International Conference on AI and Law (ICAIL -17)

https://icail2019-cyberjustice.com/

Monday

Workshop on AI in Legal Practice

- Companies pitching their "AI" products
  - KIRA
  - Irosoft
  - BlueJay Legal

- Panel discussions/Talks on AI
  - AI in Law Firms: What lawyers want and what lawyers need - Eric Lavallée
  - What AI can and can't do? - Marc Lauritsen, Bart Verheij
  - Regulatory and ethical issues - Dominic Martin, Charles Morgan
  - Collaborative Matchmaking - Katie Atkinson

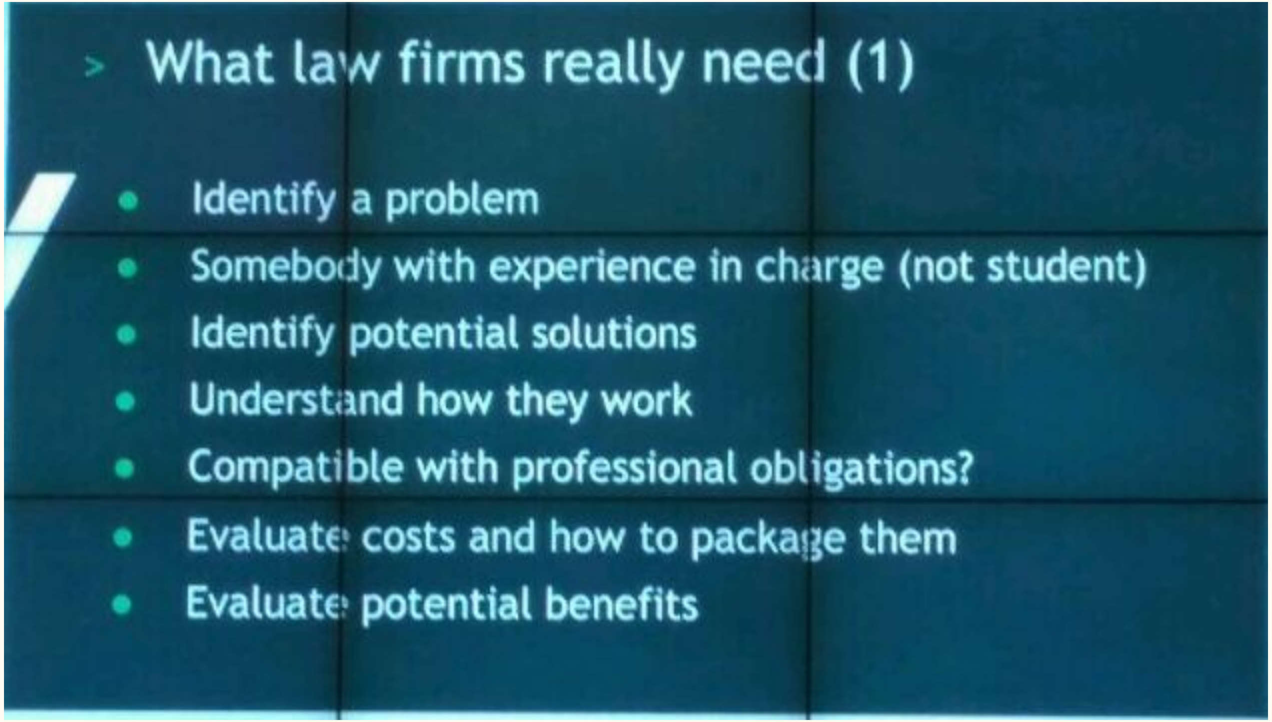
KIRA - Dr. Alexander Hudek

1. Publish - Patent - Make research a part of the job
2. Upto 50 employees - Research that can fail
3. Tech is just one part of building a product and UX exposes new constraints
4. Deep Learning - The Black Box Conundrum
    a. Users want explainability indicators to be predictable, match their gut instincts. They want to relate the indicator to what they know about the result. Result with a high confidence score which doesn't match their intuition causes lot of confusion/anger.
    b. It is not just about providing explainable insights about the underlying model (Even though the user says they wants to understand the model), but it is more about these insights showing a pattern that makes the most sense to the user.

BlueJay Legal

1. Accuracy explained
    a. What accuracy means
    b. What was the dataset
2. AI - Explain what it is and How good it is?!

AI in Law Firms: What lawyers want and what lawyers need



AI in Law Firms: What lawyers want and what lawyers need

Bad Implementations: the provider's fault

1. **AI driven case law search tool**
2. Sometimes older lower court decision is no longer relevant - Good Law project
3. Sometimes very good case law….. in the wrong jurisdiction
4. Not as user-friendly as advertised
5. Comment from a colleague: waste of time!

Question:

Q: How do you get Lawyers to construct NL questions/ How often would they want to do that?
A: Lawyers often don't want/need to construct natural language questions. They often want to ask questions to understand long decisions.

"Case Law AI should not just be to search within but it should be to understand the results - At document level / Result level"

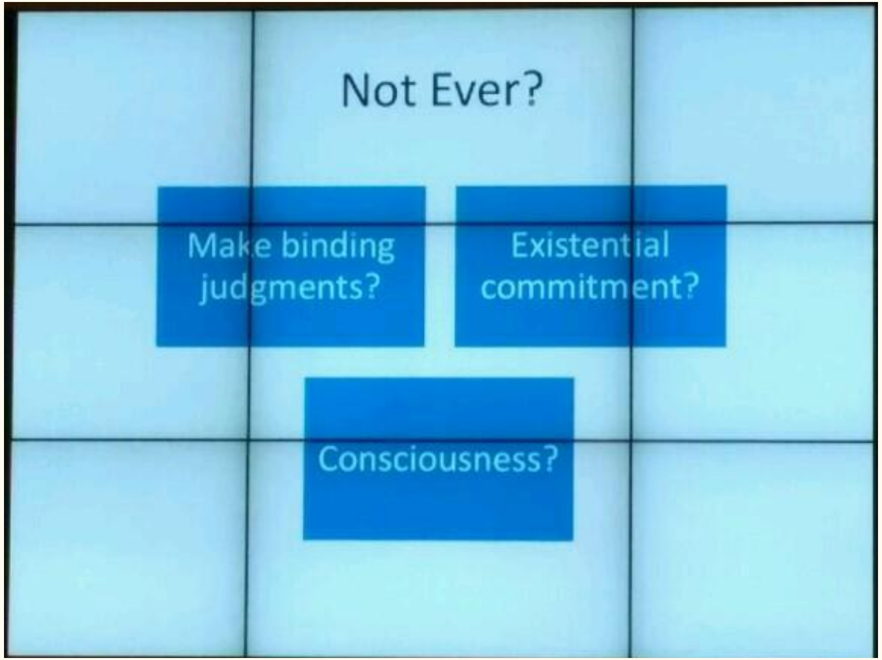
AI in Law Firms: What lawyers want and what lawyers need

Wishlist of Products:

1. Good Motion/Brief drafting tools - Voice Interfaces - Legal Context
2. Well formatted results
3. Differences in agreements
4. Draft questions - If AI search
5. Ability to quickly read long cases - CTRL F Steroids TL;DR

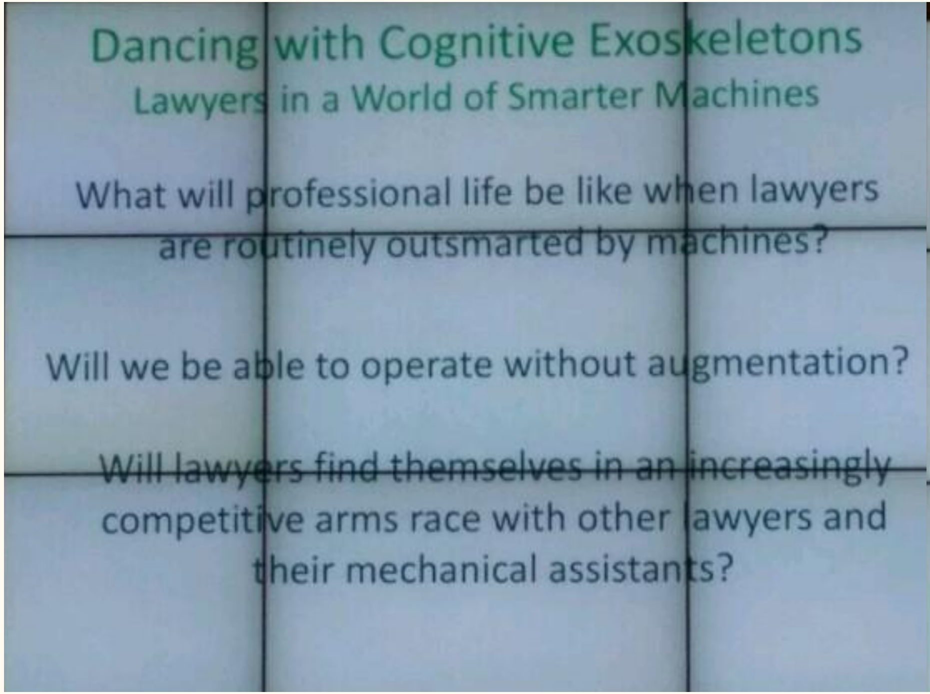
Needs someone with experience to be critical about the results

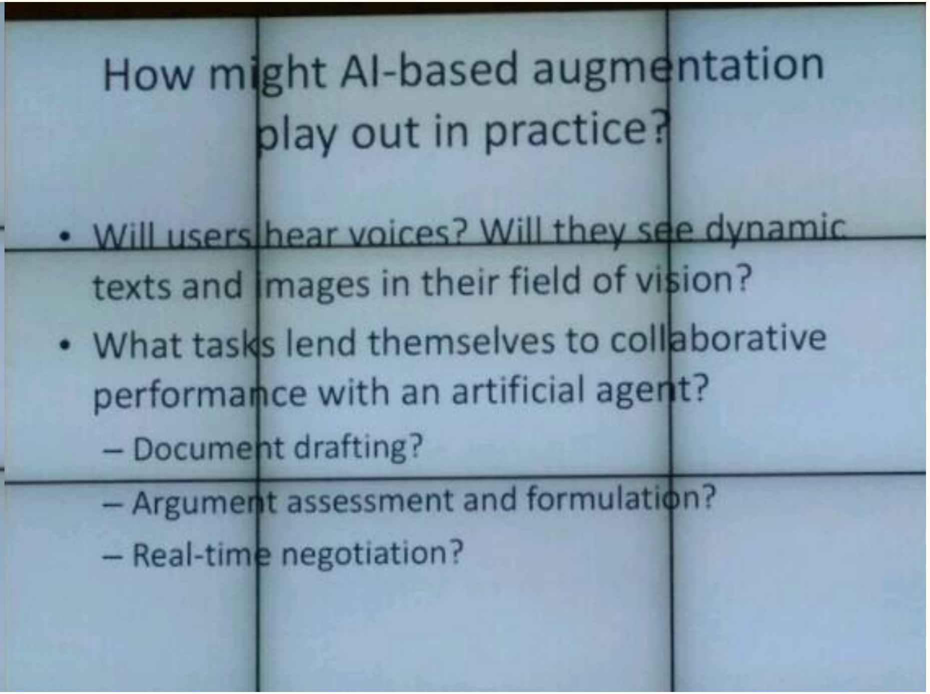
What AI can and cannot do currently



Many discussions revolved around AI autonomous Judges.

What AI can and cannot do currently





Regulatory and ethical issues

What are the key ethical issues?

1. Explainability
2. Transparency
3. Accuracy - Very imp for Law - Benchmark them to other tech alt, or human experts. It is very difficult to measure the accuracy of system wrt law
4. Discrimination / Bias

Regulatory and ethical issues

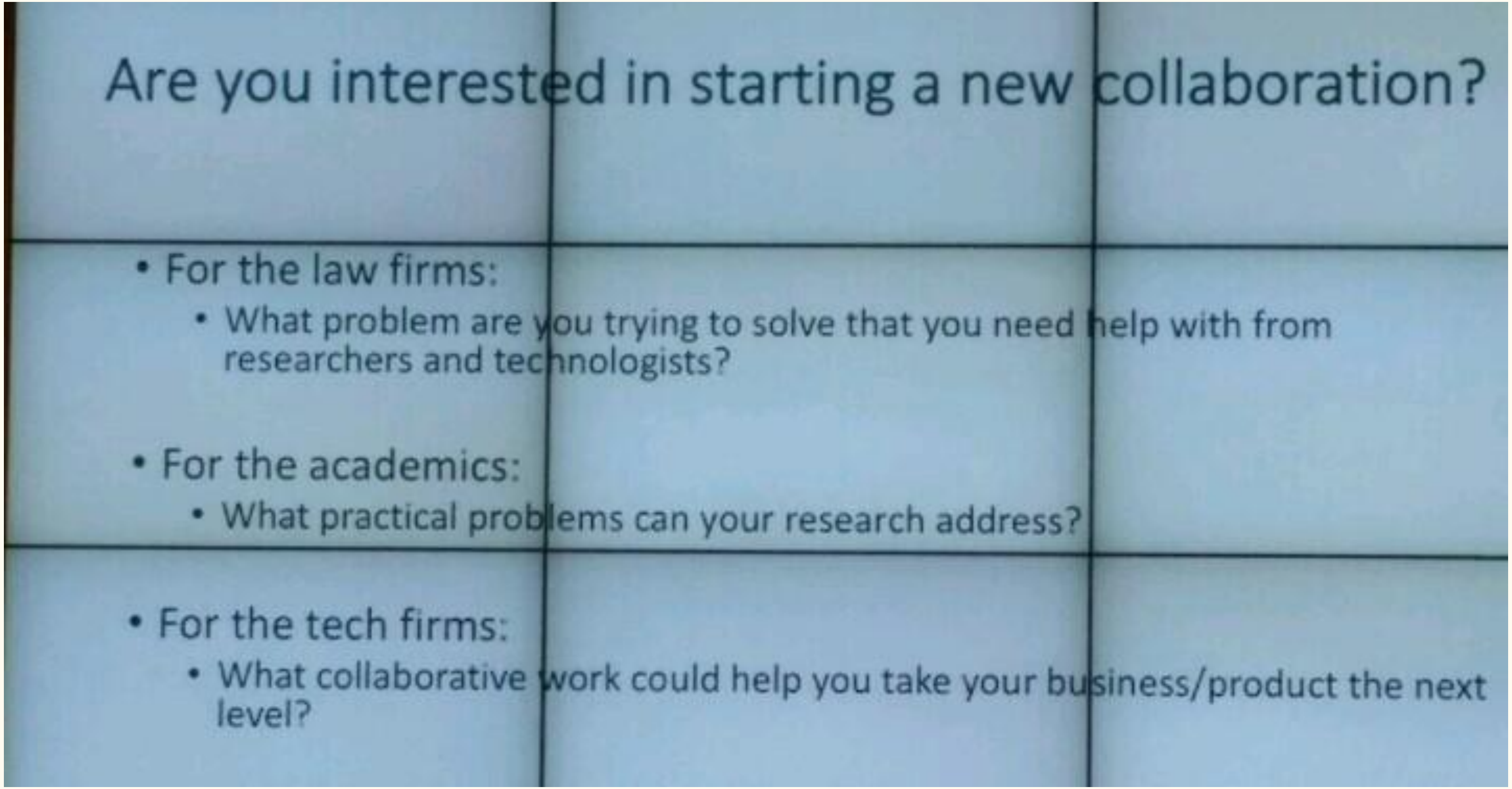
So far our response is appropriate to the changes

What is being done?

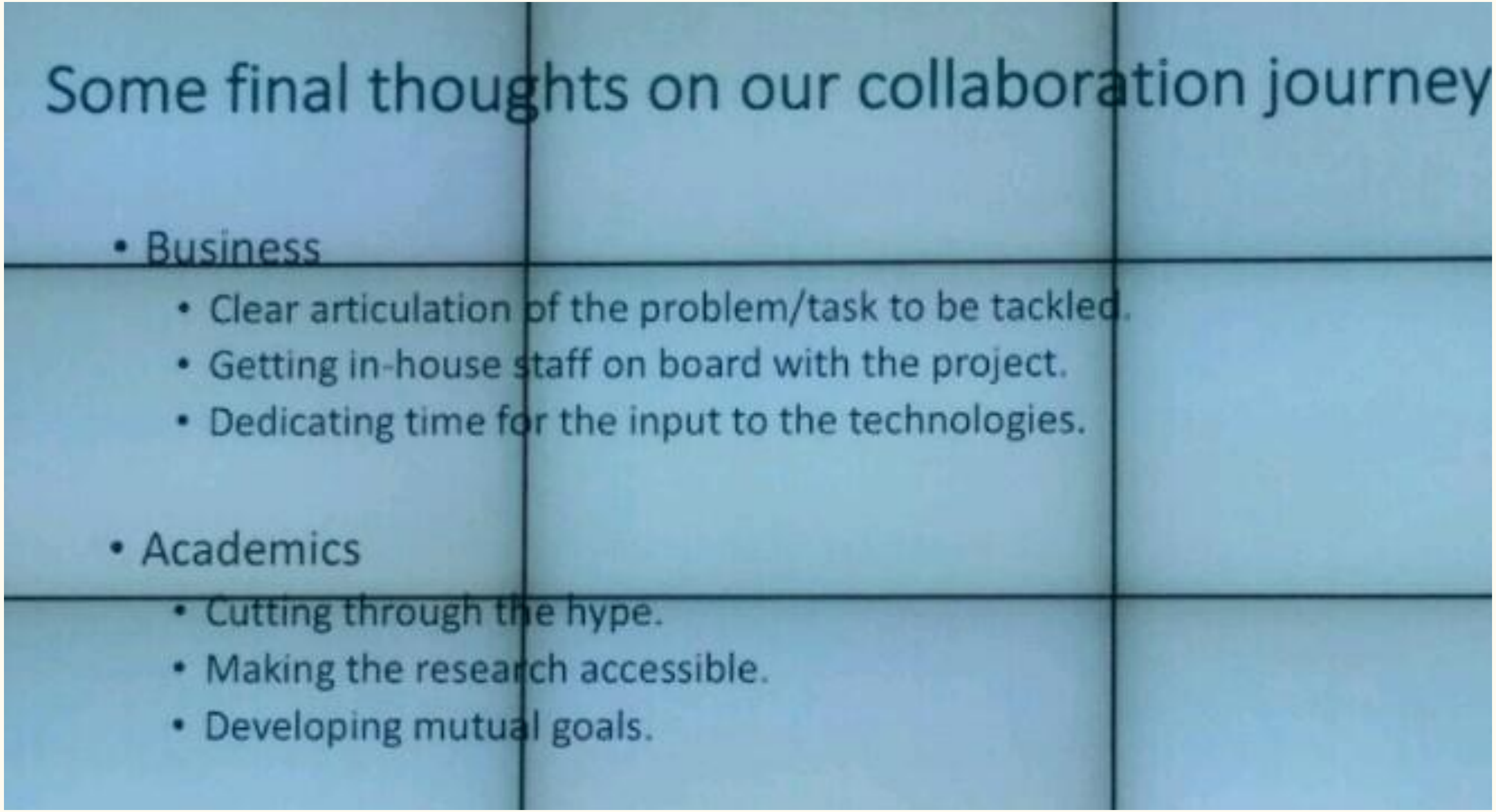
1. Fair information practices - every fairness law is based around it
2. Error/Data Analysis - Algorithmic Bias vs Data Bias
3. Bias is much more than just DS - AI - Labelling Bias?
4. Are humans better?

Devs should be aware of it and ack it

Collaborative Matchmaking



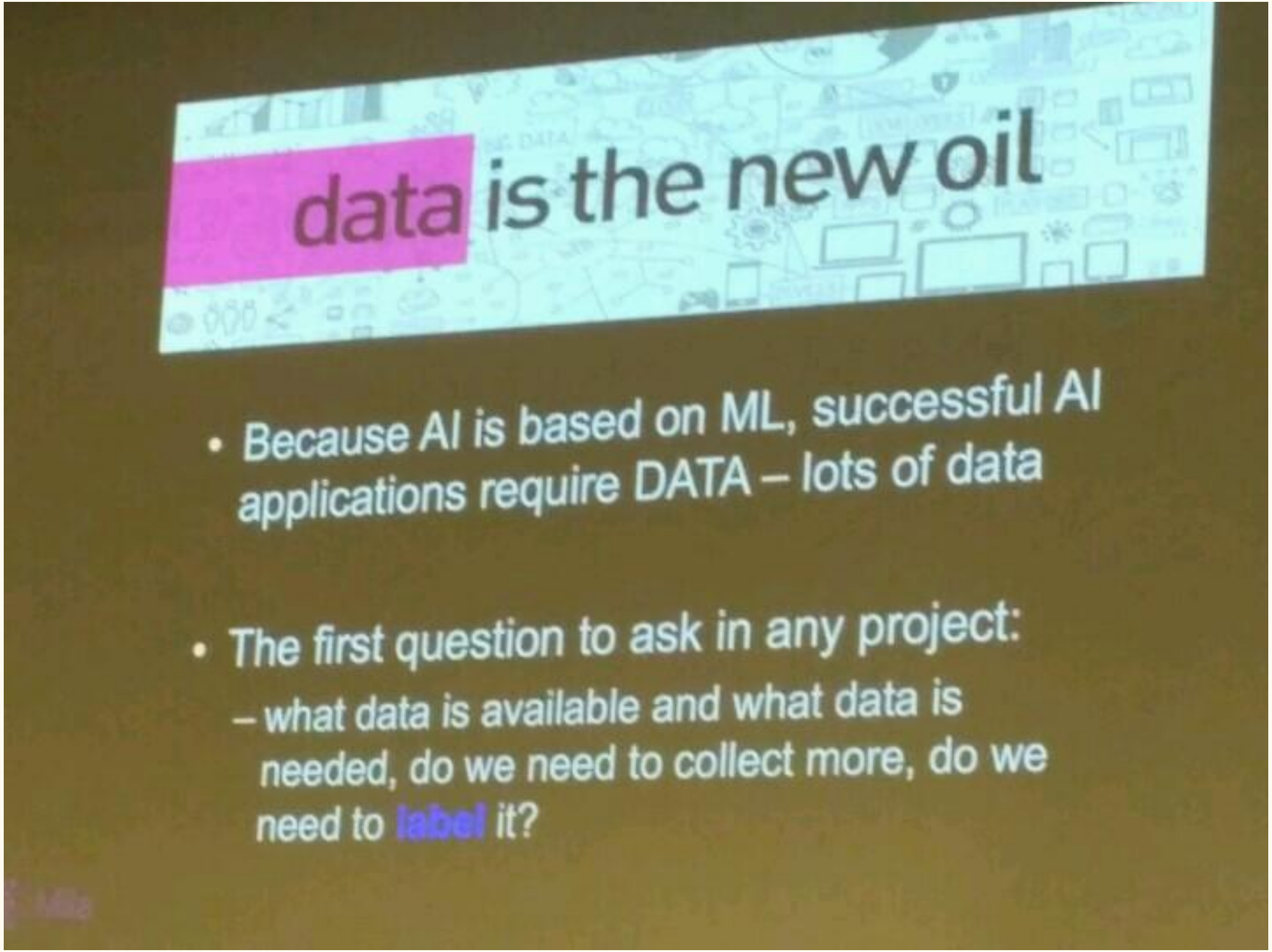
Collaborative Matchmaking



Tuesday

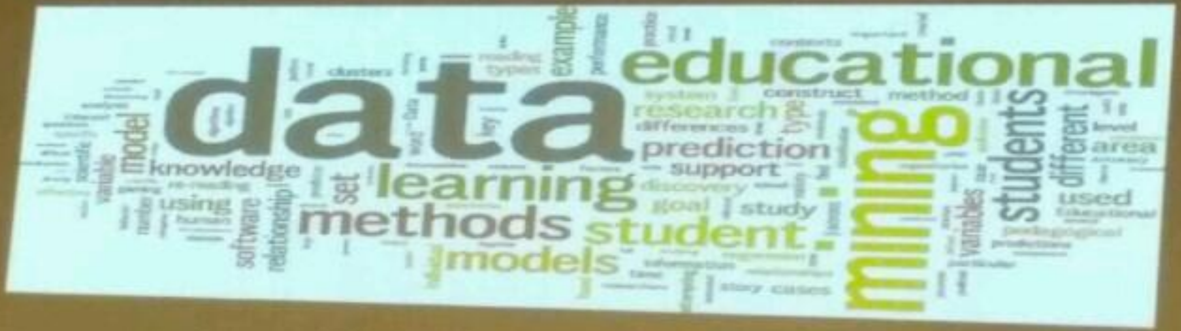
Bengio's talk

1. Intro to ML/AI/Deep Learning - Similar to Paul
2. CNNs
3. NLP
   a. Attention
   b. Neural Language Models
4. 0-Shot Generalization
5. Success shown so far is Supervised - Data Heavy - Short-Term Expectations
6. Machines extending humans' **Cognitive Power**


data is the new oil
• Because AI is based on ML, successful AI applications require DATA – lots of data
• The first question to ask in any project:
– what data is available and what data is needed, do we need to collect more, do we need to label it?

Entrepreneurs & AI

- Start thinking about how ML could be exploited to improve current services & products, and create new products and services
- Need a data strategy
- Need talent (that is tough), but smart engineers can learn the skills if given time and resources
- Connect with academic researchers and take advantage of training events & precompetitive research at institutes like MILA



Examples of Applications

- Recommendation systems, search engines, document classification
- Use computer vision in robots, factories, drones, agriculture
- Robotics, automating repetitive office & factory jobs
- Dialogue systems, personal assistants, customer service
- Design new materials, new drugs
- Diagnose and identify medical problems, predict outcomes, personalize optimized treatments
- Speed up computer graphics & other expensive simulations (climate, molecules, physics and chemistry in general)
- Generate personalized environments, games, music, text…

Dangers and Concerns with AI

- Big Brother and killer robots
- Misery for jobless people, at least in transition
- Manipulation from advertising and social media
- Reinforcement of social biases and discrimination
- Increased inequality and power concentration in few companies

Mila

Exploiting Search Logs to Aid in Training and Automating Infrastructure for Question Answering in Professional Domains

Implicit Relevance Feedback - Thomson Reuters

USE CASES

IRF useful for several important, diverse applications

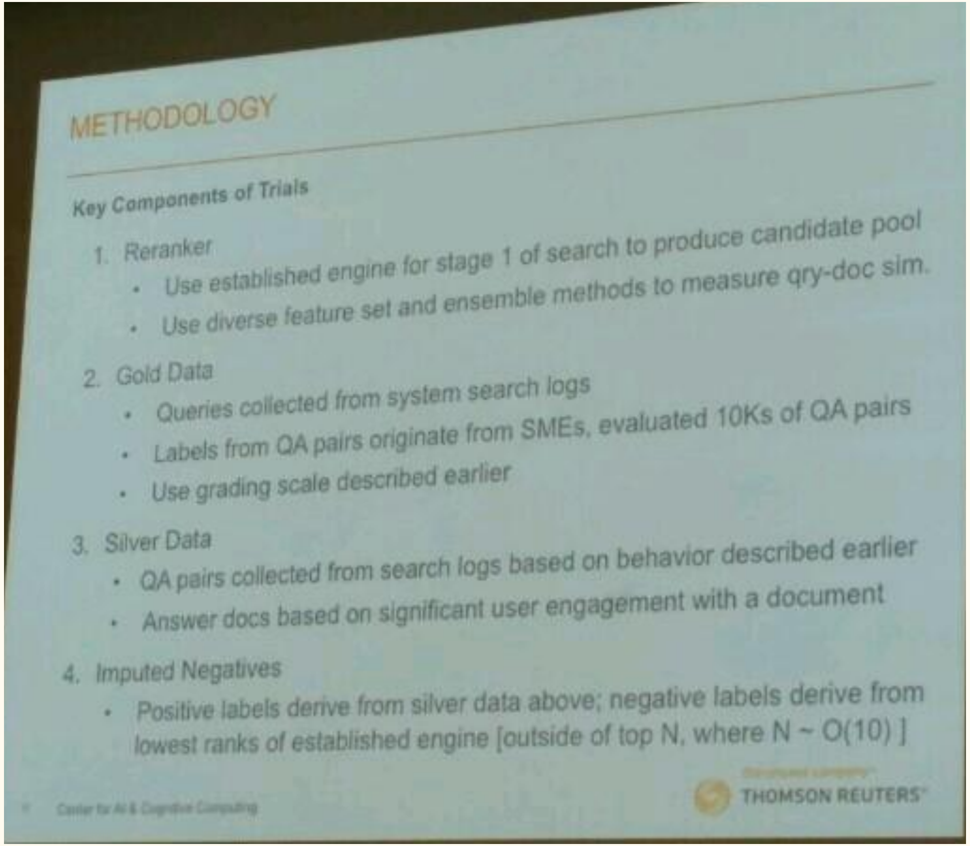
- Cold Start Problem – newly launched search or QA system w/o training data
- Automated Infrastructure – large-scale operational search or QA systems in competitive settings where time-to-market is increasingly imperative
  - Enables continuous integration, delivery, deployment
  - For production environments with system requirements including:
    - Cost-effective, on-demand scalability, flexibility, reliability all key
    - High performance systems need to reduce latencies
    - Grading of training data doesn't scale well here
    - Procurement of silver data a worthy alternative
      - No humans in the loop – fully automatable
- Auto-suggest Curation – user validated results available for auto-complete resources

THOMSON REUTERS

Implicit Relevance Feedback - Thomson Reuters

1. Can use heightened levels of interaction with content to infer relevance
   a. Just viewing doesn’t count
   b. Print/Save/Email/Export/Flag

2. User Activity Logs (UAL)
   a. Domain-specific preferences
   b. Practice Area Preferences
   c. Jurisdictional Preferences
   d. Topical Preferences



Grade Distribution for Silver Data

1. #Queries 125
2. 90% of the queries contain an answer (A&C)
3. Nearly 2/3 As
4. 10% or less non-answers
   a. Obviously not reflective of real world!!
5. Requires no human curation to obtain or apply

We need reliable clean user interactions for saving/relying on similar information.

IRF - Data

Sample Questions - Legal & Regulatory Domain

1. Who has the burden of proof for contempt
2. What are the elements of a criminal trespass
3. What is a certificate of appealability

Implicit Relevance Feedback - Thomson Reuters

1. A
   a. Fully and completely answers the question, flows fluidly as a response
   b. Could immediately appear as an answer for users
2. C
   a. Basically answers the question, but doesn't flow fluidly, not as good as A
   b. Answer may be found amidst extraneous material
   c. Exceptions or circumstances may be presented that are not addressed in the question
3. D
   a. Doesn't answer the question, but is related to the question
   b. Users would understand why it may appear in the result set
4. F
   a. Doesn't answer the question and is unrelated to the issue
   b. Would be a serious deficiency to include in a user's result set

Implicit Relevance Feedback - Thomson Reuters

Principal Research Questions

1. Does silver data provide a useful starting point in the absence of gold data? [Cold Start]
2. How much gold data is needed to reach a performance level comparable to silver data?
3. Does adding silver data to graduated amounts of gold data still add value?
4. Ablation studies based on negatives and practice area

Implicit Relevance Feedback - Thomson Reuters

Principal Research Questions - Answers

1. Does silver data provide a useful starting point in the absence of gold data? [Cold Start]
   a. Yes - Even in the ablation studies
2. How much gold data is needed to reach a performance level comparable to silver data?
   a. Answer was ablation specific
3. Does adding silver data to graduated amounts of gold data still add value?
   a. Yes
4. Ablation studies based on negatives and practice area
   a. Impact question 2 & 3, Negatives presence helps

Implicit Relevance Feedback - Thomson Reuters

Future Work

1. Contribution of silver data when harnessing SoA - BERT
2. More sophisticated imputation strategies from query logs
3. Continuous integration of implicit feedback from interacting users
4. Integration of user-profile modelling and recommendation systems
5. Multi-turn conversational systems

Synthetic Minority Over-sampling Technique (SMOTE)

Unbalanced Classes - Predicting one class all the time good accuracy

There are 4 ways of addressing class imbalance problems like these:

1. Synthesis of new minority class instances
2. Over-sampling of minority class - SMOTE
3. Under-sampling of majority class
4. Tweak the cost function to make misclassification of minority instances more important than misclassification of majority instances

Extracting the Gist of Chinese Judgments of the SC

Extractive Summary - Pick sentences from the original text - No paraphrasing

Chinese SC decisions - Reasoning Section - How the judges reached their decisions

Dice Coefficient - Divide Longest Common Subsequence by the average length of the segments - http://www.algomation.com/algorithm/sorensen-dice-string-similarity

Extracting the Gist of Chinese Judgments of the SC

Features

1. Quantitative
   a. # Characters, # Words, # Unique Words
   b. Absolute and Relative positions of the segment
   c. Length of the Reasoning section and Number of segments
2. Nature of Judgement - Criminal/Civil, Category - Specific words
3. Presence of Legal terms - Statute number, segments related to argument structure (ToA)
4. Embeddings - Average and Concat - Segment level and Document level
5. PoS Tags
6. Word Embeddings for opening words

Extracting the Gist of Chinese Judgments of the SC

Classifiers

1. GBT - Gradient Boosted Trees
2. Deep Learning - LSTMs BiLSTMs
3. Ensembles

Different feature combinations

Different contextual PoS combinations

Improving Sentence Retrieval from Case Law for Statutory Interpretation

1. There are sentences in case law decisions which help understand complex statutory language
    a. E.g. “Enterprise” means the related activities performed…. For a ***common business purpose***
2. Types of sentences
    a. Definitional Sentences
    b. Sentences that state explicitly in a different way what the statutory phrase means or state what it doesn’t mean
    c. Sentences that provide an example, instance, or counter-example of the phrase
    d. Sentences that show how a court determines whether something is such an example, instance or counterexample
3. Paraphrasing or Quoting are less helpful

Improving Sentence Retrieval from Case Law for Statutory Interpretation

1. BM25- TF-IDF - Cosine Similarity for KW based sentence retrieval
2. Smoothing with context - Get nearby sentences and weighted similarity
3. Query expansion - Neighbors to the statutory phrase based on Word2Vec - Similarity to sentences and cases
4. Novelty - How much new information is provided. Word overlap/ Proportion of overlap, Word Mover's Distance
5. Ensembles

Small datasets - So result discussion might not be relevant

Novelty related features aren't well suited for small sentences.

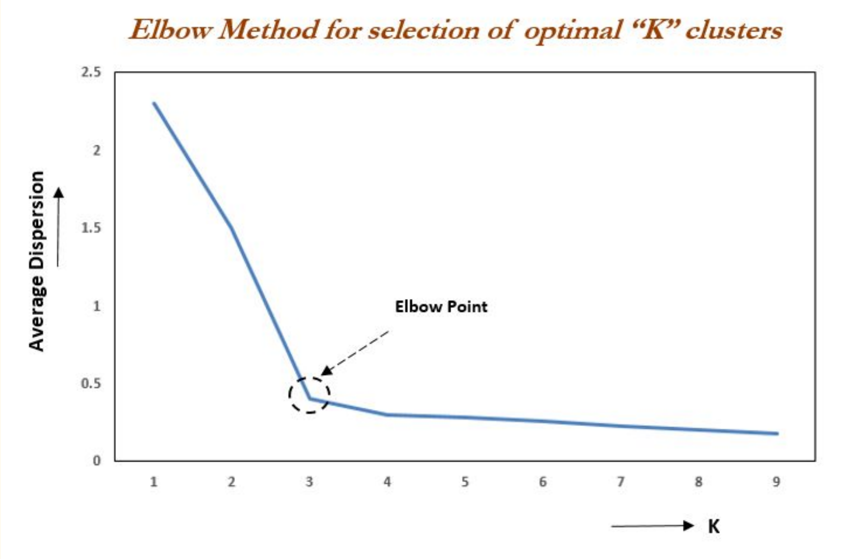
Wednesday

Automated Bundled Pagination

1. Paginate unstructured legal documents according to subjects or topics
2. Supervised topic classification and Unsupervised Clustering
3. Each page is tagged with reference data and a short summary
4. Six types of feature vectors - to represent the bundle
   a. BoW - Normalized BoW - TF-IDF Bow
   b. LDA on all the above
5. Preprocessing - Stop words, Stem, Punctuation, OCR clean up, Pruning
6. BoW - Frequency Histogram of selected n-grams (n - length n vector)
7. Gibbs Sampling based approach
8. LDA - Topic Modelling - LDA

Automated Bundled Pagination

1. Unsupervised - k Means Clustering
2. Choosing k?
   a. Elbow Method
   b. [Avg. Silhouette Analysis]
3. Elbow Method -WCSS
4. Silhouette Coefficient - Ensures that intra cluster variance is better than inter cluster similarities
5. How to get topic labels:
   a. Select popular words and use pre-built topic modelling to get the topic



Automatic Construction of a Polish Legal Dictionary

1. Map non legal terms (extra-legal terms) to legal terms
2. Dictionaries from generic docs and legal docs
3. Plain Text - Sentences/Tokens - Lemmatization - PoS
4. LLR - (Log Likelihood Ratio) Figure out co-occurring bi and tri grams
5. GloVe - word2vec
6. Process
    a. Pick an extra-legal term
    b. Most similar legal term using cosine sim - recommend

Thursday

COLIEE

1. Three Tasks
    a. Retrieval
    b. Case Law Entailment
    c. Statutory Entailment
2. Retrieval
    a. Summarization and Lexical Matching - use COLIEE 2018 data
        i. Latent Summary Feature Representation
        ii. Lexical Features
        iii. Explore different parts of the question to different documents
        iv. Using the summaries given in the dataset - didn't work
        v. Phrase Scoring based Summary Encoding - Summary phrases get higher score
        vi. Generate missing summaries using this model
        vii. Latent Representations - Combine Sentence and Document level pooling
        viii. Ngram, Skip Gram overlap + SVM

COLIEE - Statute IR and Entailment

1. TF-IDF & Language Modelling (LM) based retrieval
2. Inter-paragraph Entailment - Structure Analysis - Detect Negation
3. SVM - Heuristics

“… a common suggestion for coming up with good queries is to think of words that would likely appear in a relevant document and to use those words as the query”

LM - A document/(statute) is a good match if it is likely to generate the query

Performance LM $<$ TF-IDF

COLIEE- Entailment

1. Entailment
    a. Given two sentences, can one be inferred from the other?
    b. Can be used for QA systems or Summarization
2. Recommend reading the first two sections of this paper for ideas
3. Binary Classification
4. Multi-word tokens similarity - stemmed ngram BoW - Cosine distance
5. Noun Phrase token similarity
6. BERT confidence scores (USE and ULMFit)
7. Combine these scores

Legal and Ethical Issues of AI and Law

1. Human in loop - Contestability
    a. Risk related education
    b. Under what conditions does the prediction hold good
    c. What is needed to falsify
2. Interpretable Predictions - https://github.com/marcotcr/lime
3. AI as a Legal Person - Legal Personhood for AI systems
4. Ethical Matrix - Who respects what?
    a. Parties: Judges, Lawyers, Litigants, Legal Tech companies, Law Research & Knowledge production
    b. Values: Justice, Equality, Privacy, Competency, Transparency, Usefulness

Automatic Summarization of Legal Decisions Using Iterative Masking of Predictive Sentences*

1. Construct summaries for Board of Veterans' Appeals to help decide the outcome
2. Summarization
    a. Abstractive - Paraphrase
    b. **Generative Summarization** - Pick sentences directly - No paraphrasing
3. A good summary
    a. Include prediction sentences - Sentences that contribute or help predict the outcome, along with the outcome
    b. Express the important legal issue - facts - background - reasoning - evidence - plaintiff information
    c. Be narrative - tell a story
    d. Types of evidence
        i. Common sense experience in evaluating witness statements
        ii. Novel theories and causation paths

Automatic Summarization of Legal Decisions Using Iterative Masking of Predictive Sentences

1. Train-Attribute-Mask Pipeline - CNN - Random Forest - USE
2. Regex for extracting pattern sentences
3. MMR (Maximum Marginal Relevance) - Rouge-1, Rouge-2
4. Extensive Error Analysis
    a. Compare scores/metrics for different user annotated/generated summaries
    b. Per question analysis

Automatic Summarization of Legal Decisions Using Iterative Masking of Predictive Sentences

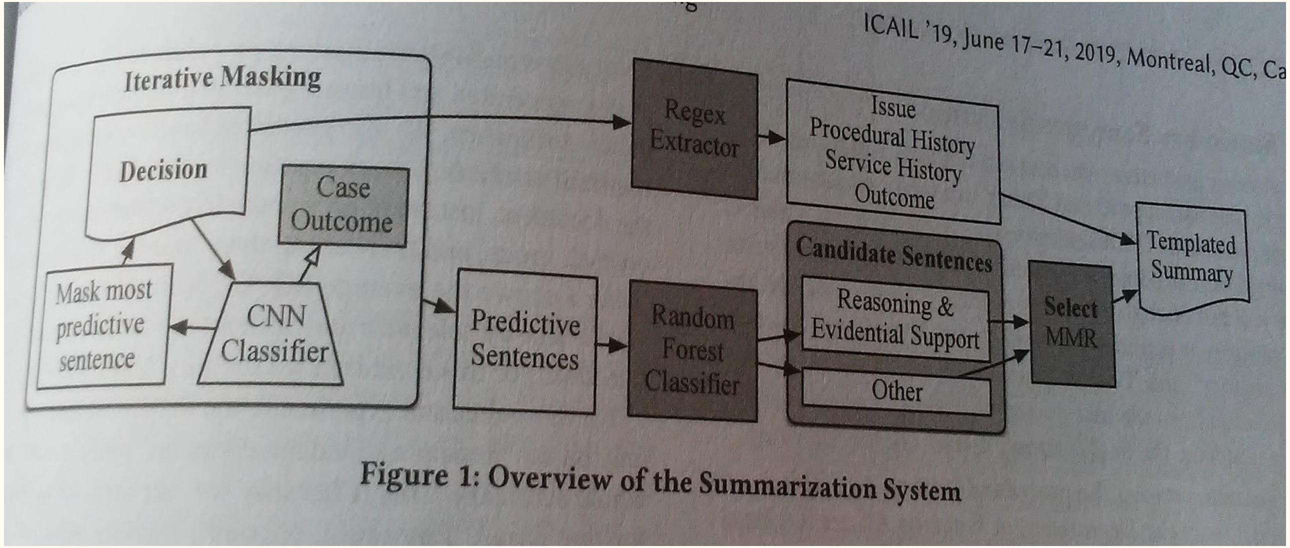


Figure 1: Overview of the Summarization System

Argument Identification in Public Comments from Erulemaking

1. Identify argumentative text at the sentence level - Classify - Determine the stance
2. Seed words to automatic annotation using rule-based model - CKY (context-free parsing algorithm)
3. BoW + Logistic Regression
4. Usable taxonomy

Semi-Supervised Methods for Explainable Legal Prediction

Highlight text of the decision that explains the outcome/prediction

Paradigm	K-E artifacts	Execution-time representation	Output	Example
1. text → output			prediction, relevance weights	NFE
2. text → predicates → output	predicate set, rule set		prediction, per-predicate relevance weights	
3. text → features → output	feature set		case-based argumentation	
4. text → features→ predicates → output	feature set, predicate set, rule set		hybrid case/rule-based argumentation	SCALE
5. features → output	feature set	featural case representation	case-based argumentation	HYPO
6. features → output	rule set	featural case representation	rule-base argumentation	
7. features→ predicates → output	feature set, predicate set, rule set	featural case representation	hybrid case/rule-based argumentation	NIHL (Angelic methodology)

Semi-Supervised Methods for Explainable Legal Prediction

1. Attention based networks - SCALE (Semi-supervised Case Annotations)
2. Attention based highlight didn't help attorneys
3. Annotate sentences - Find nearest neighbors - XGBoost on labels predictions
4. Models to predict outcome - unclear

Machine Learning for Explaining and Ranking the Most Influential Matters of Law

1. Rank legal principle citations that support a motion
2. Graph for legal feature extraction
3. SHAP - SHapely Additive exPlanations
4. SHAP values is an estimator that measures the importance of each feature for the overall model prediction of a simple sample
5. XGBoost - Bagging to fix class imbalance
6. Didn’t show any visualizations for explanations

Friday

WestLaw Edge AI - KeyCite Overruling Risk, Litigation Analysis and WestSearch Plus

"WSP is a non-factoid QA system that provides legally correct, jurisdictionally relevant and conversationally responsive answers to user-entered questions in the legal domain."

1. KOR - Detect implicit transitive overruling - If 3 overrules 1 and 2 cites 1, does 3 overrule 2?
2. Litigation Analysis - Scrape dockets for all metadata - Manual review
   a. Focuses on motion and outcome detection
   b. Grammar - Map to queries?!
3. WSP - 200k QA - A: human generated summaries
   a. "Our system relies on NLP models targeted at the tokens, syntax, semantics, and discourse structure of legal language, in addition to other machine-learned models to classify QA intents, generate search queries based on those intents, identify named entities and legal concepts, and for classifying, ranking and thresholding and final answer candidates"

The Mix - 1

1. A Regularization Approach to Combining Keywords and Training Data in Technology-Assisted Review
    a. New method of regularization, that is not well explained - Too much math

2. Adaptive Covariate Shift for Legal AI - Covariate Shift - Training and Test samples follow different input distributions - Lexis
    a. Detect and Adapt
    b. Combine current monthly user queries to the training set - Classify query labels
    c. If classification is difficult then no covariate shift
    d. Incremental Learning - Label data for the new types of queries

The Mix - 2

1. Towards Computer-aided Analysis of Readability and Comprehensibility of Patient information in the Context of Clinical Research Projects
   a. Patients need to give **informed** consent for clinical projects - if they can't understand the implications of their consent, how is it informed?
   b. Help lay person understand the contract
   c. Detect usage of any technical terms (NER) - Detect explanations provided to disambiguate and explain the term (Regex)
   d. Readability metrics - Flesch-Reading-Ease, gSMOG, Weiner Sach-textformel
2. LitiLens - When provided a case number and court, find the case and match the profile of that case against our massive database of cases, provide invaluable insights - FreeLaw project data

The Mix - 3

1. Legal Case Based Reasoning System - Issue Based Prediction
    a. Hierarchical Tree - Root Legal Issue
    b. “Knock-out” factors relevant to the legal issue
    c. Database of decided legal cases specifying the outcome of that case relative to the root issue ad the factors which were deemed to be present in that case
2. Neural Attention learning for Legal Query Reformulation
    a. Reformulation as Neural Machine Translation (NMT): Error to Correct
        i. Suggestions & Expansion
        ii. Character Level embeddings - OOV
        iii. Vocab from headnotes

New Topics

1. Deontic/Normative Logic
   a. Obligatory - Permissible - Optional - Impermissible - Omissible
2. Bayesian Inference
3. Argument Mining

If you are still awake - Thank You