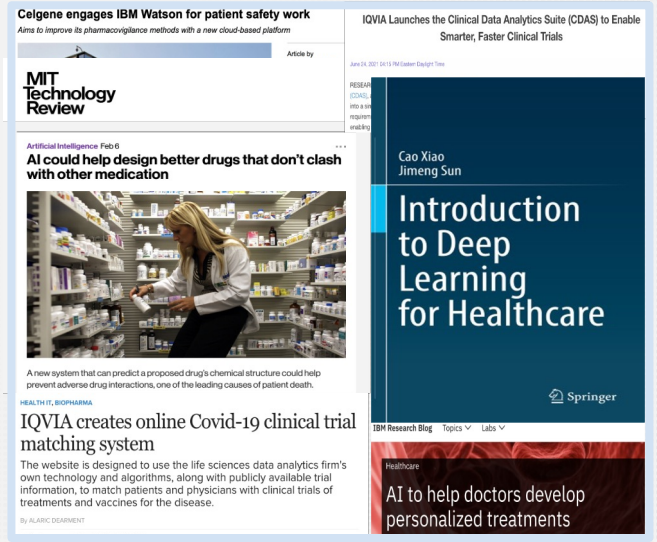
Machine Learning for Digital Transformation of Clinical Trials and Legal Review

Cao (Danica) Xiao
VP of AI Research
Relativity
Feb-27-2023

Cao (Danica) Xiao, Ph.D.

- 2022.8-now VP of AI Research@ **Relativity**
  AI/ML for legal tech
- 2021.5-2022.8 Sr.Director of ML@ **Amplitude**
  AI/ML for User Behavior Modeling
- 2019-2021.5 Global Head of ML@ **IQVIA**
  AI/ML for Drug Discovery and Development
- 2017-2019 Research Staff Member @ **IBM Research** & Principal Investigator @ **MIT-IBM Watson AI Lab**
  AI/ML for Healthcare and Medicine
- 2016.12 Ph.D. from **UW-Seattle**



Deep learning	Tensor decomposition	Graph neural networks	Natural language processing	Generative models

Agenda

1. Introduction
2. ML for Clinical Trials
3. ML for Legal Review
4. Trade-offs and Other Considerations

The Evolution of AI and Its Success in Industry

Success in industry

Virtual assistant	Machine translation	Preventive Maintenance	Drug discovery	Drug development
Self-driving car	Fraud detection	Question Answering	Disease diagnosis	Legal applications

Challenges of ML for Drug Development and Legal Tech

- **Data** is noisy and in multiple modalities, is often siloed and difficult to access due to its privacy concerns.
- **Users** are highly skilled domain experts who are reluctant to provide labor-intensive labeling.
- **Tasks** are decision-centric and thus demand high standards for model quality and interpretability.
- **Balance** between practical value and research impact.

How can we better design ML models and workflows that are tailored to the unique needs of these real-world applications and enable us to unlock the full potential of AI in these domains?

Agenda

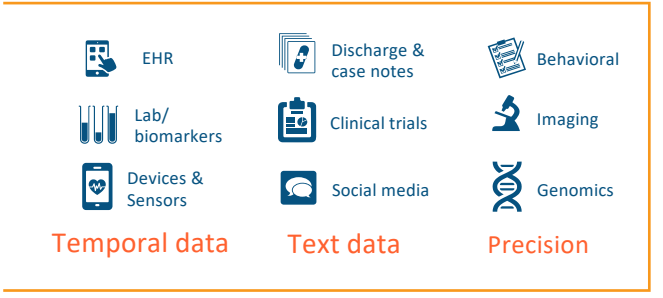
2 ML for Clinical Trials

The Clinical Trial Workflow and Its Challenges

Challenges: High Cost, Long Duration, and Low Success

	Phase 1	Phase 2	Phase 3
Time & Cost	1-2 years $225M	1-2 years $250M	2-3 years $250M
Recruitment	50-100 patients	Patients (hundreds)	Precision cohorts (thousands)
Task Design	Safety	Efficacy & Side Effect	Efficacy & Adverse reaction

Real-World Health Data



Key Tasks

Identify drug Candidates → Design trial protocols → Identify trial Sites/doctors → Find qualified Patients → Deploy within Hospitals

Machine Learning for Clinical Trial Workflow

Site selection (JAMIA 2021)
Doctor recommendation (AAAI 2020)
Fair site selection (in submission)

Patient-Trial Matching
(WWW'20, KDD'20)

Model privacy
(WWW' 21)

Identify drug Candidates → Design trial protocols → Identify trial Sites/doctors → Find qualified Patients → Deploy within Hospitals

Outcome prediction
(Cell Pattern 2022)

Language Model for Trials
(COLING' 22)

Trial protocol generation
(in submission)

Identify Rare Disease
(AAAI' 20)

Quantify Uncertainty
(WWW' 21 and AAAI' 22)

Data Quality
(KDD' 21)
(NeurIPS' 22)

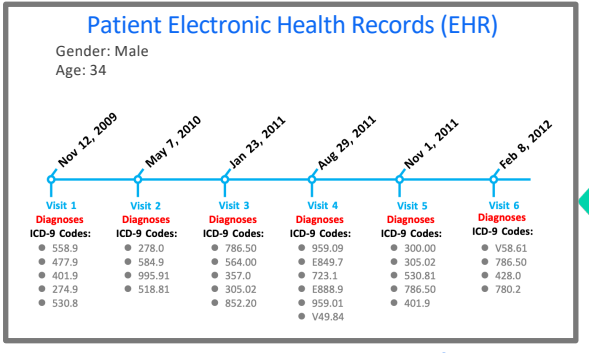
COMPOSE: Cross-Modal Pseudo-Siamese Network for Patient Trial Matching

Junyi Gao, Cao Xiao, Lucas Glass, Jimeng Sun

KDD 2020

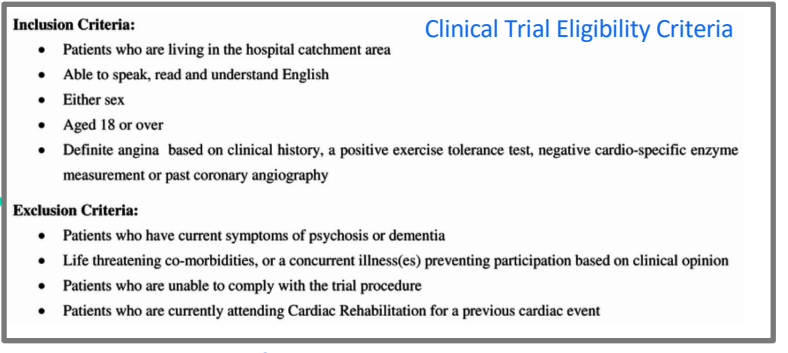
Patient-Trial Matching

Goal: to find right patients for clinical trials or find right clinical trials for patients.



Encode essential patient health information

Match?



Describe what type of patients are needed and excluded by the trial

Existing Data Driven Solutions

(1) Rule-based systems (expert annotated or via supervised learning);
(2) Combine ML with rules to generate SQL query for extracting criteria

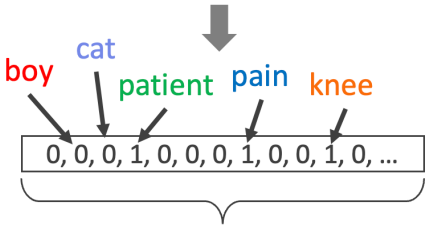
Health and Legal Data Encoding

Human language

Breaks in service may be regarded as a temporary cessation of service and not a break in contract. This includes where the employee is unable to work due to sickness or is absent due to a temporary cessation of work (section 212(1), ERA 1996).

Doctor Notes

"The patient reported a severe pain in the knee ."



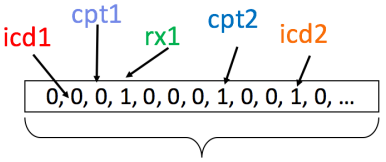
Number of unique words in the data

Machine readable codes

employee = {1, 0, 0, 0, 0,0,0}
cessation = {0, 1, 0, 0, 0,0,0}
temporary = {0, 0, 1, 0, 0,0,0}
absent = {0, 0, 0, 1, 0,0,0}
unable = {0, 0, 0, 0, 1,0,0}

One-hot encoding

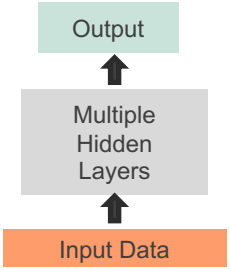
Medical Codes



Number of unique medical codes in the data

Learning with Deep Learning Models

Learning a simple binary classification task with multilayer perceptron (MLP)

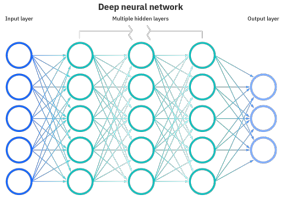


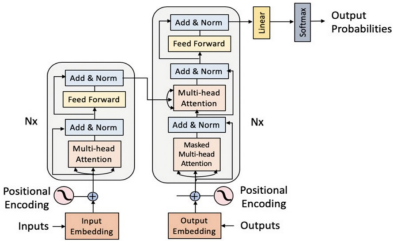
$\hat{y} = \sigma(\mathbf{w_o}^T\mathbf{h})$
(generate an outcome 0.0~1.0)

$\mathbf{h} = \sigma(\mathbf{W_h}^T\mathbf{x})$
(transform **x** for prediction)

x
(input medical code vectors)

This multiple levels of representation allows deep learning models to capture complex patterns and relationships in the data.





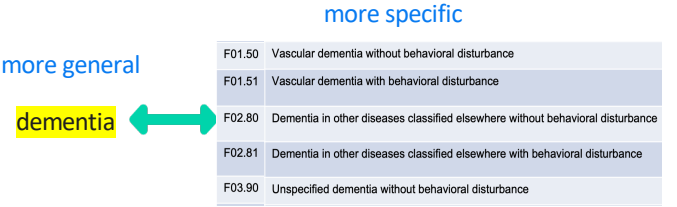
Learn model parameters using gradient descent with backpropagation.

New architectures were introduced to address challenges in pattern extraction and prediction.

Challenges of Patient-Trial Matching

1: Embed **different concept granularity.**

Observation: Trial text encode more general disease concept, while patient data use more specific medical codes.
ML Challenge: hard for embedding to connect concepts across different granularities.



2: Match **many-to-many relations.**

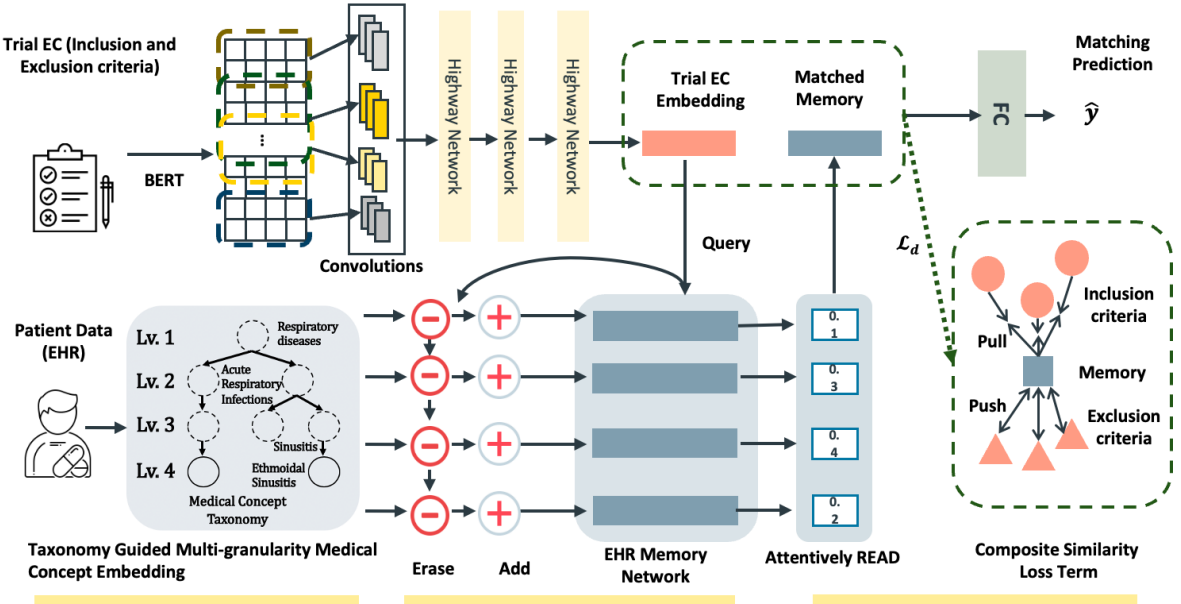
Observation: A patient may qualify for multiple trials; each trial can accept patients who share few similarity.

ML Challenge: Align the patient embedding to different trial embeddings may confuse the embedding function

3: Model **inclusion/exclusion criteria.**

Observation: Inclusion/Exclusion criteria describe desired and unwanted conditions from the targeted patients.
ML Challenge: cause patient trial mismatch if not properly handle the “negation”.

The COMPOSE Model for Patient-Trial Matching



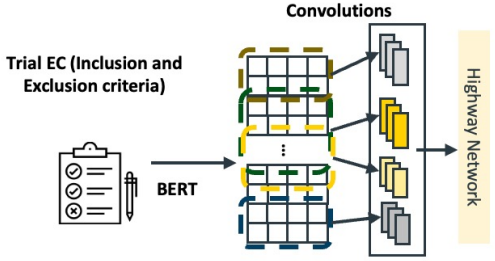
Embedding: multiple modality and concept granularity

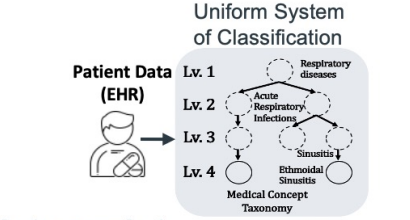
Matching: multi-to-multi alignment and dynamic matching

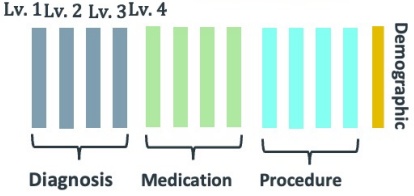
Constraints: explicitly handle inclusion/exclusion criteria

Junyi Gao, **Cao Xiao**, Lucas Glass, Jimeng Sun, COMPOSE: Cross-Modal Pseudo-Siamese Network for Patient Trial Matching, KDD 2020

Patient-Trial Matching: Data Embedding







Multi-Granularity Trial EC Text Embedding

Use BERT to learn contextual embeddings for each word of EC sentence $[w_1, \ldots, w_N]$

$$\widetilde{\boldsymbol{c}} = [\widetilde{\boldsymbol{w}}_1, \ldots, \widetilde{\boldsymbol{w}}_N] = BERT([\boldsymbol{w}_1, \ldots, \boldsymbol{w}_N])$$

Use different kernel sizes to capture different semantic granularity.

$$\boldsymbol{x} = [Conv(\widetilde{\boldsymbol{c}}, k_1), Conv(\widetilde{\boldsymbol{c}}, k_2), Conv(\widetilde{\boldsymbol{c}}, k_3), Conv(\widetilde{\boldsymbol{c}}, k_4)]$$

Use highway network and max pooling to obtain the final EC embedding

$$\boldsymbol{u} = \sigma(Conv(\boldsymbol{x}, k))$$

$$\boldsymbol{v} = \boldsymbol{u} \cdot Conv(\boldsymbol{x}, k) + \boldsymbol{x} \cdot (1 - \boldsymbol{u})$$

$$\boldsymbol{e} = MaxPool(\boldsymbol{v})$$

Taxonomy Guided Hierarchical Memory Network Based Patient Embedding

We use a Medical concept taxonomy system to save each concept into 4 levels.

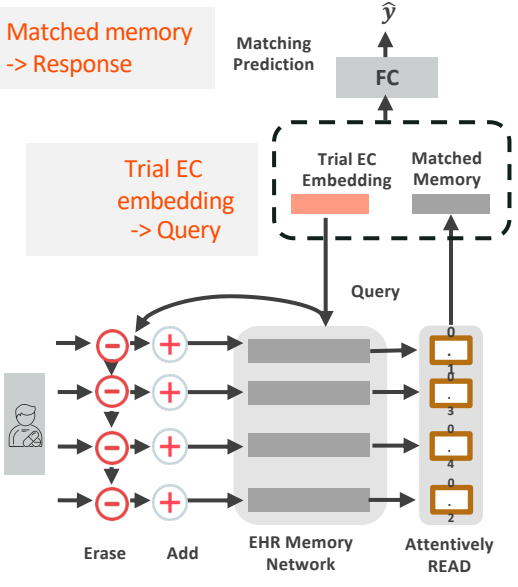
Separate memory networks to store diagnosis, medications and procedures

$$\mathbf{m} = [\mathbf{m}_{\mathcal{D}}, \mathbf{m}_{\mathcal{O}}, \mathbf{m}_{\mathcal{P}}]$$

$$= [\mathbf{m}^1_{\mathcal{D}}, \ldots, \mathbf{m}^4_{\mathcal{D}}, \mathbf{m}^1_{\mathcal{O}}, \ldots, \mathbf{m}^4_{\mathcal{O}}, \mathbf{m}^1_{\mathcal{P}}, \ldots, \mathbf{m}^4_{\mathcal{P}}]$$

Augment medical codes with textual description: For example, Code 692.9 -> "Contact dermatitis and other eczema"

Patient-Trial Matching: Matching



Attentional Alignment and Dynamic Matching

Let each EC correspond to the sub-memories, we perform attentional alignment and dynamic matching

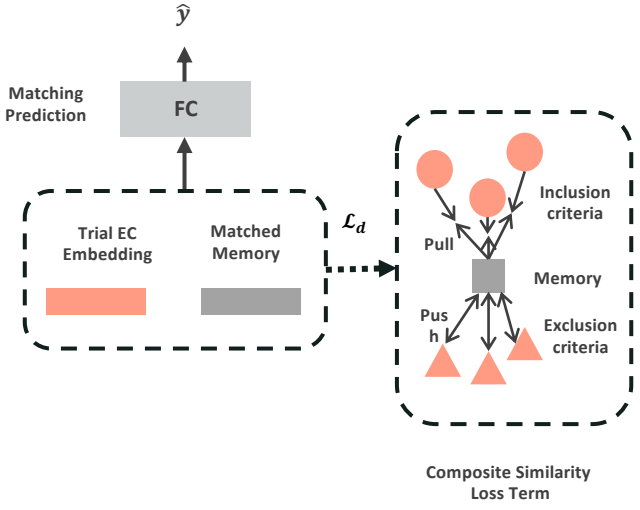
$$a_{k,G} = \frac{exp(\boldsymbol{m}_G^{k\,\mathrm{T}} MLP(\boldsymbol{e}))}{\sum_{x\in\{\mathcal{D},\mathcal{O},\mathcal{P}\}}\sum_{i=1}^{4} exp(\boldsymbol{m}_x^{i\,\mathrm{T}} MLP(\boldsymbol{e}))}$$

$$\widetilde{\boldsymbol{m}} = \sum_{x\in\{\mathcal{D},\mathcal{O},\mathcal{P}\}}\sum_{i=1}^{4} a_{i,x}\boldsymbol{m}_x^i$$

Predicting Patient-Trial Matching

$$\hat{\boldsymbol{y}} = softmax(\widetilde{\boldsymbol{m}} \oplus MLP(\boldsymbol{m}_d \oplus \boldsymbol{e}))$$

Patient-Trial Matching: Optimization



Joint Learning Using a Composite Similarity Loss Term

Cross-entropy classification Loss

$$\mathcal{L}_c = -(\boldsymbol{y}^{\mathrm{T}} log(\hat{\boldsymbol{y}}) + (1-\boldsymbol{y})^{\mathrm{T}} log(1-\hat{\boldsymbol{y}}))$$

minimize the distance between patient memory and the embedding for inclusion criteria

$$\mathcal{L}_d = \begin{cases} 1 - d(\boldsymbol{e}, \widetilde{\boldsymbol{m}}_I)), & if\ \boldsymbol{e}\ is\ \boldsymbol{e}_I \\ max(0, d(\boldsymbol{e}, \widetilde{\boldsymbol{m}}_E) - \alpha), & if\ \boldsymbol{e}\ is\ \boldsymbol{e}_E \end{cases}$$

-> 0

>= α

maximize the distance between the memory and the embedding for exclusion criteria

Jointly minimize $\mathcal{L} = \mathcal{L}_c + \mathcal{L}_d$

Patient-Trial Matching: Evaluation

Clinical Trial Data

590 trials from publicly available data source (clinicaltrials.gov)
12,445 criteria-level EC statements

Patient EHR Data

83,371 patients from 2002 to 2018

30% testing Set

Statistic	
# of trials	590
Avg. # of Inclusion criteria per trial	6.67
Avg. # of Exclusion criteria per trial	10.56
Avg. criteria sentence length	31.80
# of patients	83,371
# of unique medications	286
# of unique diagnosis	421
# of labeled pairs	397,321
Avg. # of trials per patient	1.20
Avg. # of patients per trial	169.49

trial level matching (100% criteria match)

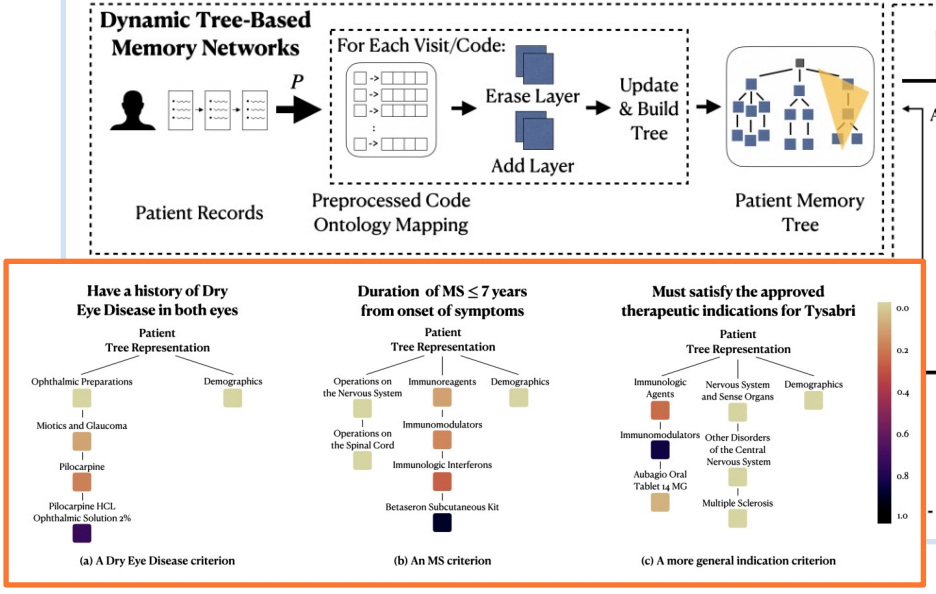
	Model	Accuracy
Baselines	LSTM+GloVe	0.4294±0.010
	LSTM+BERT	0.5460±0.008
	Criteria2Query	0.6147±-
	DeepEnroll	0.6737±0.021
Reduced	COMPOSE-MN	0.7833±0.011
	COMPOSE-Highway	0.8102±0.009
	COMPOSE-$\mathcal{L}_d$	0.8212±0.010
Proposed	COMPOSE	**0.8373±0.012**

criteria level matching

	Model	Accuracy	AUROC	AUPRC
Baselines	LSTM+GloVe	0.722±0.010	0.789±0.009	0.784±0.009
	LSTM+BERT	0.834±0.008	0.845±0.007	0.840±0.007
	DeepEnroll	0.869±0.012	0.936±0.013	0.947±0.011
Reduced	COMPOSE-MN	0.899±0.012	0.955±0.013	0.960±0.010
	COMPOSE-Highway	0.912±0.007	0.965±0.007	0.967±0.009
	COMPOSE-$\mathcal{L}_d$	0.939±0.010	0.976±0.009	0.973±0.007
Proposed	COMPOSE	**0.945±0.008**	**0.980±0.007**	**0.979±0.008**

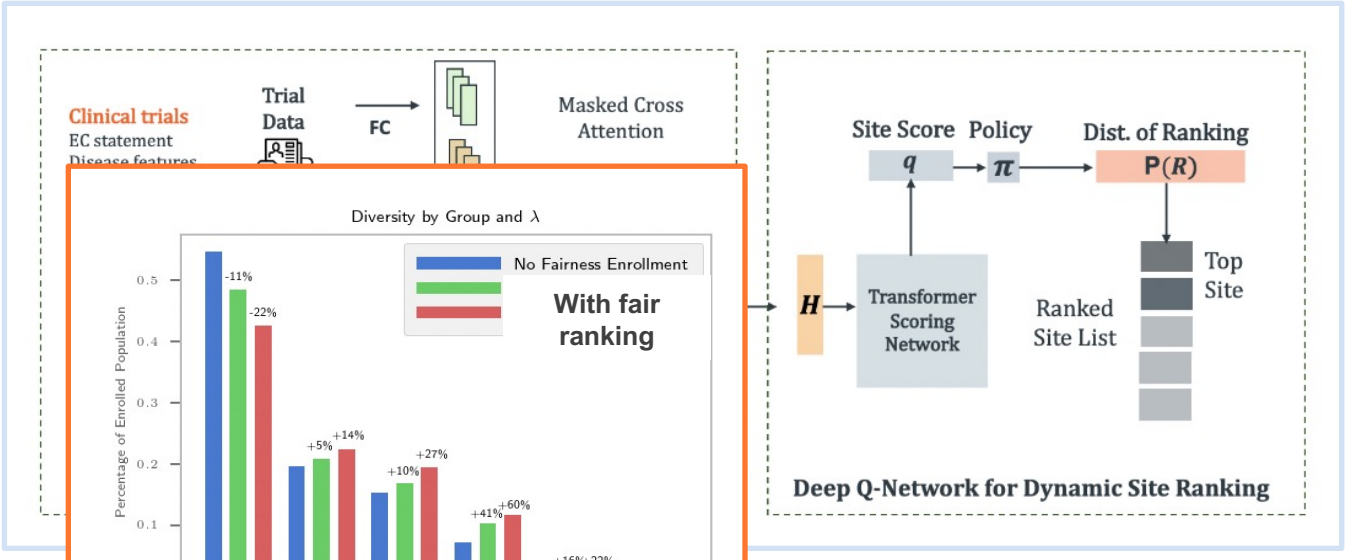
Model	Phase I	Phase II	Phase III
LSTM+GloVe	0.0008	0.5865	0.3743
LSTM+BERT	0.0025	0.6045	0.4862
Criteria2Query	0.3025	0.6433	0.5870
DeepEnroll	0.2034	0.7493	0.6329
COMPOSE	**0.5189**	**0.8939**	**0.8005**

Interpretable Patient Trial Matching



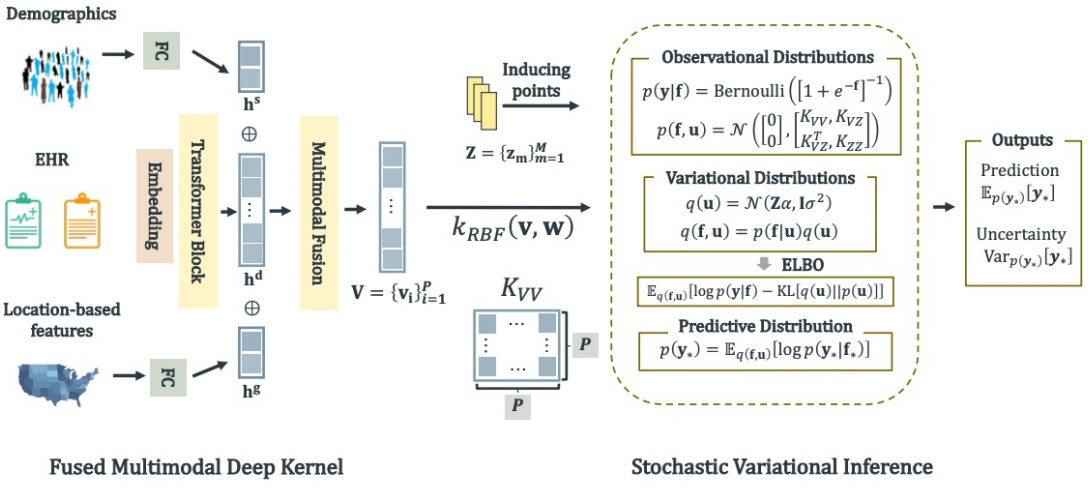
in submission

Fair Ranking for Trial Site Selection



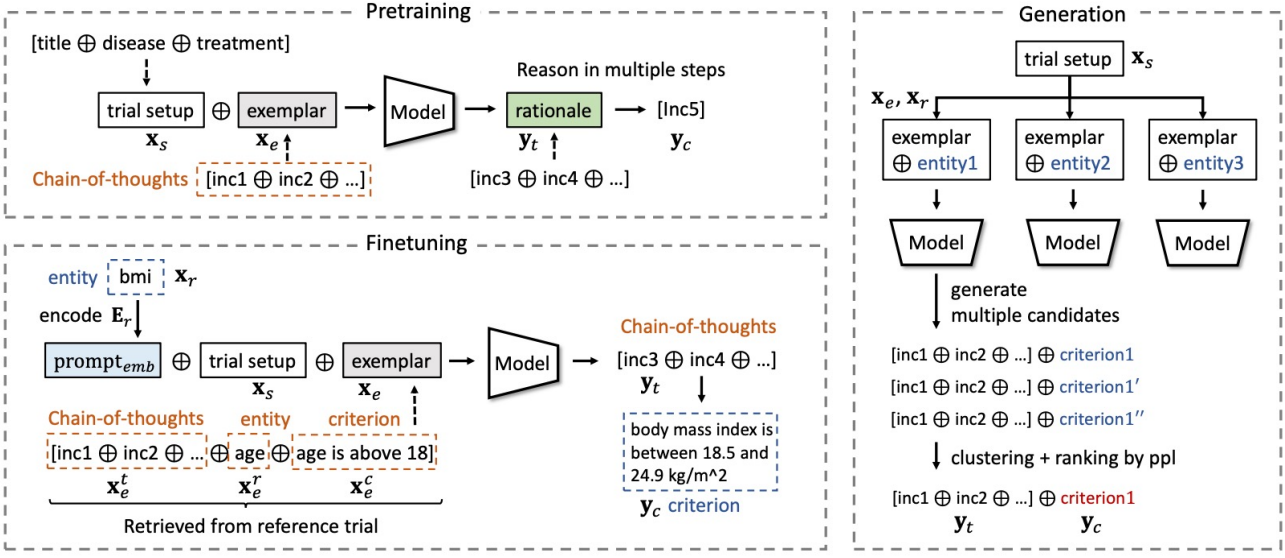
in submission

Patient-Retrieval: Model the Uncertainty of Individual Prediction



Chacha Chen, Junjie Liang, Fenglong Ma, Lucas Glass, Jimeng Sun and **Cao Xiao**. UNITE: Uncertainty-based Health Risk Prediction Leveraging Multi-sourced Data, The World Wide Web Conference (WWW' 21)

New Directions: Large Language Model for Trial Design



in submission

Agenda

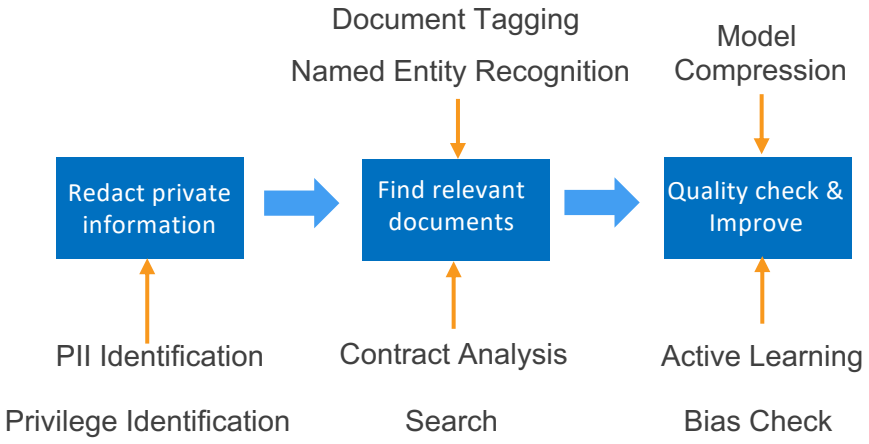
1 Introduction

2 ML for Clinical Trials

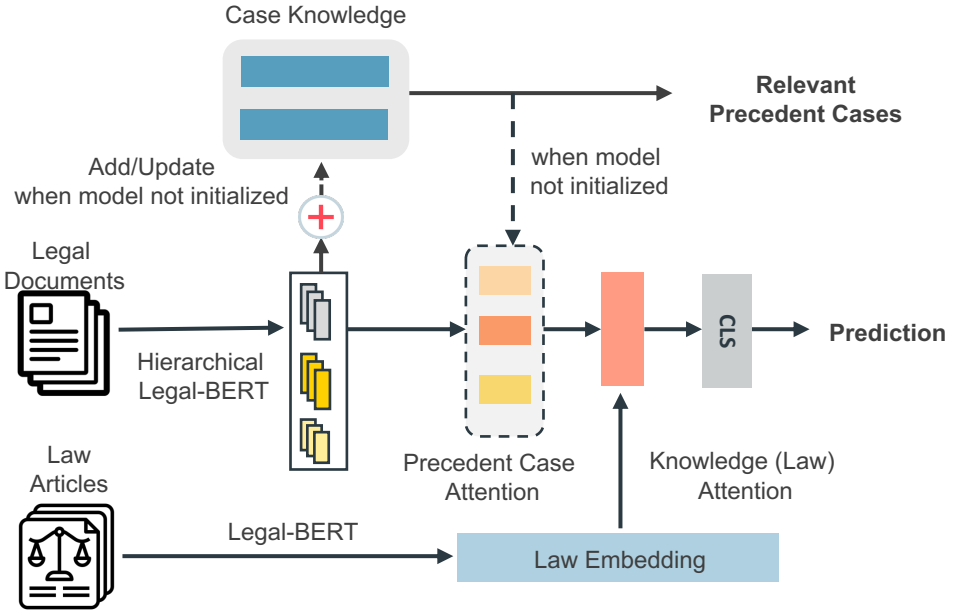
3 ML for Legal Review

4

Machine Learning for Legal Review Workflow



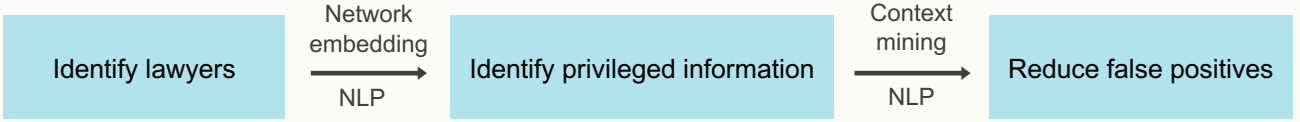
Document Review: Document Tagging and Relevant Document Retrieval



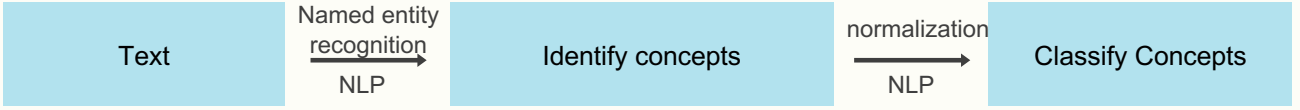
Beyond Document Review: Privilege Identification and Privacy Detection

Relativity

The Privilege Identification Workflow



The Privacy Detection Workflow



Please find attached the report of the investigation. For further information, please contact Olivia and Jim at their email addresses, olivia@sample.com and jim@sample.com, respectively. You may also reach Olivia at (212)456-7890 if necessary.

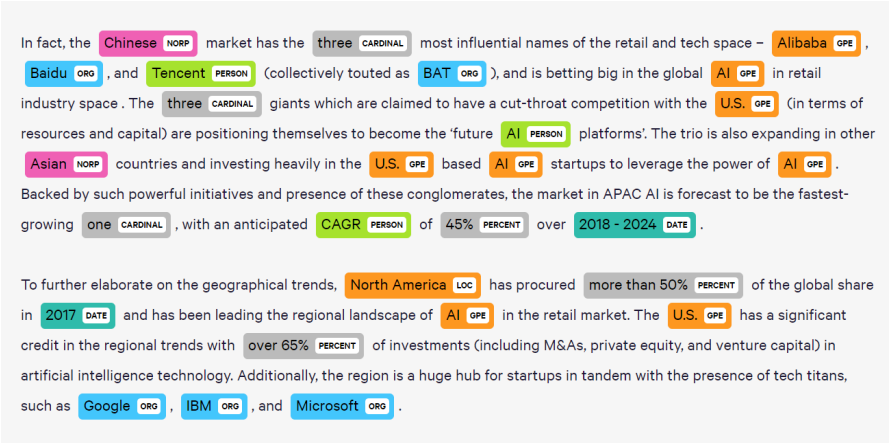
name

Phone number

Email address

Foundational Task and Its Challenge

Named Entity Recognition (NER) is an NLP technique that involves identifying and extracting named entities from unstructured text data.



Legal NER models are developed to automatically identify, and extract and link named entities from legal documents, such as contracts, court opinions, and legislation, and provides key information for legal decision making.

Bias in the Legal NER

Example 1. lower accuracy rates in identifying the names of people with non-European names, compared to European names. This can lead to the exclusion of certain racial or ethnic groups from legal analyses or decisions

Example 2. failed to connect people's name with their roles due to gender bias.

Sources of Bias

- Ambiguity in Legal Language
- Historical and Cultural Biases
- Imbalanced Training Data
- Lack of Diversity in Training Data
- Bias in the Testing Data

Agenda

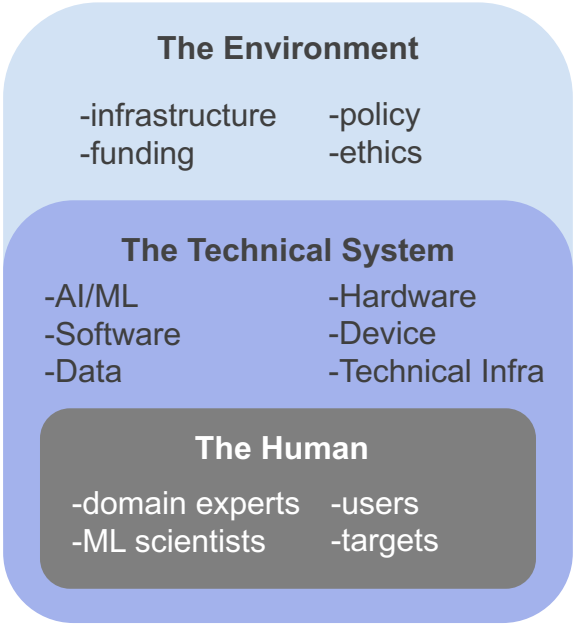
1 Introduction

2 ML for Clinical Trials

3 ML for Legal Review

4 Trade-offs and Other Considerations

Trade-offs and Other Considerations



Trade-offs

- Data quality vs. data privacy
- Model quality vs. query cost
- Human-in-the-loop vs. user experience
- .....

Update

- New data being piped into the system
- New feedbacks received from experts
- Data removal
- ……

The trade-offs and needs constrain AI/ML model development and deployment, however, also become the challenges that motivate more research works.

Q&A

https://sites.google.com/view/danicaxiao/home
danicaxiao@gmail.com