REUTERS/Jason Lee

Legal IR & NLP: Future Directions

Jack G. Conrad, TR Labs
Kripabandhu Ghosh, IISER Kolkata

The intelligence, technology and human expertise you need to find trusted answers.

the answer company™
THOMSON REUTERS®

ECIR 2023, 6 April 2023
Dublin, Ireland

OUTLINE ON FUTURE DIRECTIONS

- Traditional Perspective on Future Directions
- Recent Developments / Disruptions in AI & Law
- Everything seen through the lens of Large Language Model (LLM) applications
- The Pros & Cons of Generative AI & LLMs
  - suitable & unsuitable applications
- Explainability
  - Experiences and approaches

TRADITIONAL PERSPECTIVE ON FUTURE DIRECTIONS

For IR

- More robust models
  - Trained with larger & larger amounts of gold or silver data
- More hybrid models
  - E.g., combinations of sparse & dense retrieval techniques
  - Former being more explainable; latter being more conceptual, less brittle
- Growing use of neural retrieval approaches to information retrieval
- Enlisting more sophisticated user-centric models for information retrieval

For NLP

- Trends in Generative Models like GPT-3.5 & 4.0 will continue
- Conversational AI tools like ChatBots will become more powerful, sophisticated
- Additional applications of transfer learning in the training of such models
- Few-shot, One-shot, Zero-shot learning paradigms
  - Can apply above as well
- Advances in inference & prediction at scale
- A role for IR to empower NLP & generative models via approaches like REML

On Large Language Models as a Disruptive Force in the Field of AI & Law

RECENT DEVELOPMENTS / DISRUPTIONS IN AI & LAW (1/2)

Launch of Large Language Models like ChatGPT powered by GPT-3.5 / 4.0 & Bard

- Tremendous impact on the field, if only via the hype and upheaval
- Much written already of the benefits and risks associated with LLMs

Experts divided over whether long-term impact will be substantive or modest

- Some see recent releases as opportunities to develop, embrace powerful productivity tools
- Others see LLMs as disruptive and serious challenges for lawyers

The different types of users impacted:

- Judges, Clerks, Judicial aids – little affected
- Practicing attorneys, partners – impact depends on field, needs, specific. applications
- Academics – a new source for comparative research, analysis, reporting
- Law students – potentially greatly affected, esp. the unscrupulous

https://www.law.upenn.edu/live/news/15538-chatgpt-and-the-law

RECENT DEVELOPMENTS / DISRUPTIONS IN AI & LAW (2/2)

Illustrations

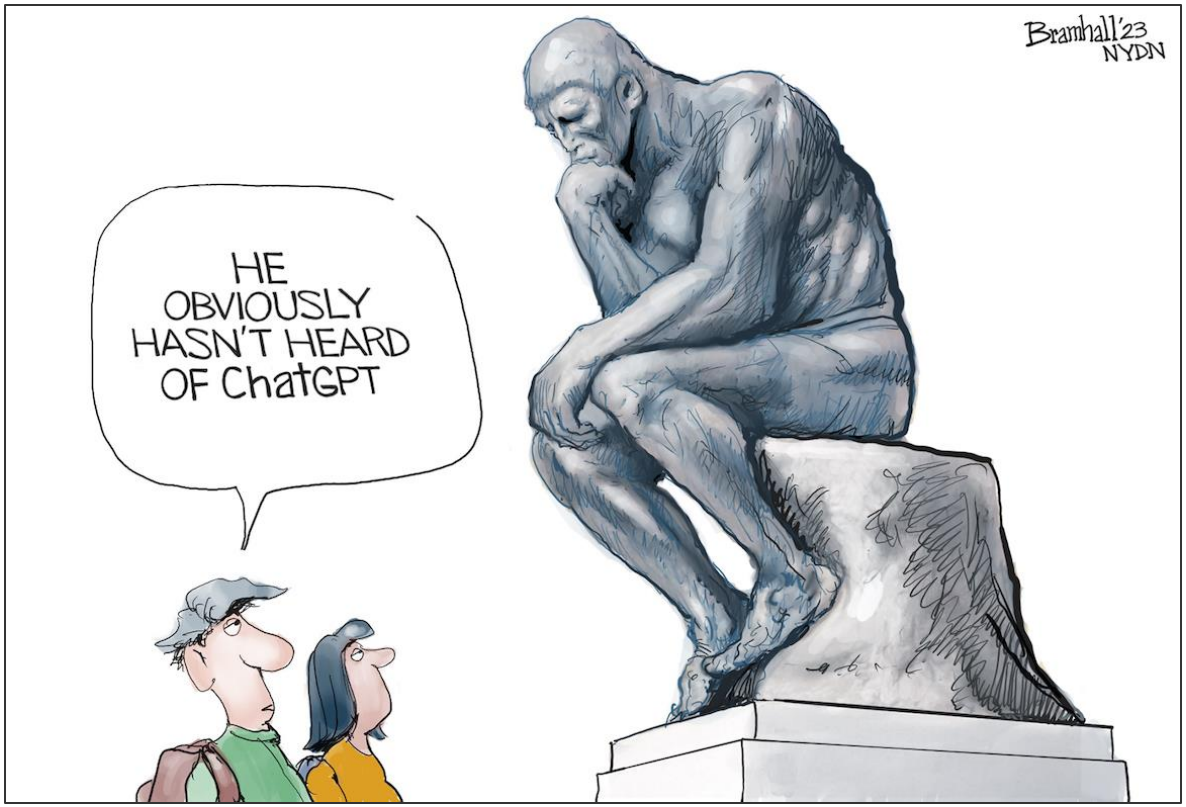
- A judge on a prominent High Court in India recently asked ChatGPT for its view on a bail in murder case. ChatGPT did not attempt to take a stance whether bail should be granted; rather, it pointed out what factors the judge should keep in mind. https://indianexpress.com/article/cities/chandigarh/in-a-first-hc-turns-to-chatgpt-for-view-on-bail-in-murder-case-8522544/

- Daniel Katz's group at the Chicago-Kent Law School at Illinois Tech has tried different versions of GPT on the task of answering legal questions from Bar Exams in the US. Their most recent paper says "GPT-4 significantly outperforms both human test-takers and prior models, demonstrating a 26% increase over ChatGPT and beating humans in five of seven subject areas."
  Papers: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4389233 and https://arxiv.org/abs/2212.14402
  Articles: https://www.iit.edu/news/gpt-4-passes-bar-exam
  https://www.lawnext.com/2023/03/gpt-takes-the-bar-exam-again-this-time-it-score-among-top-10-of-test-takers.html

- OpenAI indicates that GPT-4.0 is 40% more likely to produce factual responses than GPT-3.5 (the opposite of hallucinations) and is 82% less likely to respond to disallowed topics.

EVERYTHING THROUGH THE LENS OF LLM APPLICATIONS

Sign of the times



EVERYTHING THROUGH THE LENS OF LLM APPLICATIONS

Sign of the times

Technology group to Business group:

- "We have a proposed solution, based on principled approaches …"

Business group to Technology group:

- "That's great, but can the problem being addressed be done using ChatGPT instead?"

If it looks too good to be true, …

- "All that glitters is not gold … "

ChatGPT may be creating unrealistic expectations for AI-enhanced applications as a whole

Advisable to focus on principled evaluation of the LLM application at hand.

THE PROS & CONS OF GENERATIVE AI & LLMS

Pros

- Generally reasonable approximations to legal discourse
- Usually decent at producing first drafts, outlines, or, exchanges of ideas
- Produces substantial drafts in minutes
- Good tool for “staking out the territory” in an OpenAI-like environment
- Can lower the bar for access to justice for resource scarce parts of society

Cons

- As good as its training data – what it’s missing, without qualms, it may make it up
- Although reasonably coherent in legal discourse, it’s not a trained professional
- Currently no way for professionals (law school profs., judges, etc.) to determine when they’re used and when not
- May dumb down the level of the quality for a legal brief, contract, set of interview questions
- Can appear to be cogent, perceptive, rationale, but is not.

SOME POTENTIALLY SUITABLE USE CASES

Use Cases

- Document review
- Review & drafting
- Research & knowledge management
- Contract analysis & drafting
- Chatbots & assistance

Associated Observations

- Note the emphasis on drafts and drafting
- Not clear how much better such drafts will be for contracts et al. than templates
- Best suited for internal applications rather than for customer or client-facing content or communications
- Best to assist with accelerating the data creation or data ingestion processes

https://www.thomsonreuters.com/en-us/posts/legal/legalweek-chatgpt-expectations/

On the Importance of Explainability in the Field of AI & Law

ACCEPTANCE OF AI IN LAW

Support the Guardian

Available for everyone, funded by readers

Support us

The Guardian

News | Opinion | Sport | Culture | Lifestyle | More

World UK Coronavirus Climate crisis Environment Science Global development Football **Tech** Business Obituaries

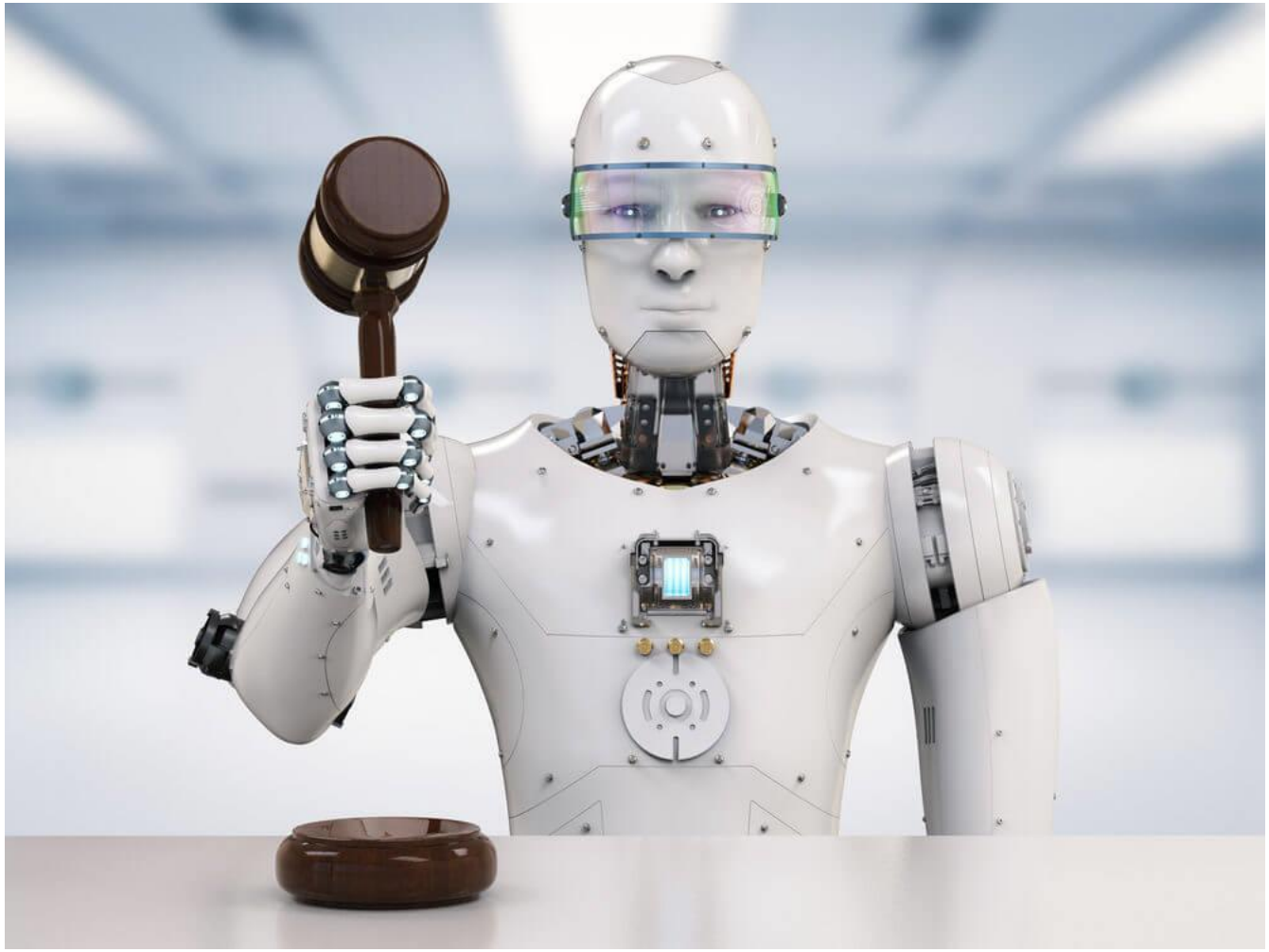
ChatGPT

Colombian judge says he used ChatGPT in ruling

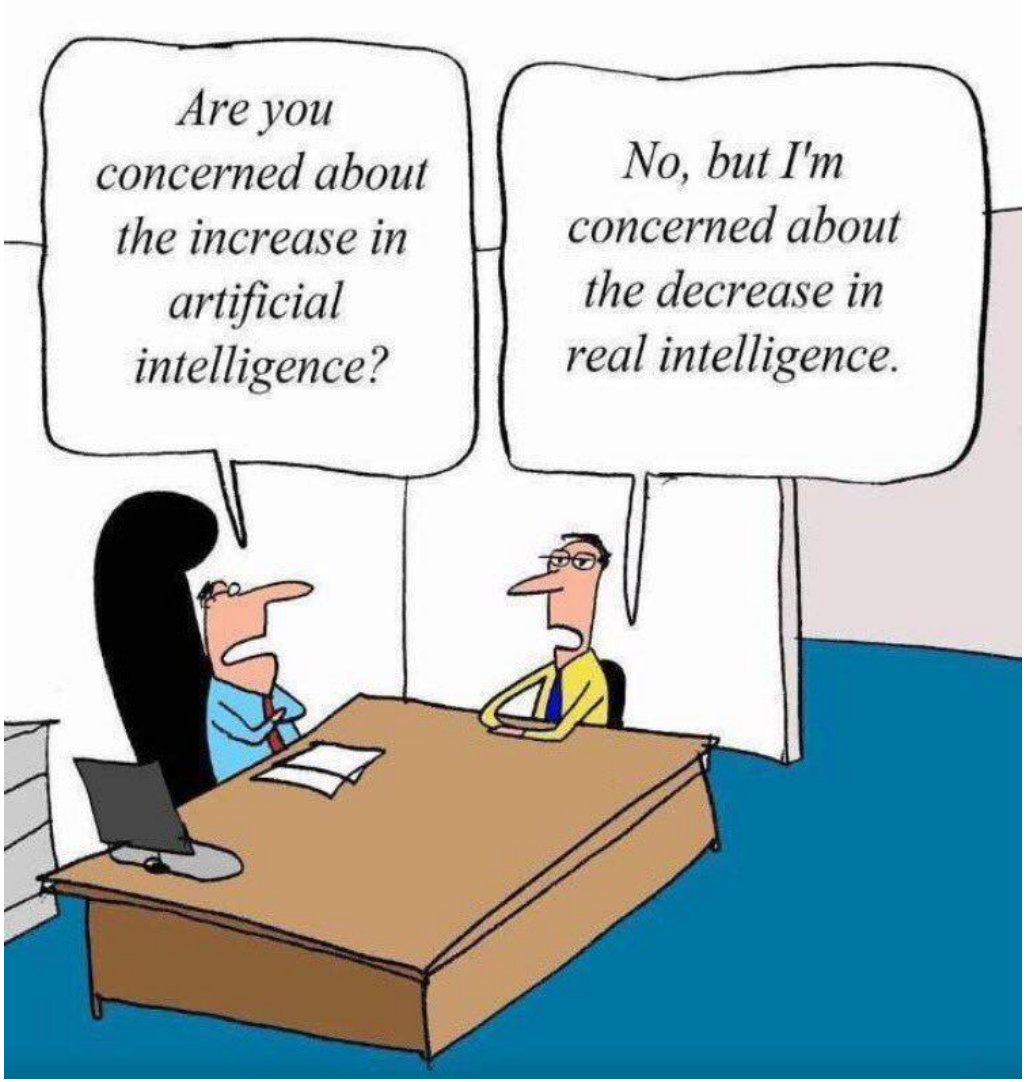
Juan Manuel Padilla asked the AI tool how laws applied in case of autistic boy's medical funding, while also using precedent to support his decision

Advertisement

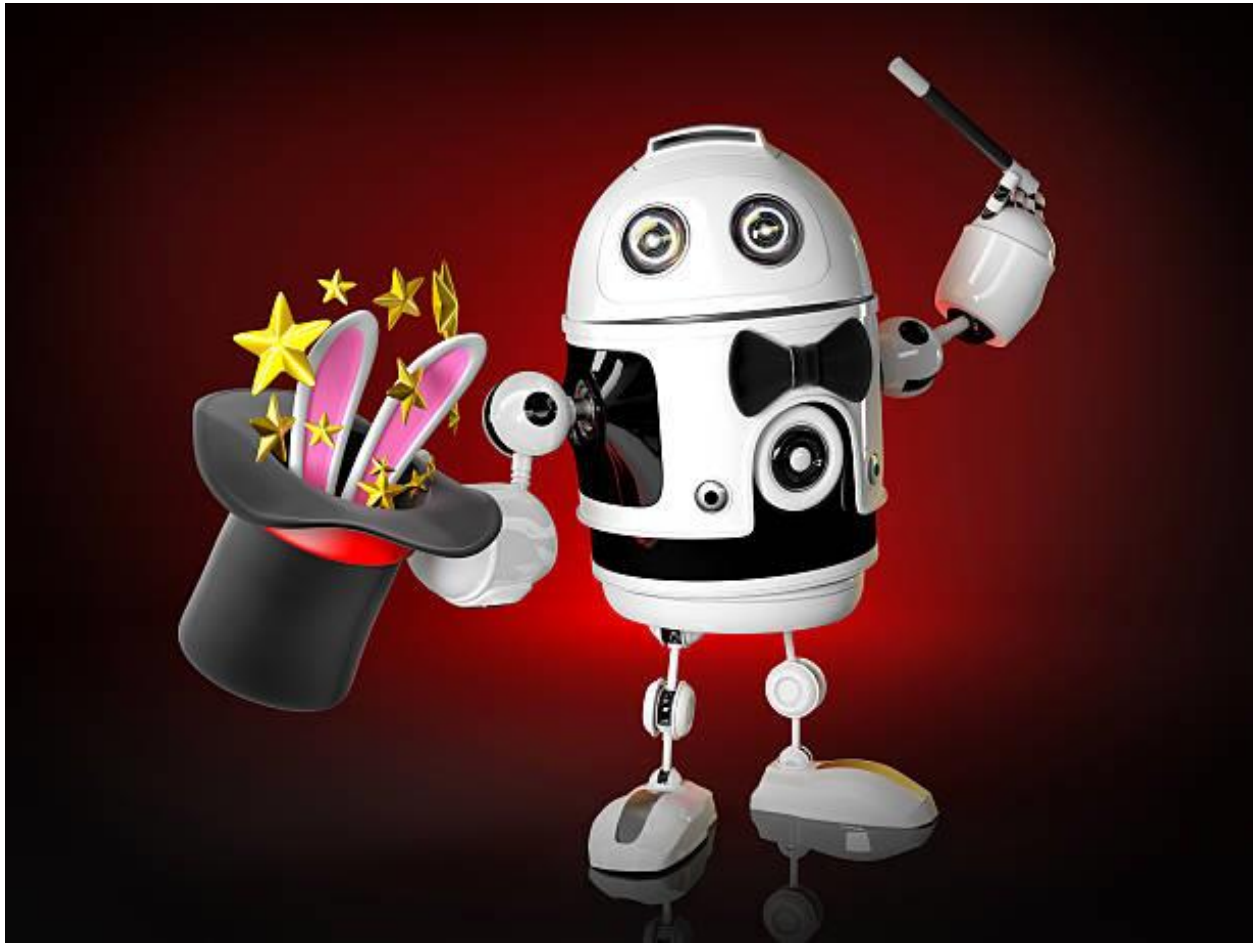
APPREHENSIONS?



SKEPTICISM



OBSCURITY



EXPLAINABILITY OF AI TO LEGAL PROFESSIONALS

AI Prediction

- AI prediction is crucial – faith on judicial system, careers, money, reputation (social image) etc. at stake.
- End users (Legal professions or the parties) are not likely to trust an absolute black box.

Explainability to legal professionals is vital!

NEURAL JUDGMENT PREDICTION IN ENGLISH[1]

Case ID: 001-148227 **Violated Articles:** Article 3 **Predicted Violation:** YES (0.97%)

1. The applicant was born in 1955 and lives in Kharkiv .

2 . On 5 May 2004 the applicant was arrested by four police officers on suspicion of bribe - taking .
The police officers took him to the Kharkiv Dzerzhynskyy District Police Station , where he was held overnight .
According to the applicant , the police officers beat him for several hours , forcing him to confess .

3. On 6 May 2004 the applicant was taken to the Kharkiv City Prosecutor 's Office . He complained of ill-treatment to a senior prosecutor from the above office . The prosecutor referred the applicant for a forensic medical examination .

4. On 7 May 2004 the applicant was diagnosed with concussion and admitted to hospital .

5. On 8 May 2004 the applicant underwent a forensic medical examination , which established that he had numerous bruises on his face , chest , legs and arms , as well as a damaged tooth .

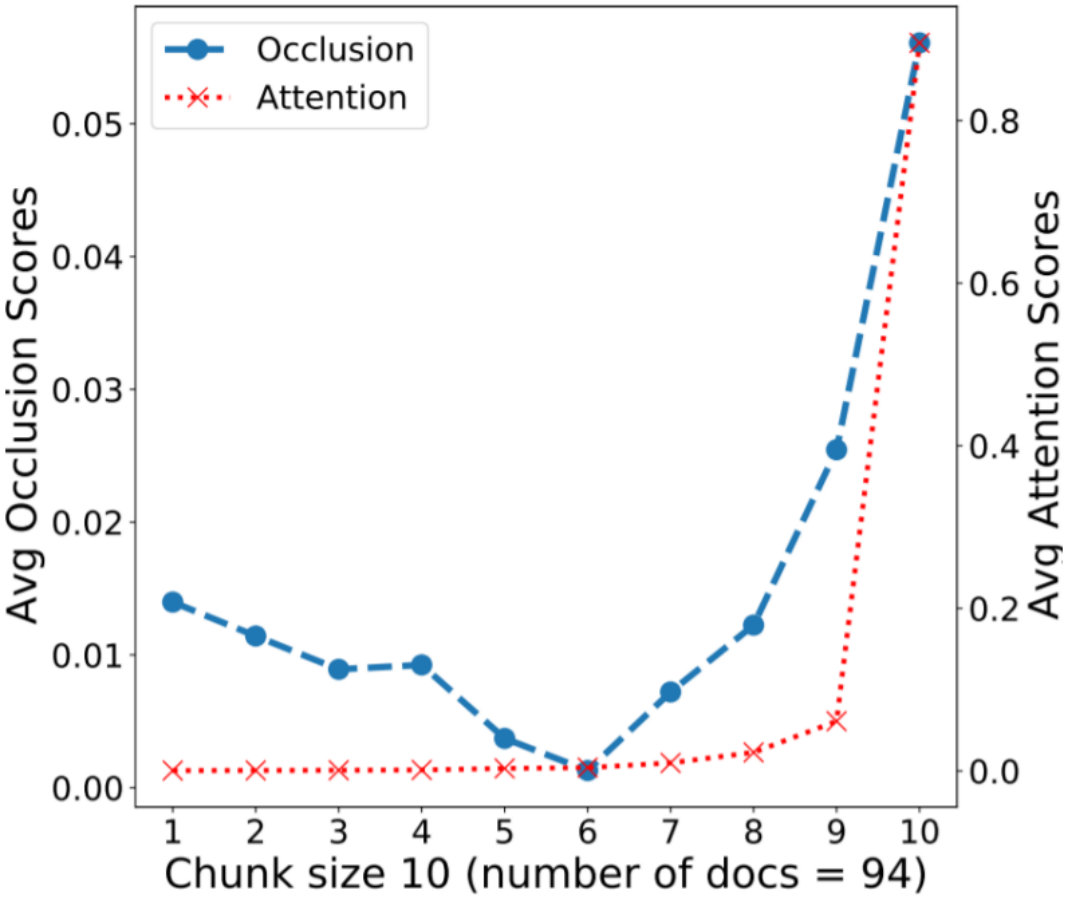
6. On 11 May 2004 criminal proceedings were instituted against the applicant on charges of bribe-taking . They were eventually terminated on 27 April 2007 for lack of corpus delicti .

7. On 2 June 2004 the applicant lodged another complaint of ill - treatment with the Kharkiv City Prosecutor 's Office .

[1] Neural Legal Judgment Prediction in English. Chalkidis et al. ACL 2019

Attention is not explaination!

ILDC FOR CJPE: INDIAN LEGAL DOCUMENTS CORPUS FOR COURT JUDGMENT PREDICTION AND EXPLANATION[2]



[2] ILDC for CJPE: Indian Legal Documents Corpus for Court Judgment Prediction and Explanation. Malik et al. ACL 2021

ILDC FOR CJPE: INDIAN LEGAL DOCUMENTS CORPUS FOR COURT JUDGMENT PREDICTION AND EXPLANATION

Metric	Explainability Model vs Experts				
	Expert				
	1	2	3	4	5
Jaccard Similarity	0.333	0.317	0.328	0.324	0.318
Overlap-Min	0.744	0.589	0.81	0.834	0.617
Overlap-Max	0.39	0.414	0.36	0.35	0.401
ROUGE-1	0.444	0.517	0.401	0.391	0.501
ROUGE-2	0.303	0.295	0.296	0.297	0.294
ROUGE-L	0.439	0.407	0.423	0.444	0.407
BLEU	0.16	0.28	0.099	0.093	0.248
Meteor	0.22	0.3	0.18	0.177	0.279

Machine explanations v/s Expert explanations

Low agreement between machine-generated explanation and human-generated explanation

PEEKING INTO THE BLACK BOX

USE CASE OF STATUTE IDENTIFICATION (AUTOMATIC SOLUTION)

Facts (Problem)

The appellants were tried for offences on the allegations that about four months prior to the incident, some quarrel had taken place between the deceased P1 and the appellant No. 4 P2 in relation to raising of boundary wall. On 12 April 1983 at about 7.00 P.M. the deceased accompanied by P3, had gone to P4 (PW-8) to engage some labourers for cutting crop in his field and while returning from the house of P4, when the deceased came near the L1 of appellant No. 4, he shouted to the remaining accused who were there that the deceased was their enemy and he should not be allowed to go and kill him. The appellant No. 6 P5 assaulted the deceased with the lathi on his head. When he fell down on the ground, all the appellants assaulted him. In spite of P3 intervening, the appellants did not stop assaulting. When they found that the deceased had died, the appellants dragged his body from the spot to place near L2 of appellant No. 1 P6. Thereafter, they ran away.

Statute (Solution)

Punishment for murder (IPC 302): Whoever commits murder shall be punished with death or imprisonment for life, and shall also be liable to fine.

USE CASE OF STATUTE IDENTIFICATION (AUTOMATIC SOLUTION WITH EXPLAINATION)

Facts (Problem)

The appellants were tried for offences on the allegations that about four months prior to the incident, some quarrel had taken place between the deceased P1 and the appellant No. 4 P2 in relation to raising of boundary wall. On 12 April 1983 at about 7.00 P.M. the deceased accompanied by P3, had gone to P4 (PW-8) to engage some labourers for cutting crop in his field and while returning from the house of P4, when the deceased came near the L1 of appellant No. 4, he shouted to the remaining accused who were there that the deceased was their enemy and he should not be allowed to go and kill him. The appellant No. 6 P5 assaulted the deceased with the lathi on his head. When he fell down on the ground, all the appellants assaulted him. In spite of P3 intervening, the appellants did not stop assaulting. When they found that the deceased had died, the appellants dragged his body from the spot to place near L2 of appellant No. 1 P6. Thereafter, they ran away.

Statute (Solution)

Punishment for murder (IPC 302): Whoever commits murder shall be punished with death or imprisonment for life, and shall also be liable to fine.

EXPLAINABLE AI & LAW FOR FUTURE APPLICATIONS

- Explanations are meaningful if they are generated by legal experts.
- An AI-Law system/algorithm is likely to be more acceptable if it is capable of generating explanations like legal experts, along with the correct prediction/recommendation of the associated task (judgement, statute etc.).

On Retrieval-enhanced Machine Learning (REML)

INTRODUCTION TO REML

Retrieval-Enhanced Machine Learning

Hamed Zamani*
University of Massachusetts Amherst
zamani@cs.umass.edu

Fernando Diaz*
Google Research
diazf@acm.org

Mostafa Dehghani
Google Research
dehghani@google.com

Donald Metzler
Google Research
metzler@google.com

Michael Bendersky
Google Research
bemike@google.com

ABSTRACT

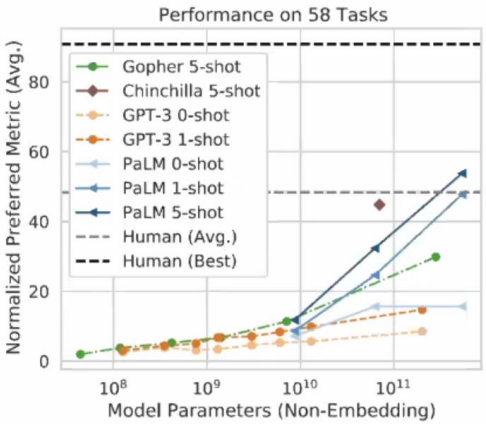
Although information access systems have long supported *people* in accomplishing a wide range of tasks, we propose broadening the scope of users of information access systems to include task-driven *machines*, such as machine learning models. In this way, the core principles of indexing, representation, retrieval, and ranking can be applied and extended to substantially improve model generalization, scalability, robustness, and interpretability. We describe a generic retrieval-enhanced machine learning (REML) framework, which includes a number of existing models as special cases. REML challenges information retrieval conventions, presenting opportunities for novel advances in core areas, including optimization. The REML research agenda lays a foundation for a new style of information access research and paves a path towards advancing machine learning and artificial intelligence.

CCS CONCEPTS

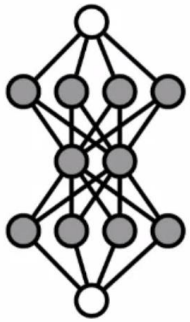
from 94 million in ELMo [45] to 1.6 trillion in Switch Transformers [13], an over 16× increase in just three years (2018 – 2021). Despite these successes, improving performance by increasing the number of model parameters can incur significant cost and limit access to a handful of organizations that have the resources to train them [4]. As such, focusing model development on the number of parameters is neither scalable nor sustainable in the long run.

Motivated by recent work demonstrating both that high capacity models memorize training data [6] and that using retrieval-style methods can offload memorization to storage [5], we propose the augmenting ML models with access to stored information through information retrieval (IR) techniques. Whereas IR has proven an effective tool to support people accessing large text corpora, we believe that IR can be extended to support machines accessing not just large text corpora but more abstractly-represented knowledge stores. By designing machine learning architectures that have explicit access to an information retrieval system, we can decouple

INTRODUCTION TO REML

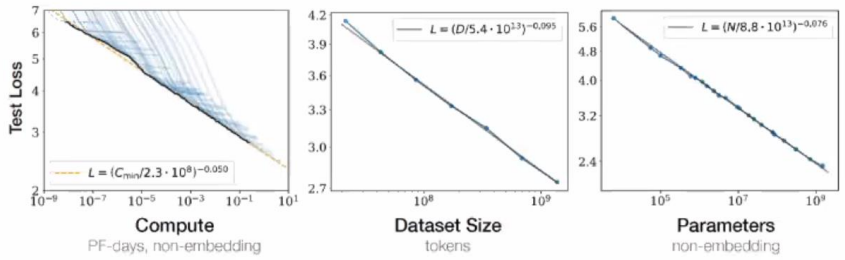


trained model

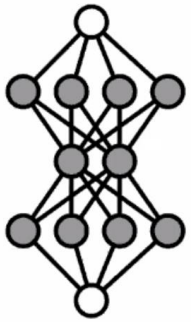


A Chowdhery, S Narang, J Devlin *et al*. PALM: Scaling Language Modeling with Pathways. 2022.

INTRODUCTION TO REML

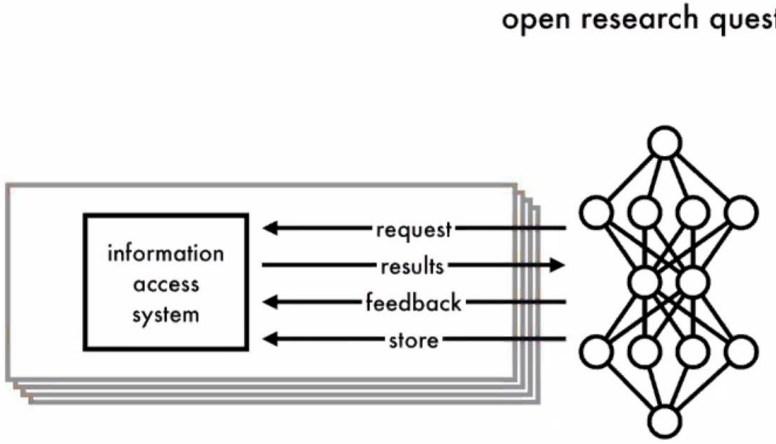


trained model



J Kaplan, S McCandlish, T Henighan, T Brown, B Chess, R Child, S Gray, A Radford, J Wu, D Amodei. Scaling laws for neural language models. 2020.

INTRODUCTION TO REML



open research questions

evaluation

- intrinsic vs extrinsic evaluation
- generalization in presence of memorization
- ensuring robust performance across multiple models (accessing a single IA system)
- development of efficiency and other cost metrics

H. Zamani, F. Diaz, M. Dehghani, D. Metzler, M. Bendersky, Retrieval Enhanced Machine Learning, SIGIR 2022.

EXPLAINABLE ARTIFICIAL INTELLIGENCE (XAI)

Role for Machine Learning

- Increasing concerns today about the explainability of technical approaches
  - Some conferences / workshops won't accept research reports that rely on "black box" technologies, where factors in the outcomes can't be explained or are not transparent
- Ability to tweak a system in a consistent and principled manner important
  - If error analysis can't be translated directly into remedies and performance improvement, how useful is the technology?
- Ethical concerns exist concerning inscrutable machine learning techniques
  - If we don't know what factors the machine is basing decisions on, how do we know it is unbiased, fair, and non-discriminatory?
  - The "social right to explanation" becoming increasingly expected
- Ultimately, systems that make life-saving or life-threatening decisions should be interpretable
  - E.g., self-driving cars

REUTERS/Jason Lee

Legal IR & NLP: Future Directions

Jack G. Conrad, TR Labs
Kripabandhu Ghosh, IISER Kolkata

Thank you. Questions?

ECIR 2023, 6 April 2023
Dublin, Ireland