Dataset: Benchmarking Case Outcome Prediction for the UK Employment Tribunal,

in: Proceedings of the Natural Legal Language Processing Workshop 2024, pp. 81–96, 16 November 2024: <https://aclanthology.org/2024.nllp-1.7/>, also available at <https://ssrn.com/abstract=4954617> and <https://arxiv.org/abs/2409.08098>,

H. Xie, F. Steffek, J. Ribeiro de Faria, C. Carter and J. Rutherford

Shortcut Learning in Legal Judgment Prediction,

In: <https://arxiv.org/abs/2607.04261>,

J. Watson, J. Ribeiro de Faria, M. Tomalin, C. Carter, J. Rutherford, M. Magnusson, H. Xie, H. Yeung and F. Steffek

Towards Explainable Adjudicative Variance: Quantifying Judicial Discretion via Gated Multi-Task Learning,

ICML 2026, AI for Law Workshop: <https://arxiv.org/abs/2606.27069>

S. Sójka, F. Steffek and M. Grabmair

Information Extraction from Employment Tribunal Judgments Using a Large Language Model,

in: (2025) Artificial Intelligence and Law (Artif Intell Law): <https://doi.org/10.1007/s10506-025-09443-z>

J. Ribeiro de Faria, H. Xie and F. Steffek

The Cambridge Law Corpus: A Corpus for Legal AI Research,

in: Neural Information Processing Systems Datasets and Benchmarks Track (NeurIPS 2023): <https://neurips.cc/virtual/2023/poster/73448>; also available at <https://ssrn.com/abstract=4763429> and <https://arXiv:2309.12269>,

A. Östling, H. Sargeant, H. Xie, L. Bull, A. Terenin, L. Jonsson, M. Magnusson and F. Steffek

Expanding the Cambridge Law Corpus: UK Legislation, Expanded Case Law, and Benchmarks for Legal AI,

Work in progress,

H. Xie, H. Sargeant, T. K. Jain, H. Waldetoft, C. Li, L. Bull, M. Tomalin, M. Magnusson and F. Steffek

Legal Innovation: Conversations about Technology, the Legal Profession and Societal Change,

Cambridge University Press, 2025, XVIII, 243 pp: <https://doi.org/10.1017/9781009427371>

F. Steffek and M. Sumida (eds)

Legal Judgment Prediction

Issues

General outcome prediction

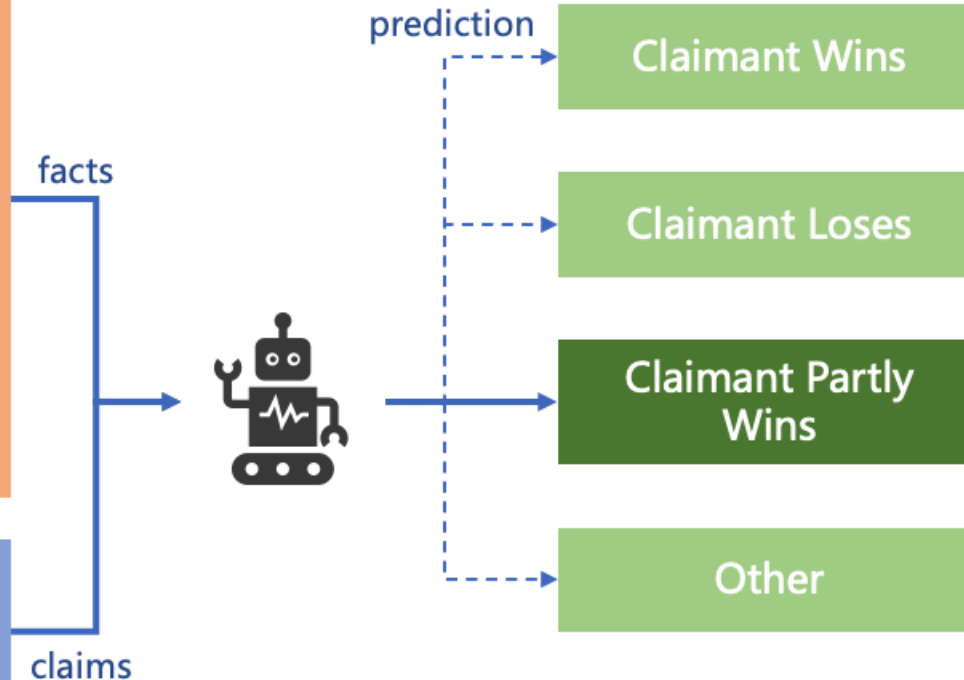
Data limitations

LLMs v other models v human experts

Employment tribunal: decision prediction

The claimant, Mr. L Tarbuck, was an employee of Peninsula Business Services Limited. He had been working for the company since 4 January 1999 as a salesman. In July 2013, the claimant's terms and conditions of employment were changed, which included the removal of a sales target and the threat of disciplinary action if targets were not met. However, in January 2014, the respondent sought to reinstate the sales target and threatened the claimant's job if he did not agree to the new terms. The claimant felt he had no choice but to accept the new terms. He resigned on 7 July 2014, citing the respondent's conduct as the reason for his resignation.

The claimant made claims for constructive unfair dismissal, age discrimination, disability discrimination, non-payment of holiday pay and non-payment of commission.



Employment tribunal decision prediction: results

Baseline	Label	Precision	Recall	F-score
BERT	wins	0.459	0.828	0.591
	loses	0.869	0.215	0.345
	partly	0.381	0.364	0.372
	other	0.036	0.455	0.067
T5	wins	0.595	0.716	0.650
	loses	0.647	0.846	0.734
	partly	0.541	0.066	0.117
	other	0	0	0
GPT-3.5 _{juris5}	wins	0.515	0.700	0.594
	loses	0.720	0.565	0.633
	partly	0.362	0.305	0.331
	other	0.143	0.455	0.217
GPT-4 _{juris2}	wins	0.588	0.700	0.639
	loses	0.778	0.430	0.554
	partly	0.359	0.541	0.431
	other	0.082	0.364	0.133
Human	wins	0.627	0.815	0.708
	loses	0.792	0.812	0.802
	partly	0.554	0.302	0.391
	other	0	0	0



- Humans and all models generally predict well and clearly outperform random guess baseline (F-score of 0.276)
- Humans expert baseline for ‘wins’ and ‘loses’ better than AI model baseline
- AI better than humans for complex cases (‘partly’ and ‘other’)
- Different models excel for different predictions and scores
- Some interesting patterns, e.g., asymmetry for wins and loses in Precision and Recall
- Precision (Recall) good if cost of false positive (negative) is high
- F-score balances Precision, Recall and uneven dataset
- Baselines only

Conclusion

Prediction of UKET decisions shows promising baseline scores for both generative and non-generative AI models

Variation between models, outcome type predicted and scores

Limitations

Claim Level Prediction and Shortcut Learning in Legal Judgment Prediction

Issues

Predicting success of claims v predicting general outcome

Shortcut learning?

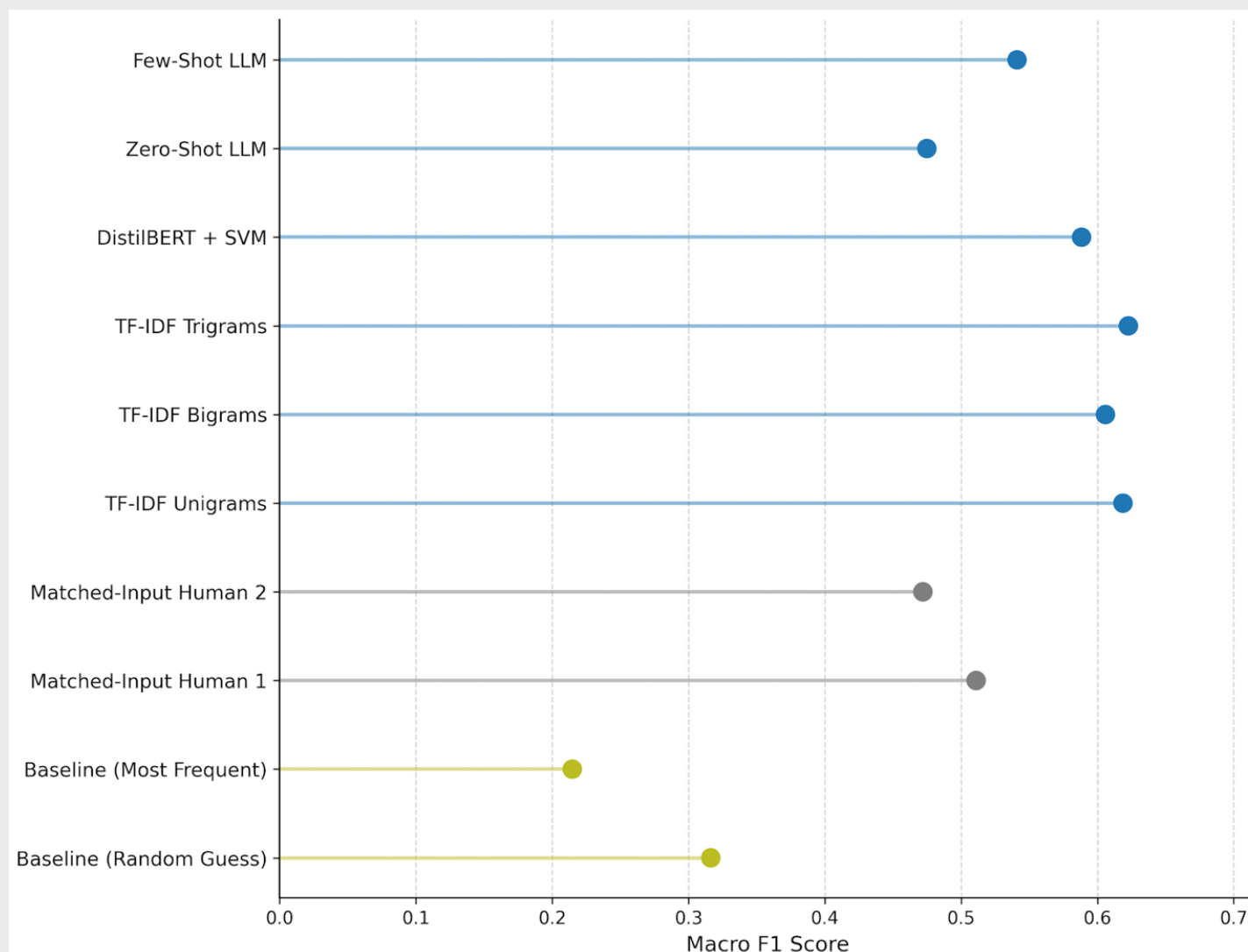
Examples: 'unauthorised deduction from' for Win predictions; 'claimant failed to' for Loss prediction

Can shortcut learning be remedied?

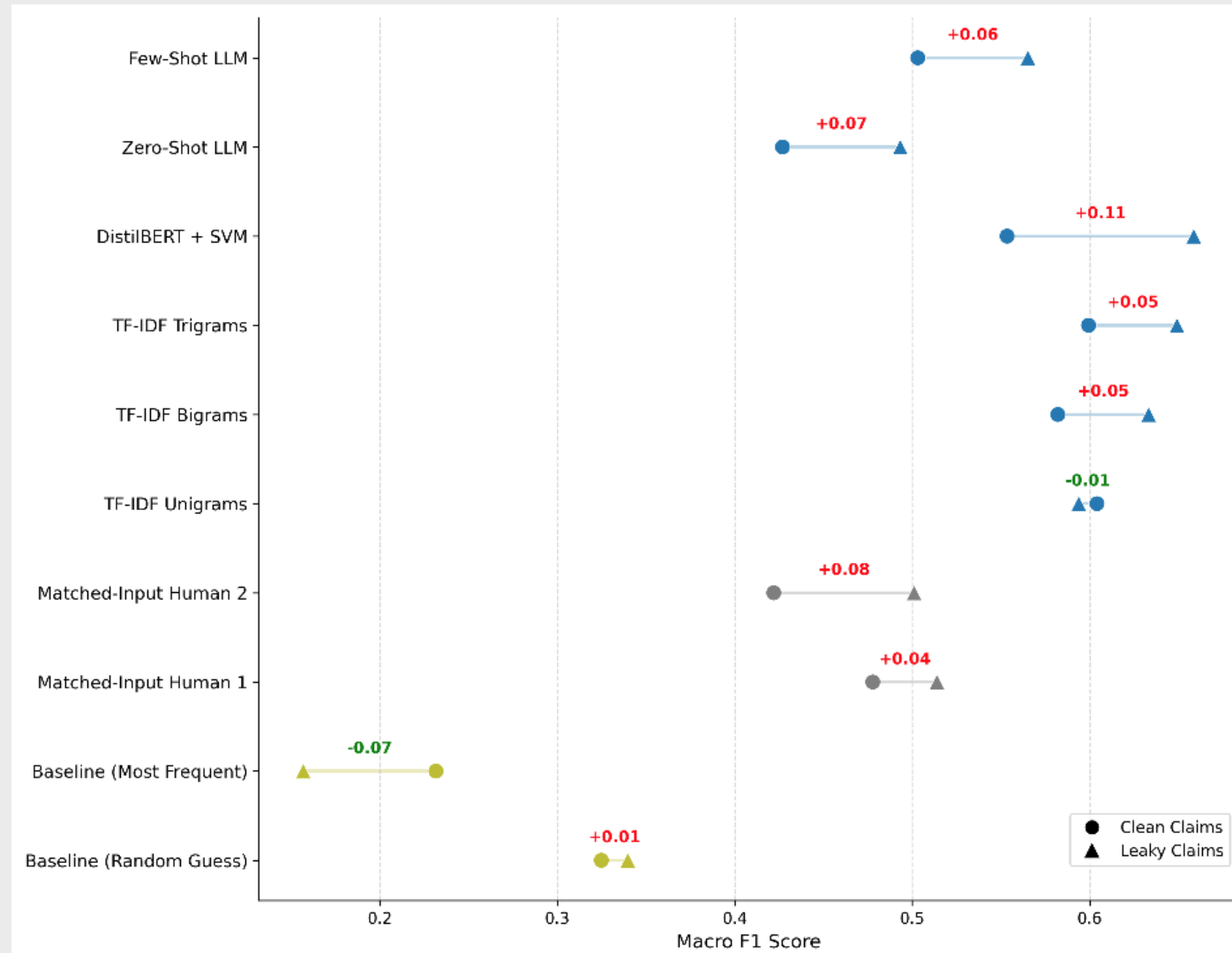
Dataset

Category	Statistic	Value
Totals	Total Judgments (Cases)	13,253
	Total Claims	33,158
	Avg. Claims per Judgment	2.50
Train/Test Split	Training Set (LLM-annotated)	32,030 (96.6%)
	Test Set (Human-annotated)	1,128 (3.4%)
Annotator Distribution	Expert-Annotated Ground-Truth Labels	1,128
	LLM-Annotated Training Labels	32,030
Class Distribution (Overall)	Win	13,104 (39.5%)
	Loss	16,174 (48.8%)
	Other / Procedural	3,880 (11.7%)
Leakage (Test Set Only)	Claims marked as leakage by human expert	289 (25.6%)
Text Length (Tokens)	Claim Summaries (Mean / Median / Range)	5.6 / 4.0 / 1–30
	Case Facts (Mean / Median / Range)	80.7 / 80.0 / 2–462

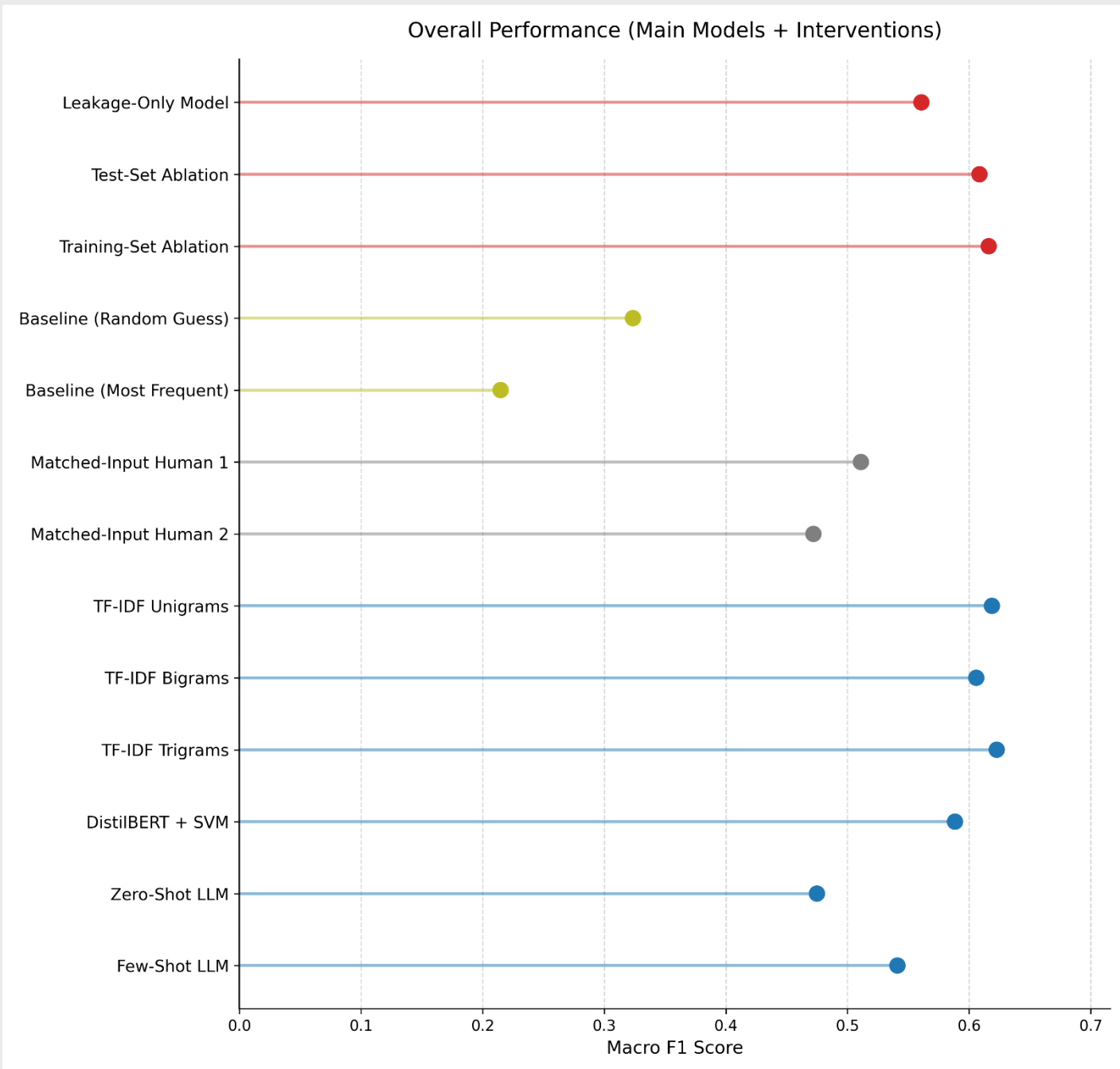
Employment tribunal decision prediction: claim level



ET claim level prediction: leakage gap (clean v leaky claims)



Employment tribunal decision prediction: intervention effect on performance



Conclusion

‘Serviceable’ claim level prediction scores

Risk of shortcut learning due to data limitations

Shortcut learning can be remedied

Decision Type and Judge Influence on Outcome Prediction

Issues

Does it matter what kind of decision is taken?

Can judge decision patterns improve outcome prediction?

Ethical questions

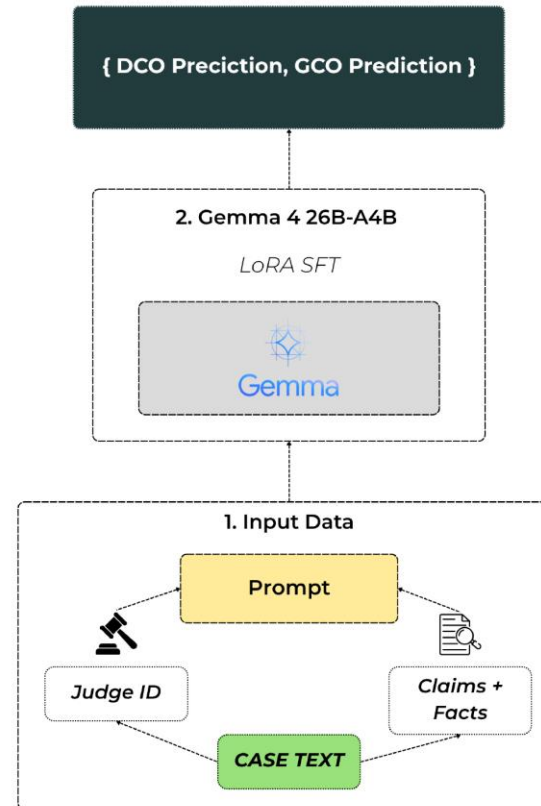
Dataset classification

Func. Cat.	DCO Class	Count	Description
3=Merit-Based	substantive win	3,463	Won on merits after evidence evaluation
	substantive loss	3,341	Lost on merits after evidence evaluation
	default judgment	1,307	(Sufficiency Check)
Mixed	claimant partly wins	2,991	Mixed outcome (Partial success)
7=Non-Merit	out of time	731	Dismissed as time-barred (Limitation)
	strike out non pursuit	579	Strike-out due to failure to pursue
	strike out other	532	Merged Strike-outs (Abuse/Compliance)
	other procedural dismissal	398	Other technical procedural defects
	lack of jurisdiction	377	Tribunal lacks jurisdiction
	withdrawal settlement	193	Withdrawal or explicit settlement
	other	698	Other miscellaneous outcomes

Prediction architecture

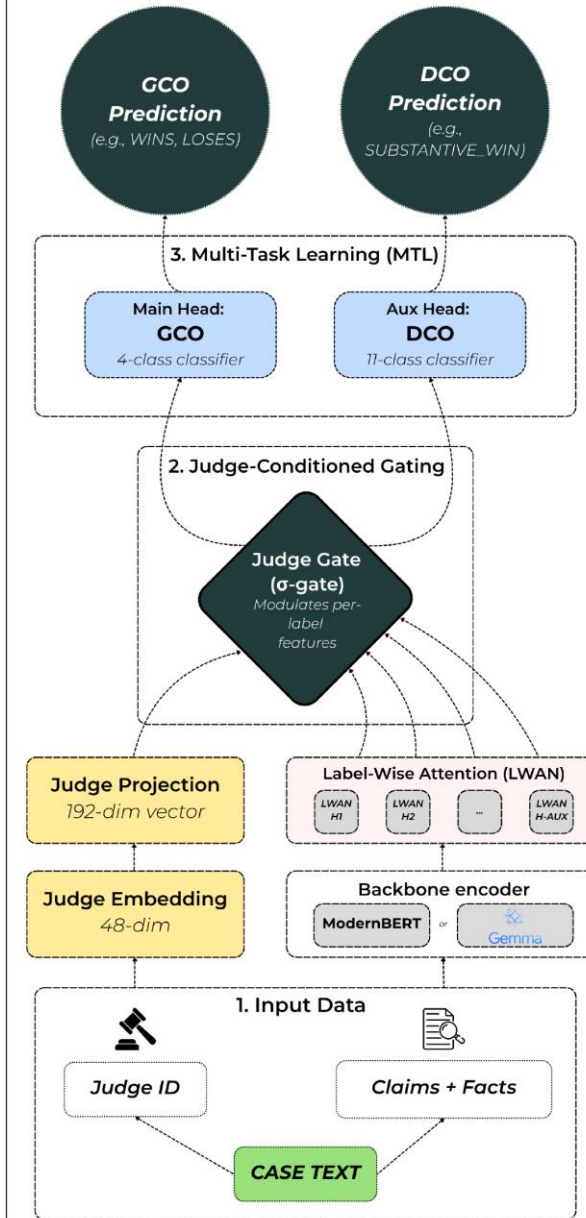
Generative SFT - Single autoregressive channel

judge as prompt token; DCO + GCO autoregressively co-supervised



Hybrid - Differentiable structured pathways

judge as learnable embedding fused via gated LWAN



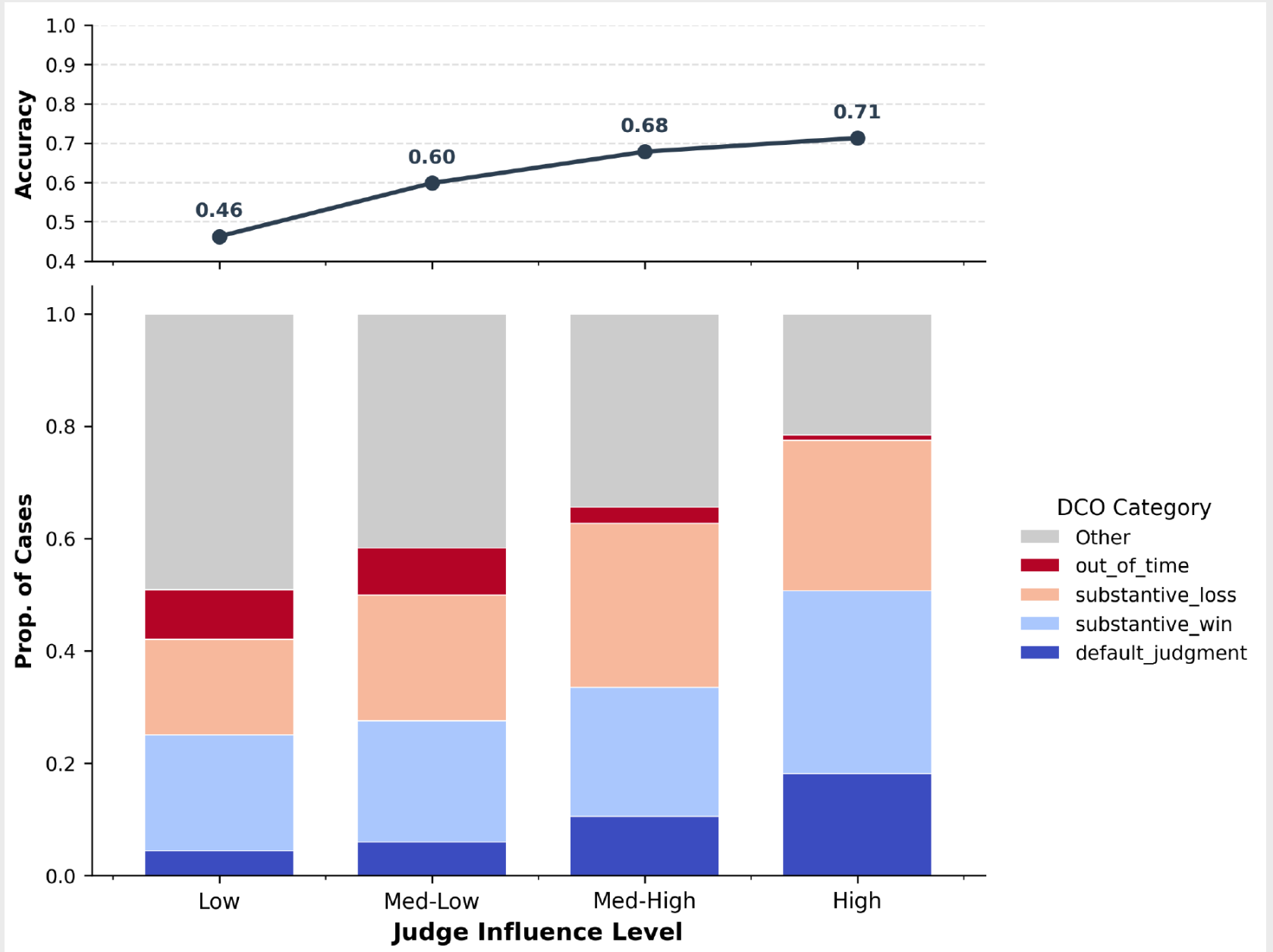
Judge map: relative specialisations and influence



Annotations indicate the top-2 outcomes with the highest relative lift per cluster. The color gradient denotes judge Influence.



Judge influence on case prediction



Conclusion

Predictions can be improved using types of decisions and judicial patterns

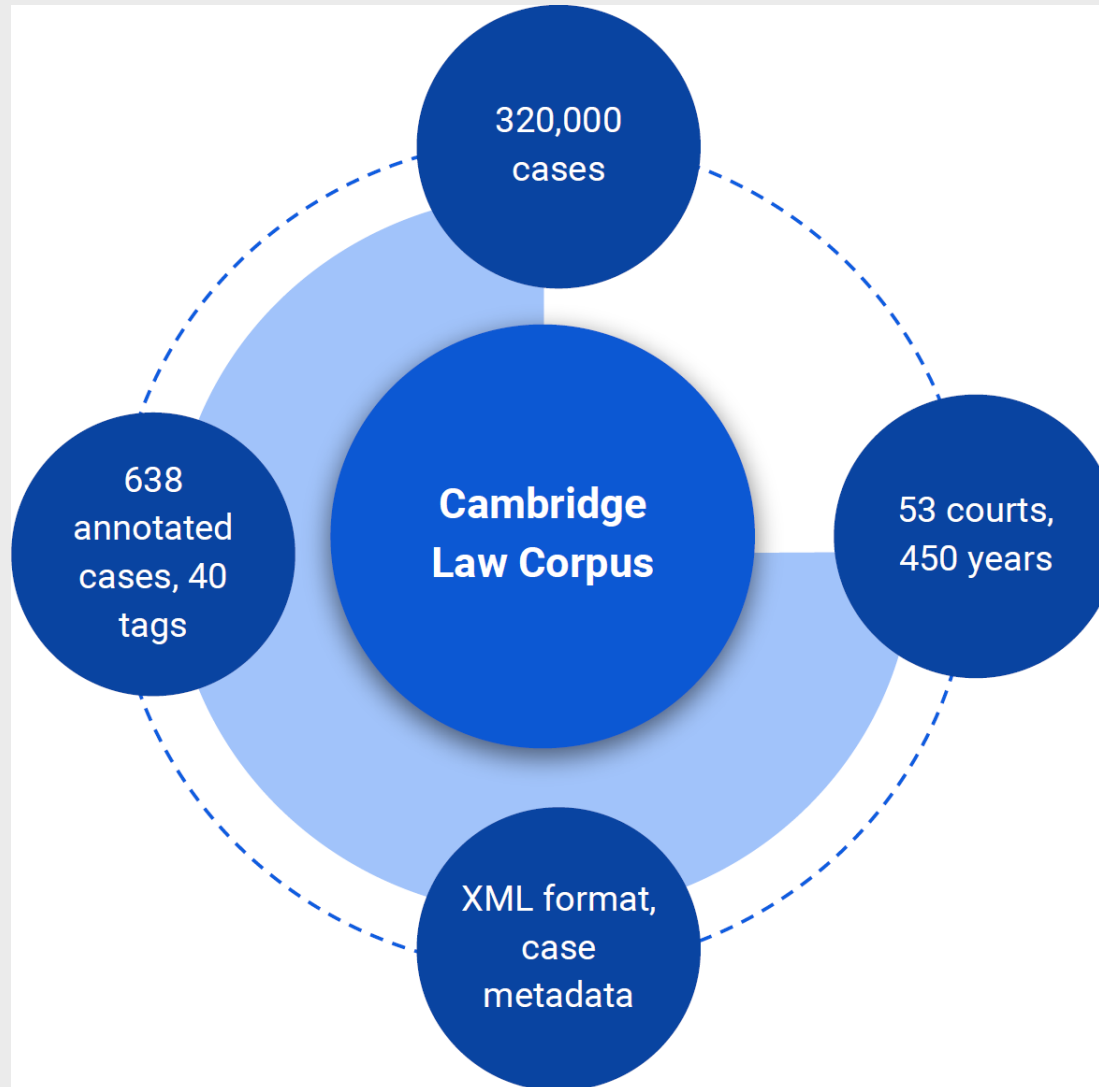
In this case the improved F1 was 65.21 (+ 5.1 over best alternative architecture)

Prediction accuracy gains concentrated on Partly Wins and Other cases

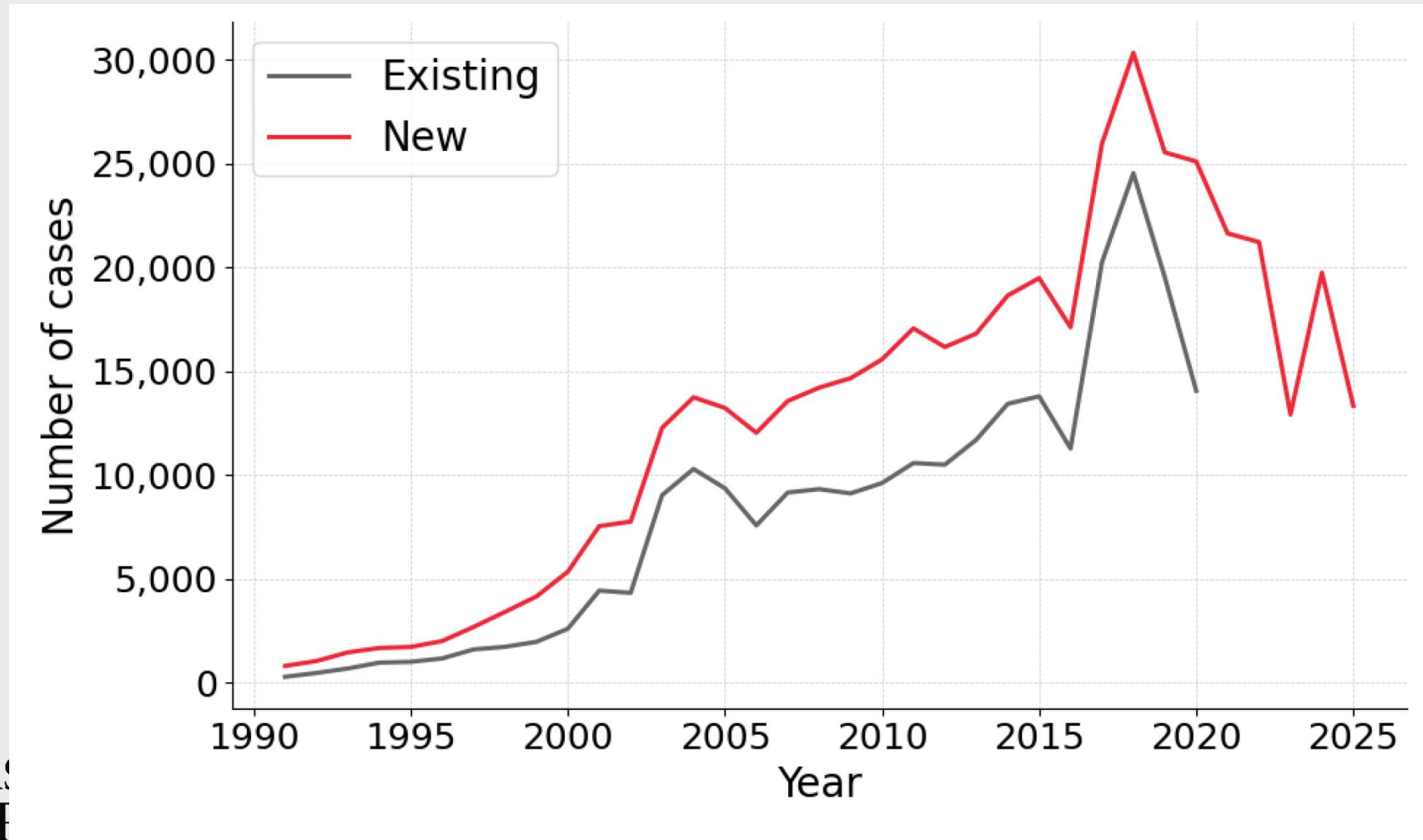
Ethical limitations

Cambridge Law Corpus

Cambridge Law Corpus (CLC): Version 1



Cambridge Law Corpus (CLC): Version 2



Cambridge Law Corpus (CLC): Version 2 – Cases

Court Abbreviation	Court Name	CLC 1.0	CLC 2.0
England and Wales			
EWCA-Civ	Court of Appeal (Civil Division)	18 461	20 276
EWCA-Crim	Court of Appeal (Criminal Division)	5 787	7 880
EWCC-Fam	County Court (Family)	111	111
EWCOF	Court of Protection	531	801
EWCSF	Care Standards Tribunal	745	745
EWFC-HCJ	Family Court (High Court Judges)	386	925
EWFC-OJ	Family Court (other Judges)	873	1 644
EWHC-Admin	High Court (Administrative Court)	11 555	14 068
EWHC-Admly	High Court (Admiralty Division)	101	133
EWHC-Ch	High Court (Chancery Division)	6 623	9 322
EWHC-Comm	High Court (Commercial Court)	2 926	3 892
EWHC-Costs	High Court (Costs)	0	781
EWHC-CP	Court of Common Pleas	6	6
EWHC-Exch	High Court (Exchequer Court)	11	13
EWHC-Fam	High Court (Family Division)	2 239	3 109
EWHC-IPEC	Intellectual Property Enterprise Court	170	278
EWHC-KB	High Court (King's Bench Division)	34	1 216
EWHC-Mercantile	Mercantile Court	22	22
EWHC-Patents	High Court (Patents Court)	642	834
EWHC-QB	High Court (Queen's Bench Division)	4 645	5 273
EWHC-TCC	High Court (Technology and Construction Court)	1 689	2 084
EWLandRA	Land Registry Adjudicator	482	482
EWLands	Lands Tribunal	570	570
EWLVT	Leasehold Valuation Tribunal	18 445	18 350
EWMisc	EW Courts – Miscellaneous	451	698
EWPBRA	The Parole Board	0	1 587
EWPCF	Patents County Court	135	135
European Union			
ECHR	European Court of Human Rights	0	31 321
EUBOA	Board of Appeal of the European Supervisory Authorities	0	20
EUECJ	Court of Justice of the European Union	0	29 998
Ireland			
IECA	Court of Appeal	0	3 935
IECCA	Court of Criminal Appeal	0	435
IECCC	Central Criminal Court	0	12
IECompA	Competition Authority	0	1 509
IEDPC	Data Protection Commission Case Studies	0	63
IEHC	High Court	0	16 123
IEIC	Information Commissioner	0	3 419
IESC	Supreme Court	0	2 363
Jersey			
JUR	Jersey Unreported Judgments	0	9 089
Northern Ireland			
NICA	Court of Appeal	0	1 693
NICC	Crown Court	0	662
NIE-Misc	NI Courts – Miscellaneous	0	108
NIFET	Fair Employment Tribunal	0	554
NIHC-Ch	High Court of Justice (Chancery Division)	0	444
NIHC-Fam	High Court of Justice (Family Division)	0	496
NIHC-KB	High Court of Justice (King's Bench Division)	0	349
NIHC-Master	High Court of Justice (Master's decisions)	0	238
NIHC-QB	High Court of Justice (Queen's Bench Division)	0	2 376
NIIT	Industrial Tribunals	0	6 837
NISSCSC	Social Security and Child Support Commissioners'	0	1 819
Scotland			
ScotCS	Court of Session	0	38 783
ScotHC	High Court of Justiciary	0	2 865
ScotIC	Information Commissioner	0	4 347
ScotJCR	Jury Court Reports	0	276

ScotSAC-Civ	Sheriff Appeal Court (Civil)	0	366
ScotSAC-Crim	Sheriff Appeal Court (Criminal)	0	155
ScotSC	Sheriff Court	0	2 028

St Helena			
SHCA	Court of Appeal	0	19
SHSC	Supreme Court	0	37

United Kingdom			
UKAITUR	Immigration and Asylum Unreported Judgments	65 034	73 370
UKCAT	Competition Appeals Tribunal	478	778
UKDRS	Nominet Dispute Resolution Service	0	3 817
UKEAT	Employment Appeal Tribunal	16 002	16 611
UKET	Employment Tribunal	36 772	57 888
UKFSM	Financial Services and Markets Tribunals	69	69
UKFTT-GRC	First-tier Tribunal (General Regulatory)	2 311	4 490
UKFTT-HESC	First-tier Tribunal (Health Education and Social Care)	437	576
UKFTT-PC	First-tier Tribunal (Property)	7 860	11 847
UKFTT-TC	First-tier Tribunal (Tax)	7 684	9 278
UKHL	House of Lords	4 929	4 929
UKIAT	Asylum and Immigration Tribunal	1 352	1 352
UKICO	Information Commissioner's Office	13 659	22 039
UKIntelP	Intellectual Property Office	0	10 591
UKIPTrib	Investigatory Powers Tribunal	51	73
UKIT	Information Tribunal	250	250
UKPC	The Judicial Committee of the Privy Council	10 334	10 548
UKSC	Supreme Court	743	955
UKSIAC	Special Immigrations Appeals Commission	174	233
UKSPC	Special Commissioners of Income Tax	439	439
UKSSCSC	Social Security and Child Support Commissioners'	2 467	2 467
UKUT-AAC	Upper Tribunal (Administrative Appeals)	3 048	3 819
UKUT-IAC	Upper Tribunal (Immigration and Asylum)	811	944
UKUT-LC	Upper Tribunal (Lands)	1 009	1 455
UKUT-TCC	Upper Tribunal (Tax and Chancery)	918	1 237
UKVAT	VAT & Duties Tribunals	2 767	2 767
UKVAT-Customs	VAT & Duties Tribunals (Customs)	104	104
UKVAT-Excise	VAT & Duties Tribunals (Excise)	785	785
UKVAT-IPT	VAT & Duties Tribunals (Insurance Premium Tax)	7	7
UKVAT-Landfill	VAT & Duties Tribunals (Landfill Tax)	12	12

Total		258 146	501 704
--------------	--	----------------	----------------

Cambridge Law Corpus (CLC): Version 2 – Legislation

Category	Statistic	Value
Overall Legislation	Legislative Documents	140,151
	Provisions	1,939,170
	Tokens	~500M
	Temporal Coverage	1267–2026
	Avg. Depth	7.39
Legislation Types	Primary	7,184
	Secondary	85,030
	EU-derived	36,951
	Draft	6,818
Primary Legislation	UK Public General Acts	4,379
	Local Acts	196
	Devolved Legislation	1,196
	Other	1,413
Secondary Legislation	UK Statutory Instruments	58,792
	Scottish Statutory Instruments	8,641
	Welsh Statutory Instruments	8,240
	NI Statutory Rules	9,295
EU-Derived Legislation	EU Regulations	24,458
	EU Directives	4,168
	EU Decisions	12,488
	EU Treaties	5
Structural Statistics	Avg. sections per document	14.26
	Avg. paragraphs per document	97.2
	Avg. tokens per provision	45.51