Large Language Models to understand biomedical text

Yuan Luo, PhD, FIAHSI, FAMIA
Chief AI Officer
Northwestern University Clinical and Translational Sciences Institute
Institute for Augmented Intelligence in Medicine
Associate Professor
Department of Preventive Medicine
Northwestern University
yuan.luo@northwestern.edu
🐦@yuanhypnosluo

9/18/2023

NORTHWESTERN UNIVERSITY FEINBERG SCHOOL OF MEDICINE

Northwestern University

Large Language Models – oversimplified review

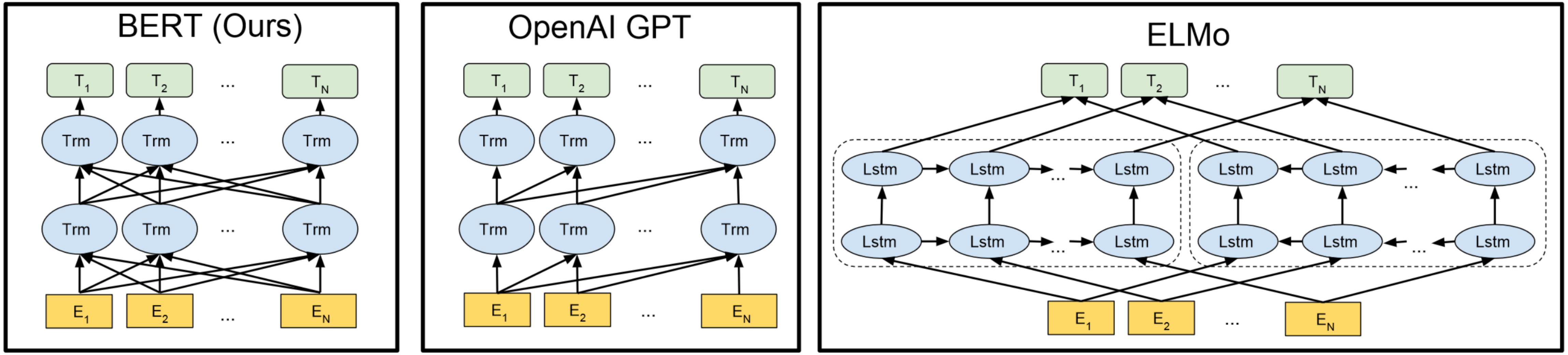


Figure credit: Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805. 2018 Oct 11.

Roadmap

- Communication
  - ChatGPT in medical and scientific writing
- Resource
  - BERT in building knowledge graphs
- Practice
  - Clinical-Longformer and Clinical-BigBird
- Surveillance
  - ChatGPT in pharmacovigilance

Study design

Prompt: ‘Please write a scientific abstract for the article [title] in the style of [journal] at [link]’

JAMA
Importance
Objective
Design, Setting, and Participants
Interventions
Main Outcomes and Measures
Results
Conclusions and Relevance

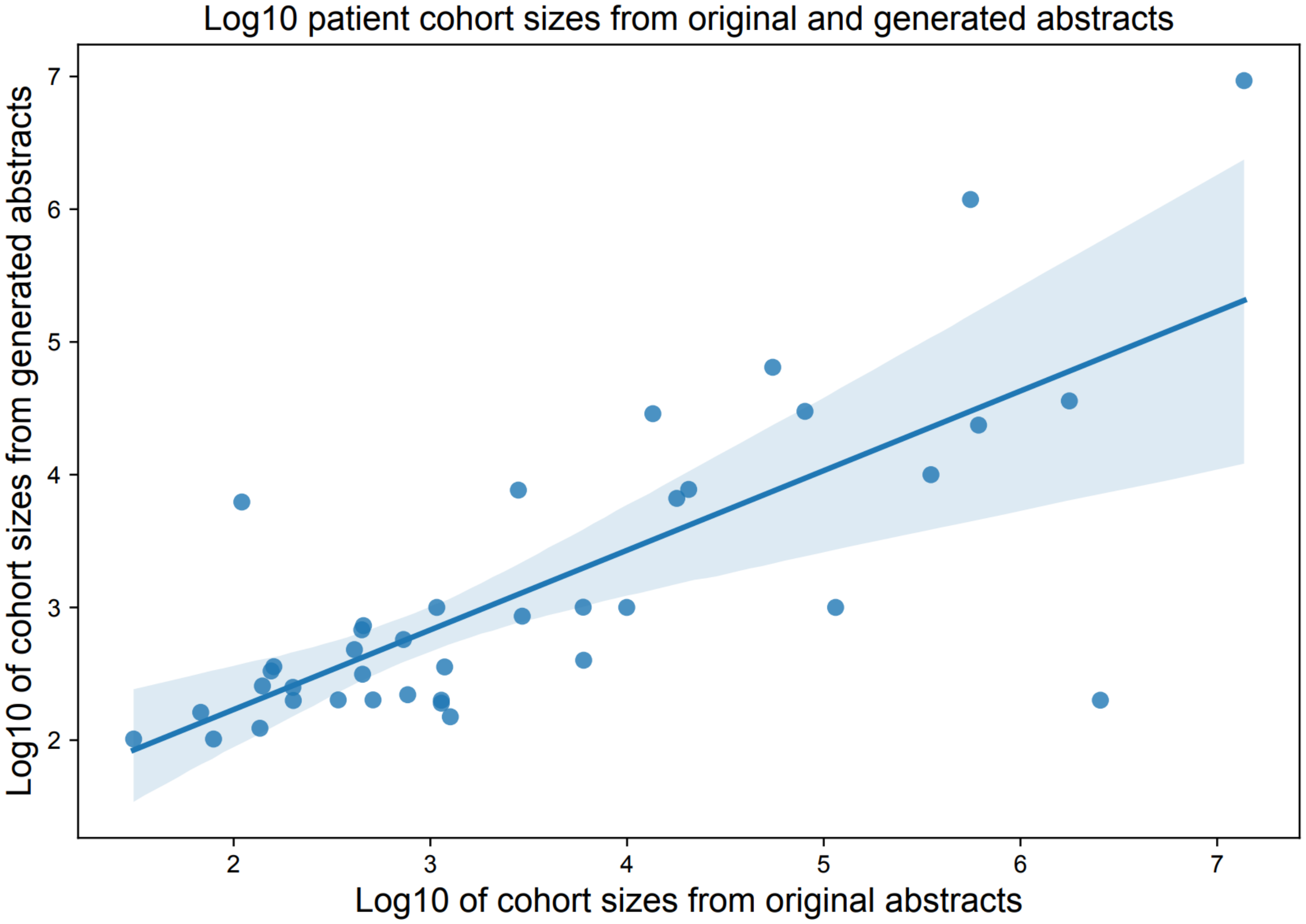
The NEJM
Background
Methods
Results
Conclusions

Nature Medicine - headerless paragraph abstract

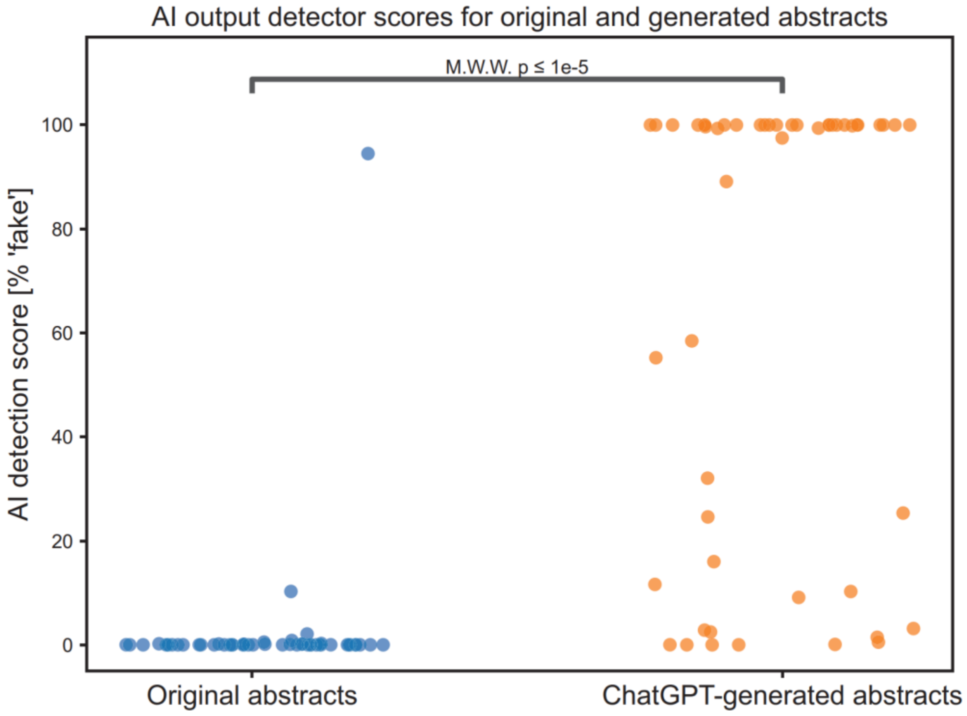
The BMJ *(variable headers depending on study type)*
Objective
Design
Data source
Setting
Participants
Data extraction and synthesis
Main outcome measures
Results
Conclusions

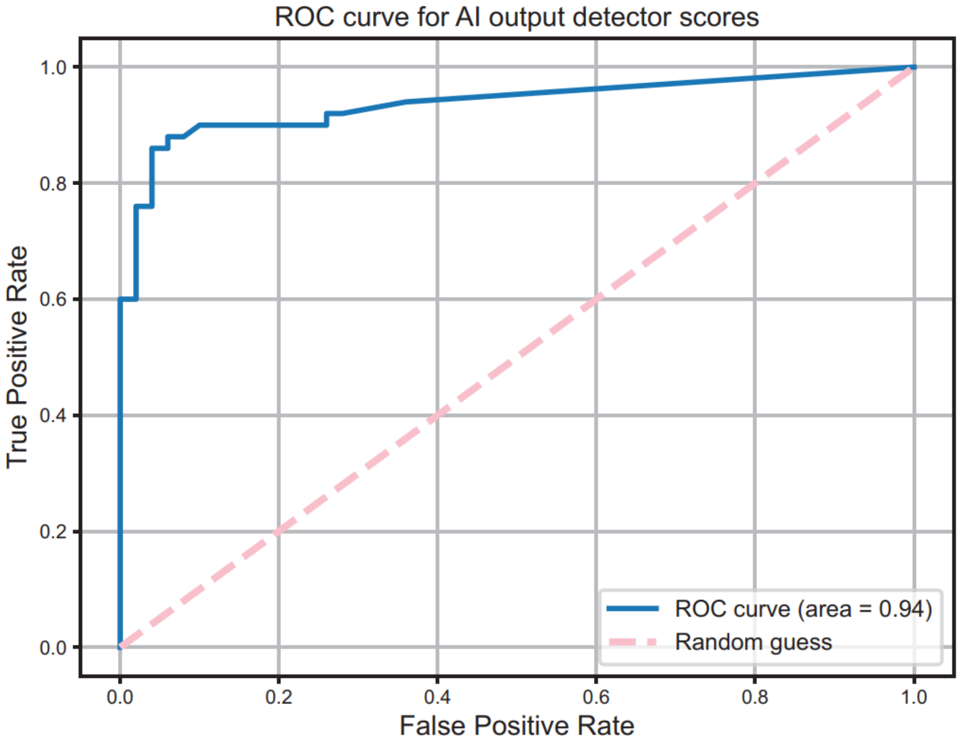
The Lancet
Background
Methods
Findings
Interpretation

Generated abstracts have a similar patient cohort size as original abstracts

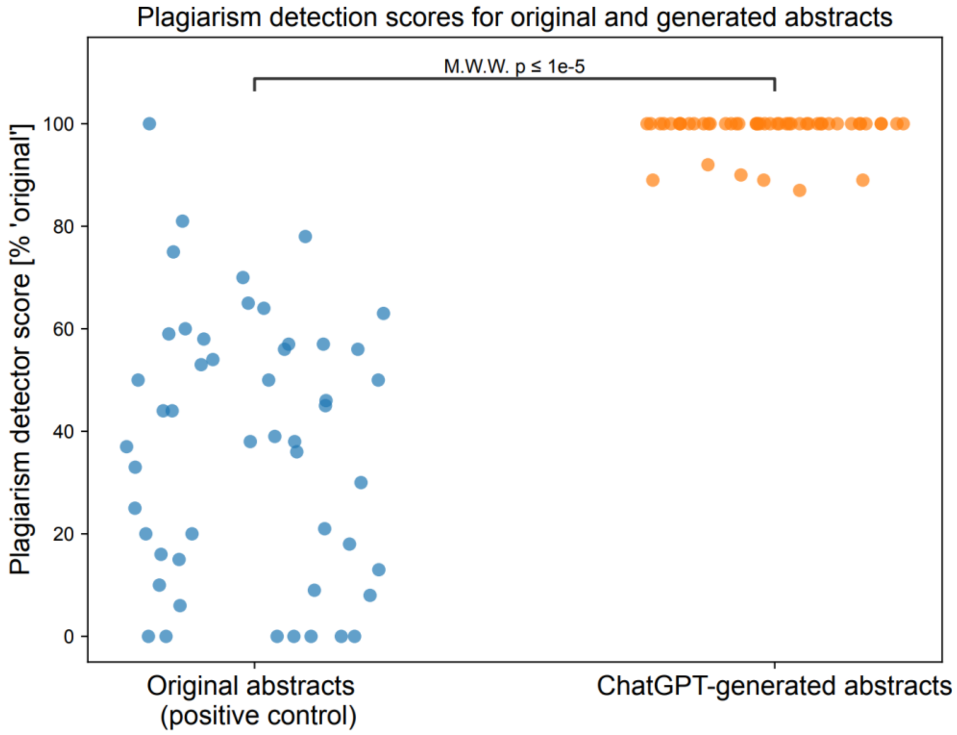


Many generated abstracts can be detected using an AI output detector



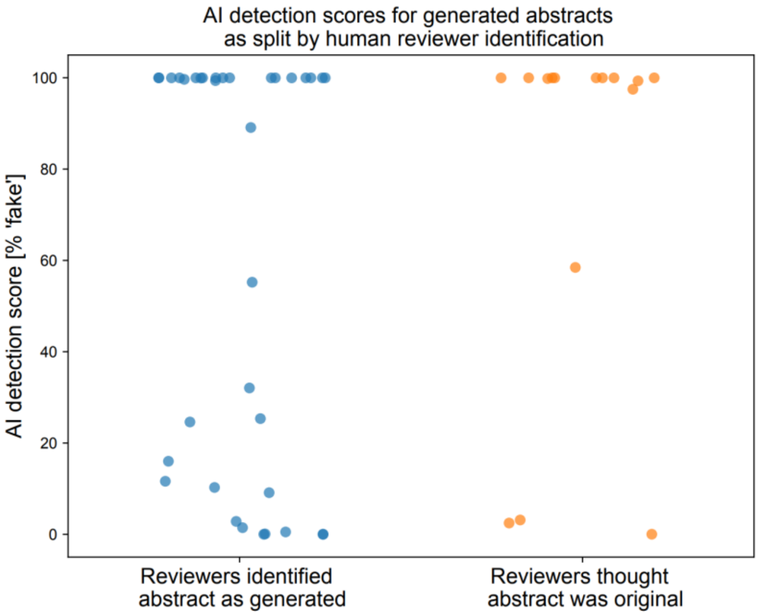


Generated abstracts are original and do not plagiarize from other written work



Reviewers use criteria different than the AI output detector for flagging abstracts

Reviewers were able to correctly identify 68% of generated abstracts as being generated by ChatGPT, but incorrectly identified 14% of original abstracts as being generated



AI detection scores for generated abstracts as split by human reviewer identification

		Truth	
		Original	Generated
Reviewer guess	Original	43	16
	Generated	7	34

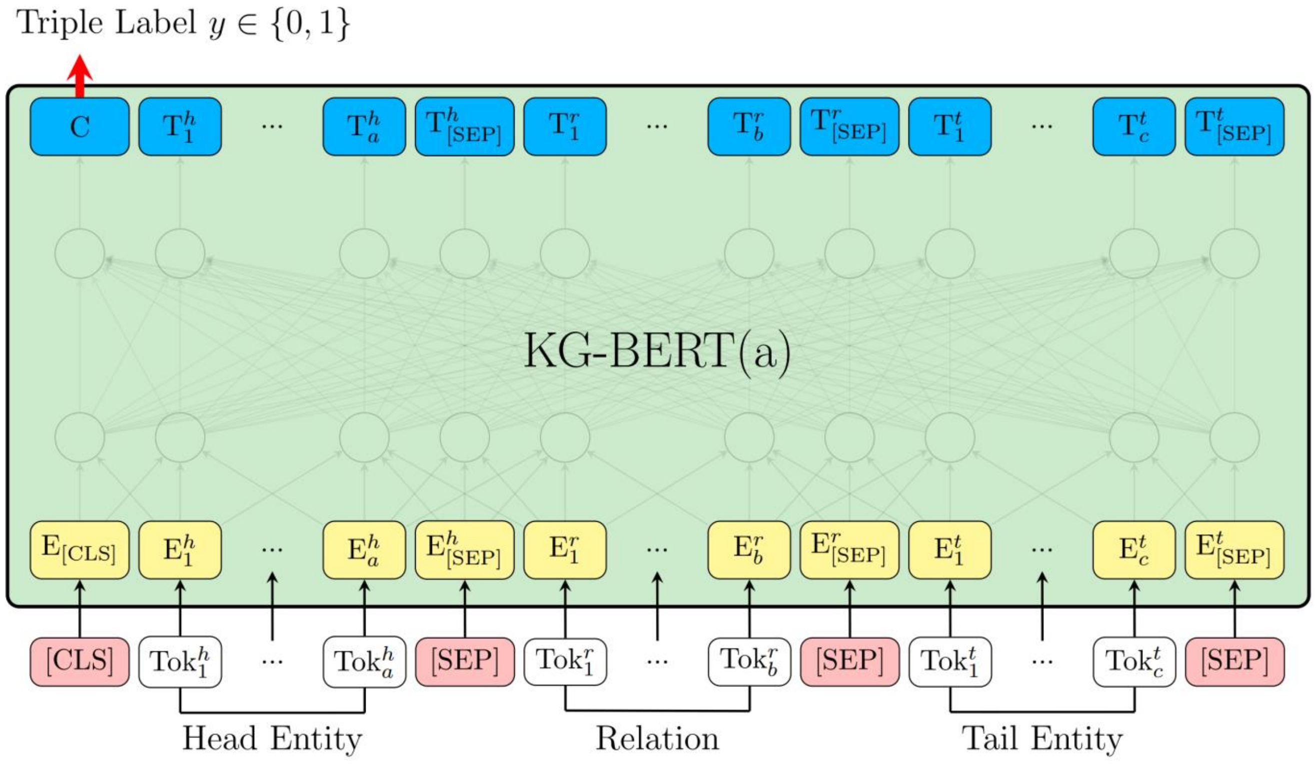
Gao CA, Howard FM, Markov NS, Dyer EC, Ramesh S, Luo Y, Pearson AT. Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. *NPJ Digital Medicine. 2023 6 (75).*

Roadmap

- Communication
  - ChatGPT in medical and scientific writing
- Resource
  - BERT in building knowledge graphs
- Practice
  - Clinical-Longformer and Clinical-BigBird
- Surveillance
  - ChatGPT in pharmacovigilance

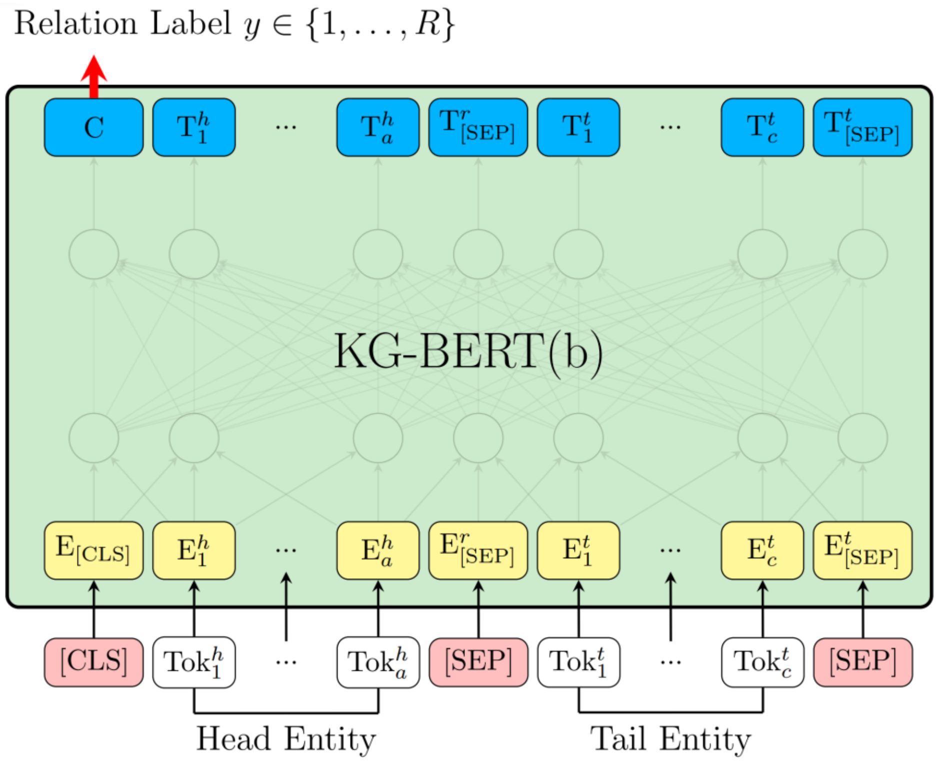
Fine-tuning KG-BERT for predicting triples, links, and relations

Predicting the plausibility of a triple or a link



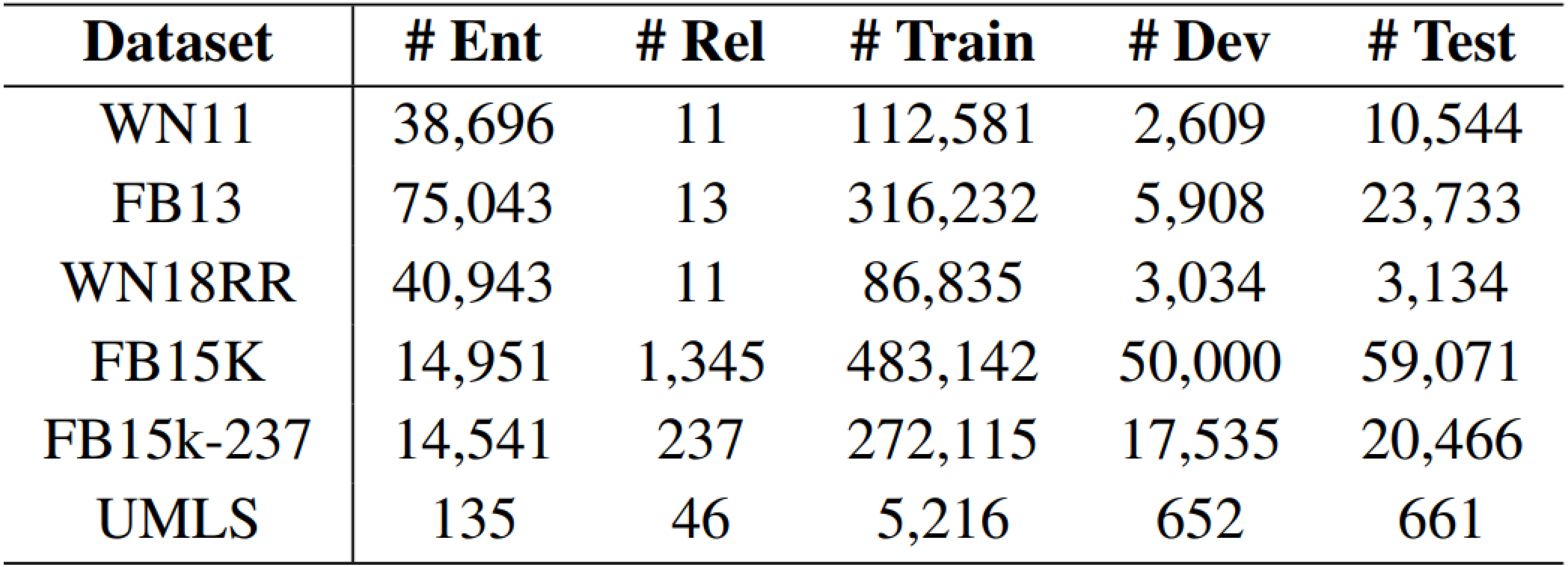
$$\mathcal{L} = -\sum_{\tau \in \mathbb{D}^+ \cup \mathbb{D}^-} (y_\tau \log(s_{\tau 0}) + (1 - y_\tau) \log(s_{\tau 1}))$$

Predicting relations



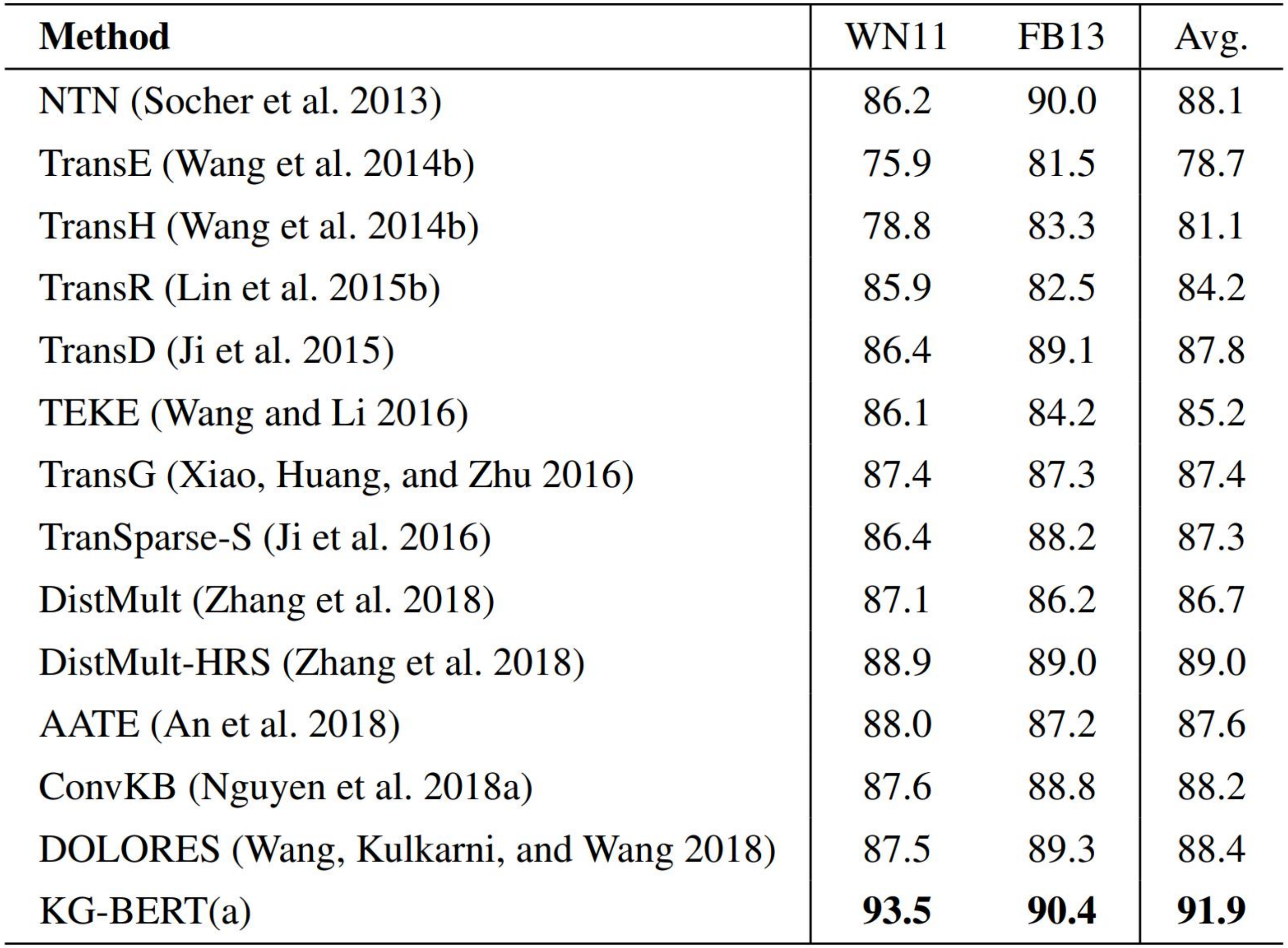
$$\mathcal{L}' = -\sum_{\tau \in \mathbb{D}^+} \sum_{i=1}^{R} y'_{\tau i} \log(s'_{\tau i})$$

Summary statistics of datasets



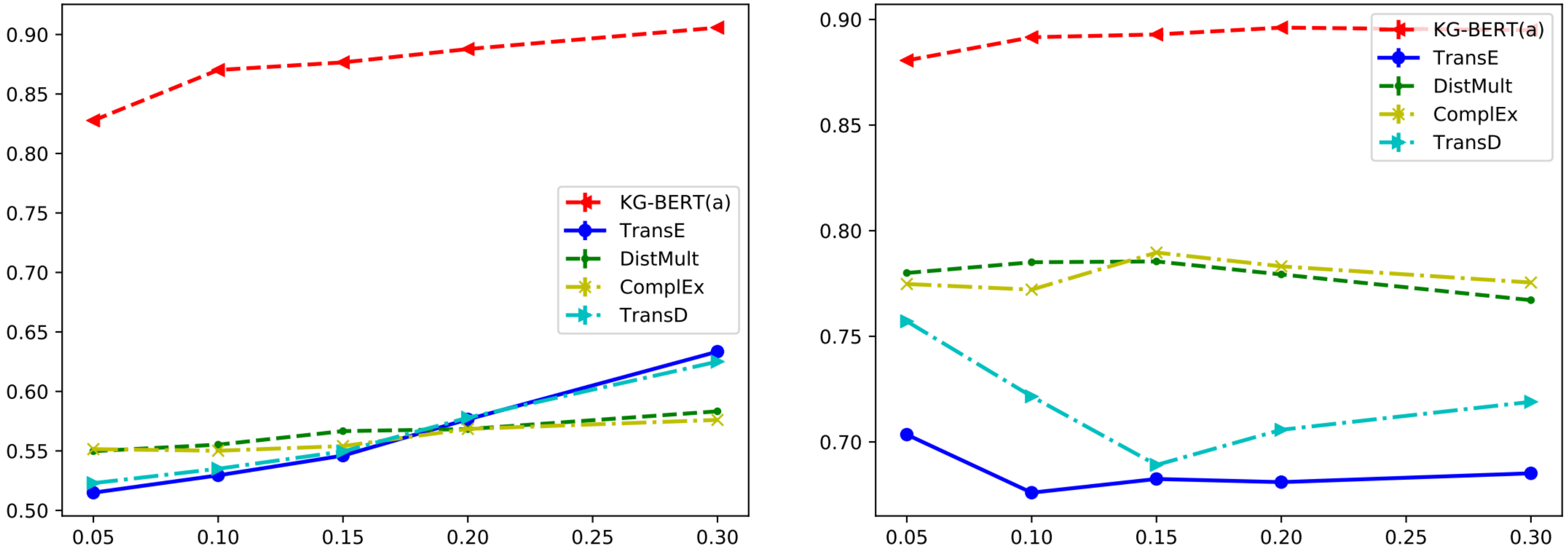
Dataset	# Ent	# Rel	# Train	# Dev	# Test
WN11	38,696	11	112,581	2,609	10,544
FB13	75,043	13	316,232	5,908	23,733
WN18RR	40,943	11	86,835	3,034	3,134
FB15K	14,951	1,345	483,142	50,000	59,071
FB15k-237	14,541	237	272,115	17,535	20,466
UMLS	135	46	5,216	652	661

Triple classification accuracy



Method	WN11	FB13	Avg.
NTN (Socher et al. 2013)	86.2	90.0	88.1
TransE (Wang et al. 2014b)	75.9	81.5	78.7
TransH (Wang et al. 2014b)	78.8	83.3	81.1
TransR (Lin et al. 2015b)	85.9	82.5	84.2
TransD (Ji et al. 2015)	86.4	89.1	87.8
TEKE (Wang and Li 2016)	86.1	84.2	85.2
TransG (Xiao, Huang, and Zhu 2016)	87.4	87.3	87.4
TranSparse-S (Ji et al. 2016)	86.4	88.2	87.3
DistMult (Zhang et al. 2018)	87.1	86.2	86.7
DistMult-HRS (Zhang et al. 2018)	88.9	89.0	89.0
AATE (An et al. 2018)	88.0	87.2	87.6
ConvKB (Nguyen et al. 2018a)	87.6	88.8	88.2
DOLORES (Wang, Kulkarni, and Wang 2018)	87.5	89.3	88.4
KG-BERT(a)	**93.5**	**90.4**	**91.9**

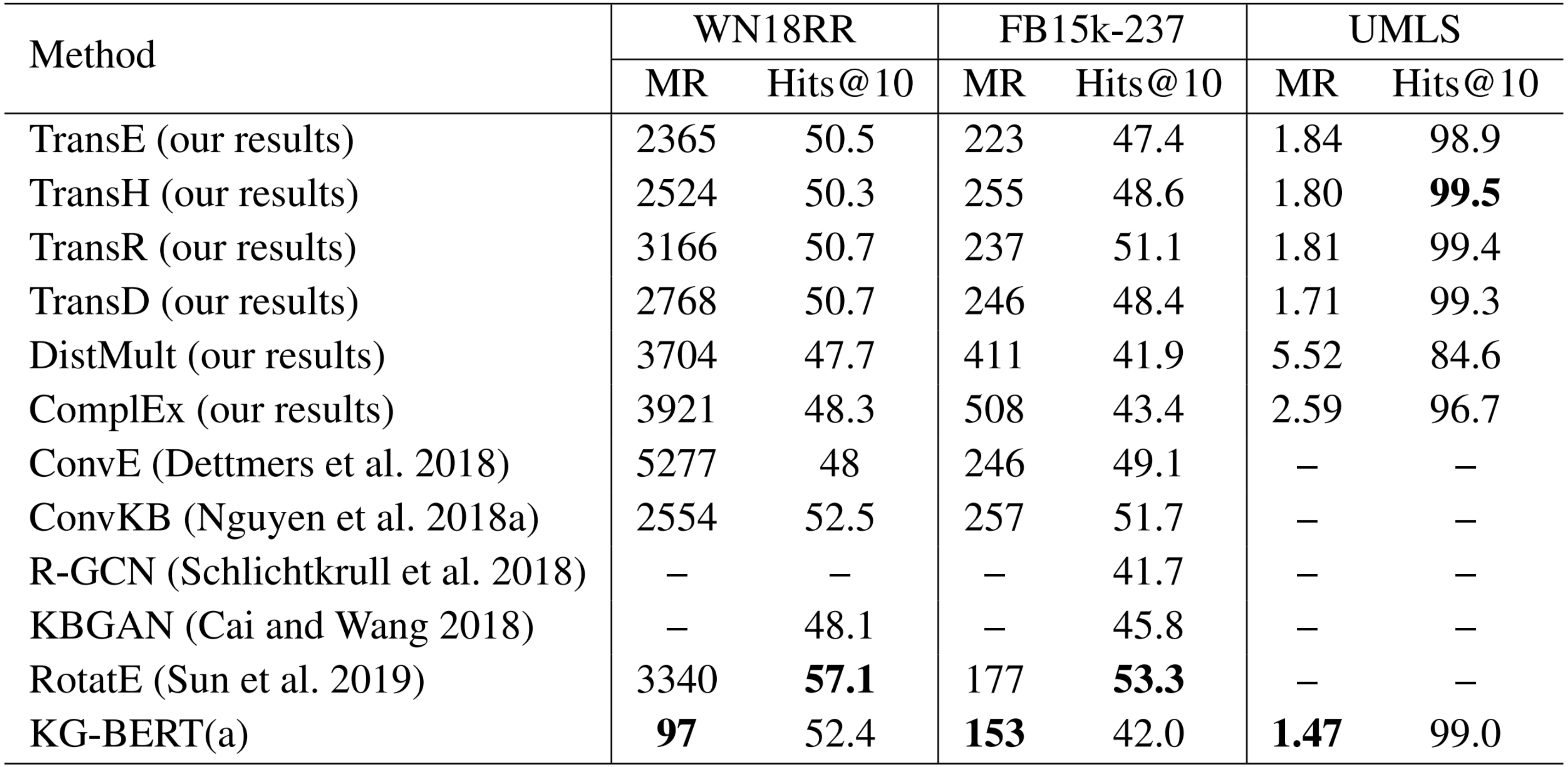
Test accuracy of triple classification by varying training data proportions



WN11

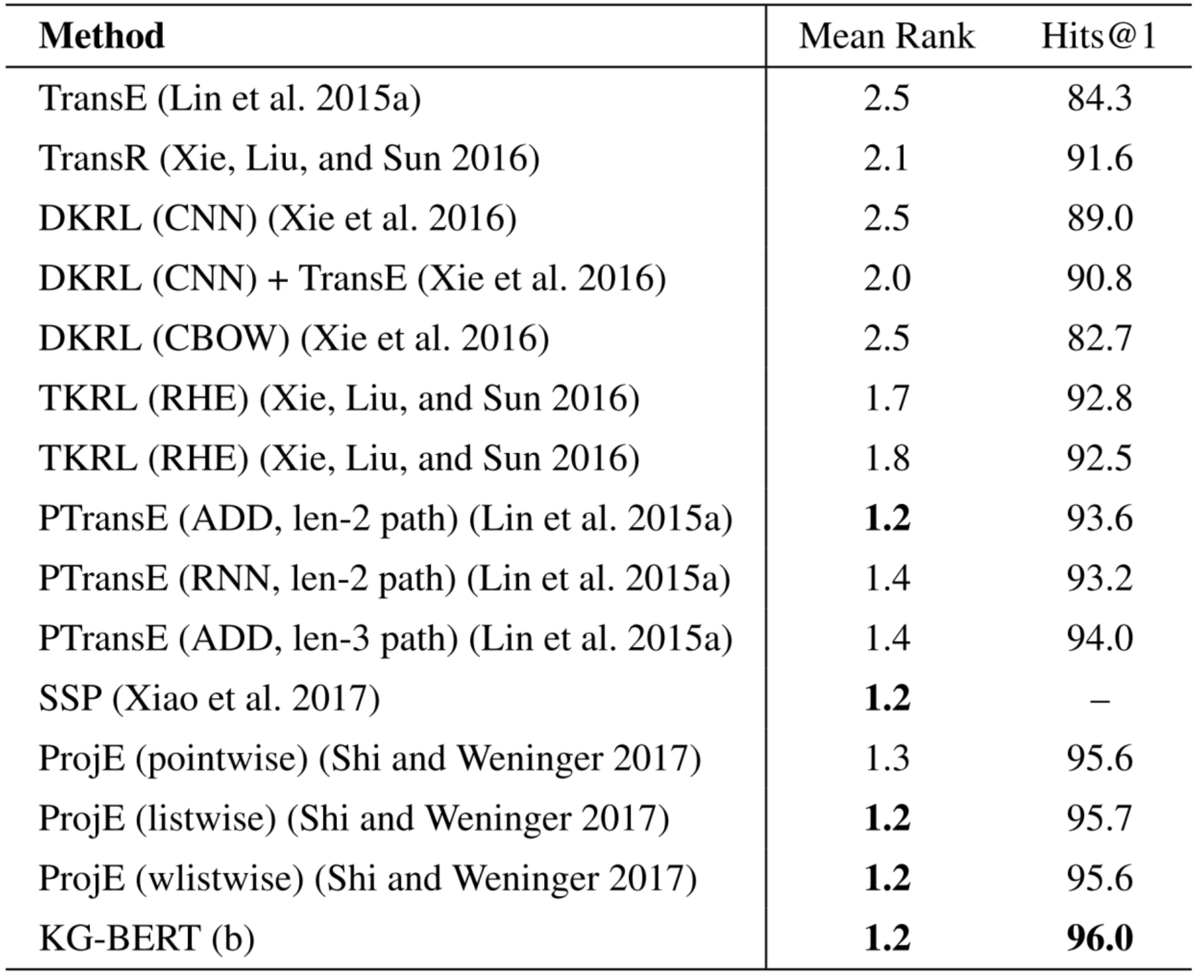
FB13

Link prediction results



Method	WN18RR		FB15k-237		UMLS	
	MR	Hits@10	MR	Hits@10	MR	Hits@10
TransE (our results)	2365	50.5	223	47.4	1.84	98.9
TransH (our results)	2524	50.3	255	48.6	1.80	**99.5**
TransR (our results)	3166	50.7	237	51.1	1.81	99.4
TransD (our results)	2768	50.7	246	48.4	1.71	99.3
DistMult (our results)	3704	47.7	411	41.9	5.52	84.6
ComplEx (our results)	3921	48.3	508	43.4	2.59	96.7
ConvE (Dettmers et al. 2018)	5277	48	246	49.1	–	–
ConvKB (Nguyen et al. 2018a)	2554	52.5	257	51.7	–	–
R-GCN (Schlichtkrull et al. 2018)	–	–	–	41.7	–	–
KBGAN (Cai and Wang 2018)	–	48.1	–	45.8	–	–
RotatE (Sun et al. 2019)	3340	**57.1**	177	**53.3**	–	–
KG-BERT(a)	**97**	52.4	**153**	42.0	**1.47**	99.0

Relation prediction results



Method	Mean Rank	Hits@1
TransE (Lin et al. 2015a)	2.5	84.3
TransR (Xie, Liu, and Sun 2016)	2.1	91.6
DKRL (CNN) (Xie et al. 2016)	2.5	89.0
DKRL (CNN) + TransE (Xie et al. 2016)	2.0	90.8
DKRL (CBOW) (Xie et al. 2016)	2.5	82.7
TKRL (RHE) (Xie, Liu, and Sun 2016)	1.7	92.8
TKRL (RHE) (Xie, Liu, and Sun 2016)	1.8	92.5
PTransE (ADD, len-2 path) (Lin et al. 2015a)	**1.2**	93.6
PTransE (RNN, len-2 path) (Lin et al. 2015a)	1.4	93.2
PTransE (ADD, len-3 path) (Lin et al. 2015a)	1.4	94.0
SSP (Xiao et al. 2017)	**1.2**	–
ProjE (pointwise) (Shi and Weninger 2017)	1.3	95.6
ProjE (listwise) (Shi and Weninger 2017)	**1.2**	95.7
ProjE (wlistwise) (Shi and Weninger 2017)	**1.2**	95.6
KG-BERT (b)	**1.2**	**96.0**

Yao L, Mao C, Luo Y. KG-BERT: BERT for knowledge graph completion. arXiv preprint arXiv:1909.03193. 2019 Sep 7.

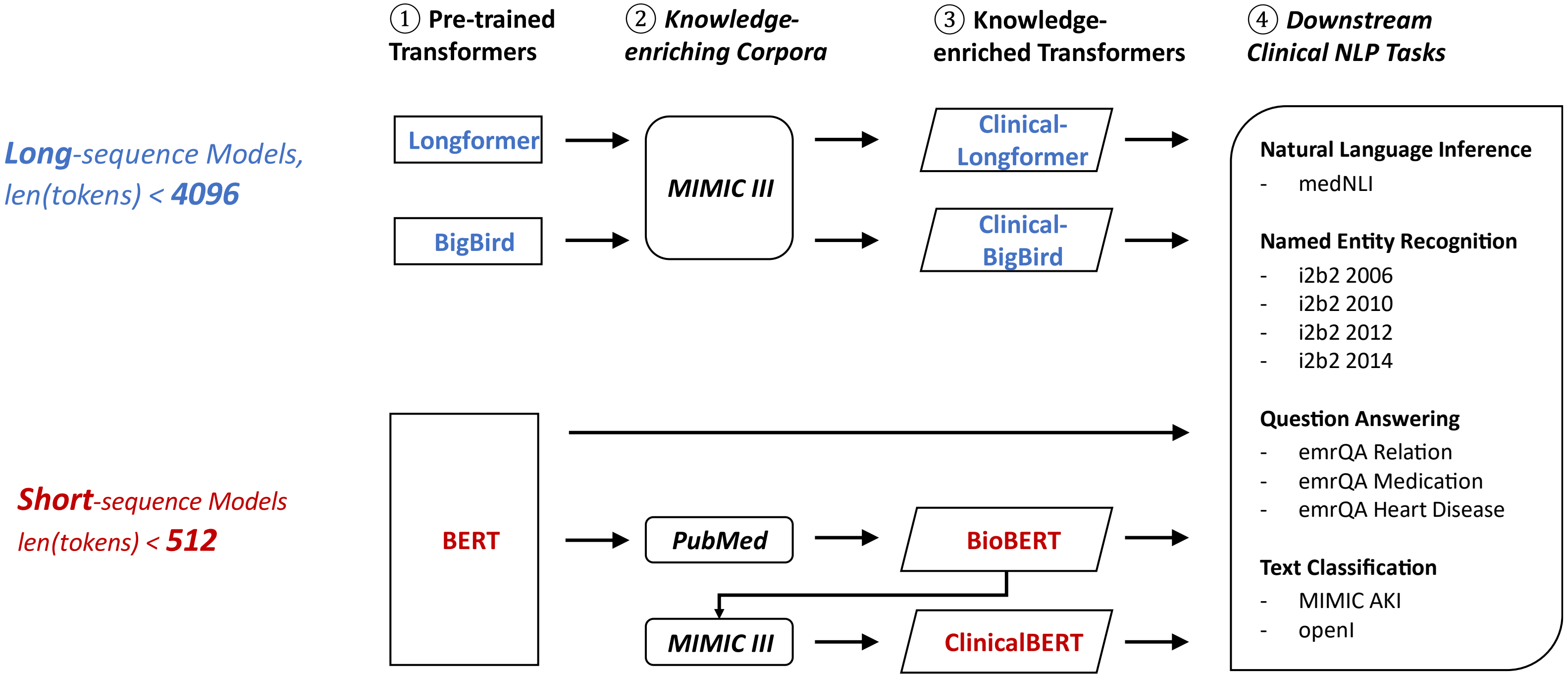
Roadmap

- Communication
  - ChatGPT in medical and scientific writing
- Resource
  - BERT in building knowledge graphs
- Practice
  - Clinical-Longformer and Clinical-BigBird
- Surveillance
  - ChatGPT in pharmacovigilance

Background and motivation

- Transformer-based models, such as BERT, ClinicalBERT, and BioBERT, are designed to handle text inputs that are up to a maximum length of 512 tokens
- This limit poses a challenge for clinical texts, such as pathology reports, which tend to be much longer
- The Longformer and BigBird models extend the maximum input length from 512 tokens to 4,096 tokens by implementing sparse attention mechanisms. Both models have achieved great success in the general domain
- However, the adaptability of both models to the clinical contexts remain unclear
- To create two clinical knowledge-enriched language models, Clinical-Longformer and Clinical-BigBird, through pre-training on large-scale clinical notes
- To compare the effectiveness of Clinical-Longformer and Clinical-BigBird with short text models in improving the performance of various downstream clinical NLP tasks

Large Language Models



Long**-sequence Models, len(tokens) < **4096

Short**-sequence Models len(tokens) < **512

① **Pre-trained Transformers**

② ***Knowledge-enriching Corpora***

③ **Knowledge-enriched Transformers**

④ ***Downstream Clinical NLP Tasks***

Longformer → ***MIMIC III*** → **Clinical-Longformer** →

BigBird → ***MIMIC III*** → **Clinical-BigBird** →

BERT →

BERT → ***PubMed*** → **BioBERT** →

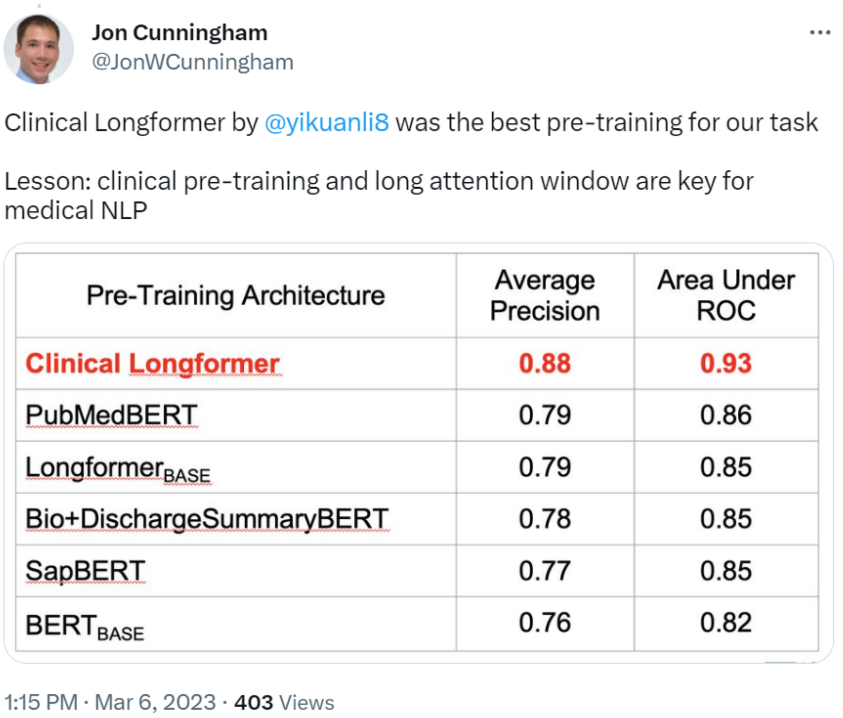
BioBERT → ***MIMIC III*** → **ClinicalBERT** →

Natural Language Inference
- medNLI

Named Entity Recognition
- i2b2 2006
- i2b2 2010
- i2b2 2012
- i2b2 2014

Question Answering
- emrQA Relation
- emrQA Medication
- emrQA Heart Disease

Text Classification
- MIMIC AKI
- openI



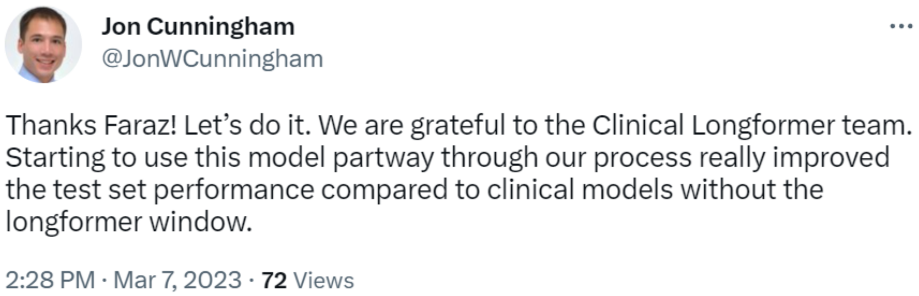
Jon Cunningham
@JonWCunningham

Clinical Longformer by @yikuanli8 was the best pre-training for our task

Lesson: clinical pre-training and long attention window are key for medical NLP

Pre-Training Architecture	Average Precision	Area Under ROC
Clinical Longformer	0.88	0.93
PubMedBERT	0.79	0.86
Longformer$_{BASE}$	0.79	0.85
Bio+DischargeSummaryBERT	0.78	0.85
SapBERT	0.77	0.85
BERT$_{BASE}$	0.76	0.82

1:15 PM · Mar 6, 2023 · **403** Views



Jon Cunningham
@JonWCunningham

Thanks Faraz! Let's do it. We are grateful to the Clinical Longformer team. Starting to use this model partway through our process really improved the test set performance compared to clinical models without the longformer window.

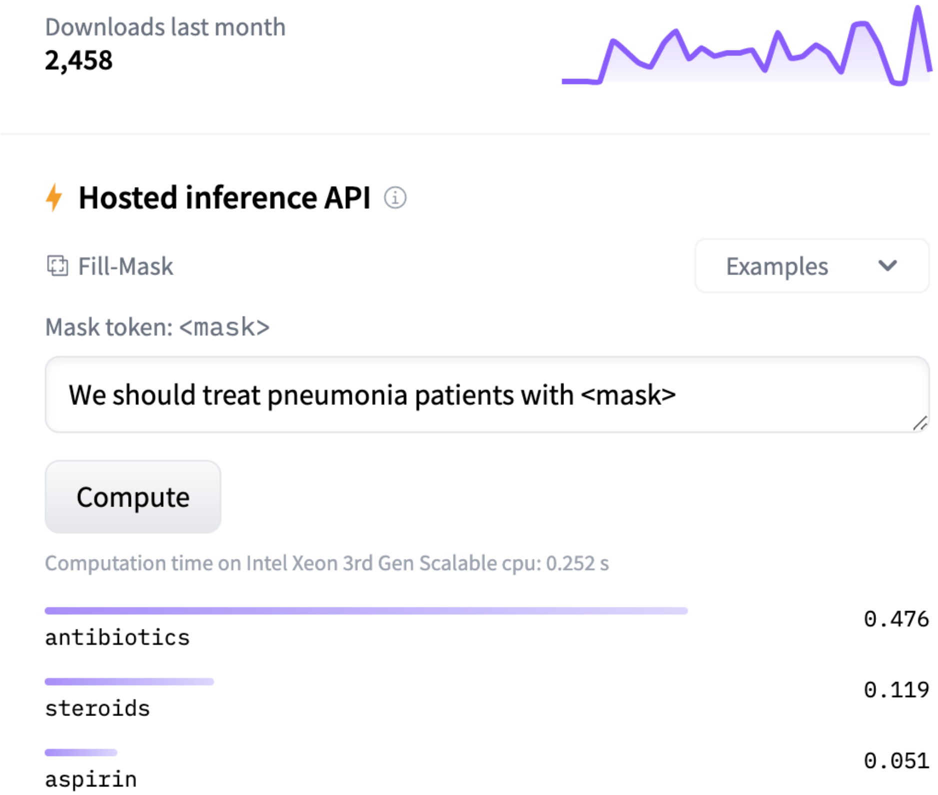
2:28 PM · Mar 7, 2023 · **72** Views

Results

	① ② ③ **Transformers**	④ ***Downstream Clinical NLP Tasks***									
		acc @ medNLI	**F1 @ i2b2 2006**	**F1 @ i2b2 2010**	**F1 @ i2b2 2012**	**F1 @ i2b2 2014**	**F1 @ emrQA Med**	**F1 @ emrQA Rela**	**F1 @ emrQA CVD**	**acc @ openI**	**AUC @ MIMIC AKI**
Long**-text Models* *len(tokens) < **4096	**Clinical-Longformer**	0.842	0.974	0.887	0.800	0.961	0.716	0.948	0.734	0.977	0.762
	Clinical-BigBird	0.827	0.967	0.872	0.787	0.952	0.715	0.944	0.711	0.972	0.755
Short**-text Models* *len(tokens) < **512	**BERT**	0.776	0.939	0.835	0.759	0.928	0.675	0.924	0.698	0.952	0.514
	BioBERT	0.808	0.948	0.865	0.789	0.93	0.700	0.926	0.702	0.954	0.534
	ClinicalBERT	0.812	0.951	0.861	0.773	0.929	0.698	0.929	0.711	0.967	0.738

Results

- We released the pre-trained models and codebase on *HuggingFace* and *GitHub*
  - https://huggingface.co/yikuan8/Clinical-Longformer
  - https://huggingface.co/yikuan8/Clinical-BigBird
  - https://github.com/luoyuanlab/Clinical-Longformer



Downloads last month
2,458

Hosted inference API

Fill-Mask

Examples

Mask token: `<mask>`

We should treat pneumonia patients with <mask>

Compute

Computation time on Intel Xeon 3rd Gen Scalable cpu: 0.252 s

antibiotics	0.476
steroids	0.119
aspirin	0.051

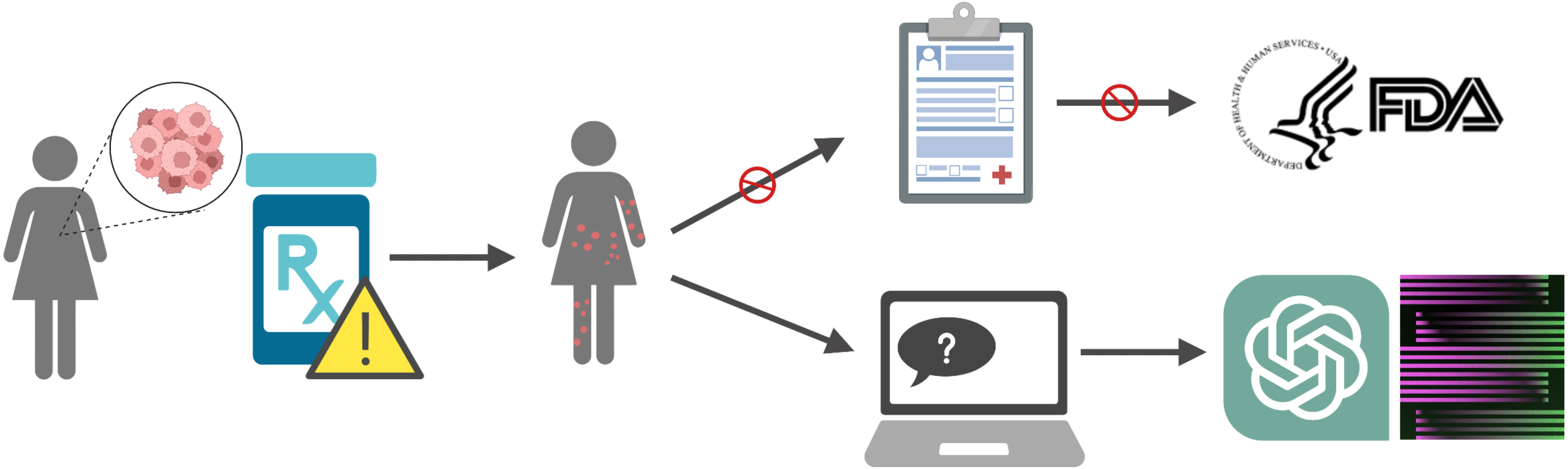
Y Li, R Wehbe, F Ahmad, H Wang, Y Luo. A Comparative Study of Pretrained Language Models for Long Clinical Text. *JAMIA 2023 30(2):340-7*

Roadmap

- Communication
  - ChatGPT in medical and scientific writing
- Resource
  - BERT in building knowledge graphs
- Practice
  - Clinical-Longformer and Clinical-BigBird
- Surveillance
  - ChatGPT in pharmacovigilance

ChatGPT-based Adverse Drug Reaction (ADR) detection

- To detect a rare ADR from clinical trials, 10,000 participants is needed[1]
- Only 1-10% of ADRS are reported to the FDA Adverse Event Reporting System[2]

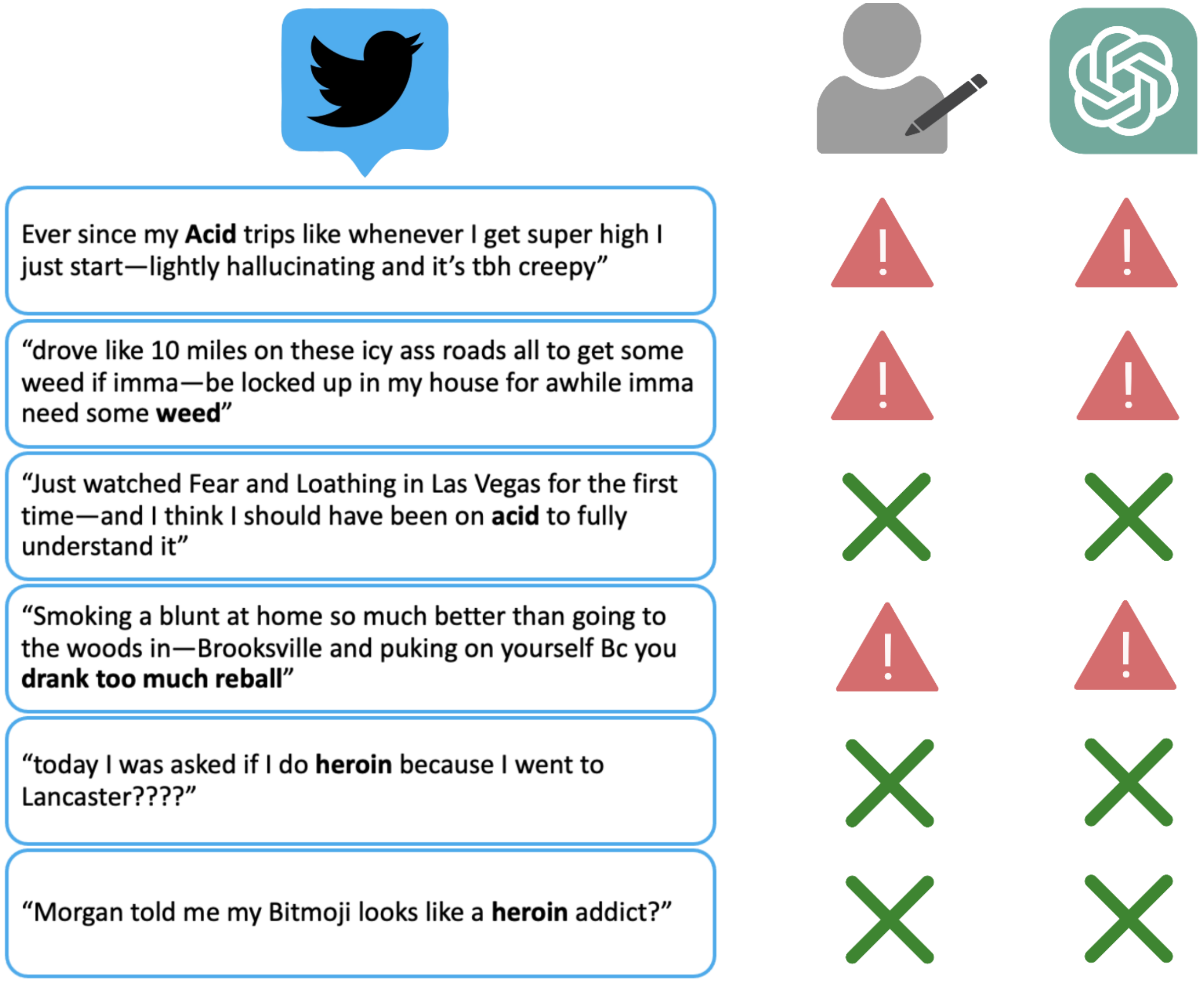


1. Lee et al., 2021
2. Meyboom et al., 1999

Drug abuse identification through social media data analysis with LLMs

GPT3 was previously shown to expand the lexicon of colloquial drug synonyms from social media posts[1]

ChatGPT can recognize drug abuse risks from tweets



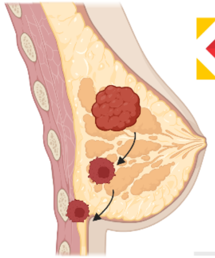
Ever since my **Acid** trips like whenever I get super high I just start—lightly hallucinating and it's tbh creepy"	⚠	⚠
"drove like 10 miles on these icy ass roads all to get some weed if imma—be locked up in my house for awhile imma need some **weed**"	⚠	⚠
"Just watched Fear and Loathing in Las Vegas for the first time—and I think I should have been on **acid** to fully understand it"	✗	✗
"Smoking a blunt at home so much better than going to the woods in—Brooksville and puking on yourself Bc you **drank too much reball**"	⚠	⚠
"today I was asked if I do **heroin** because I went to Lancaster????"	✗	✗
"Morgan told me my Bitmoji looks like a **heroin** addict?"	✗	✗

Table adapted from Hu et al., 2019

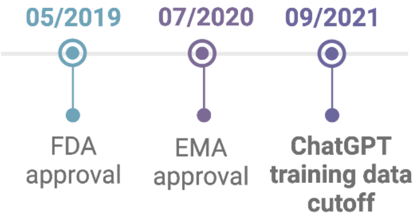
---

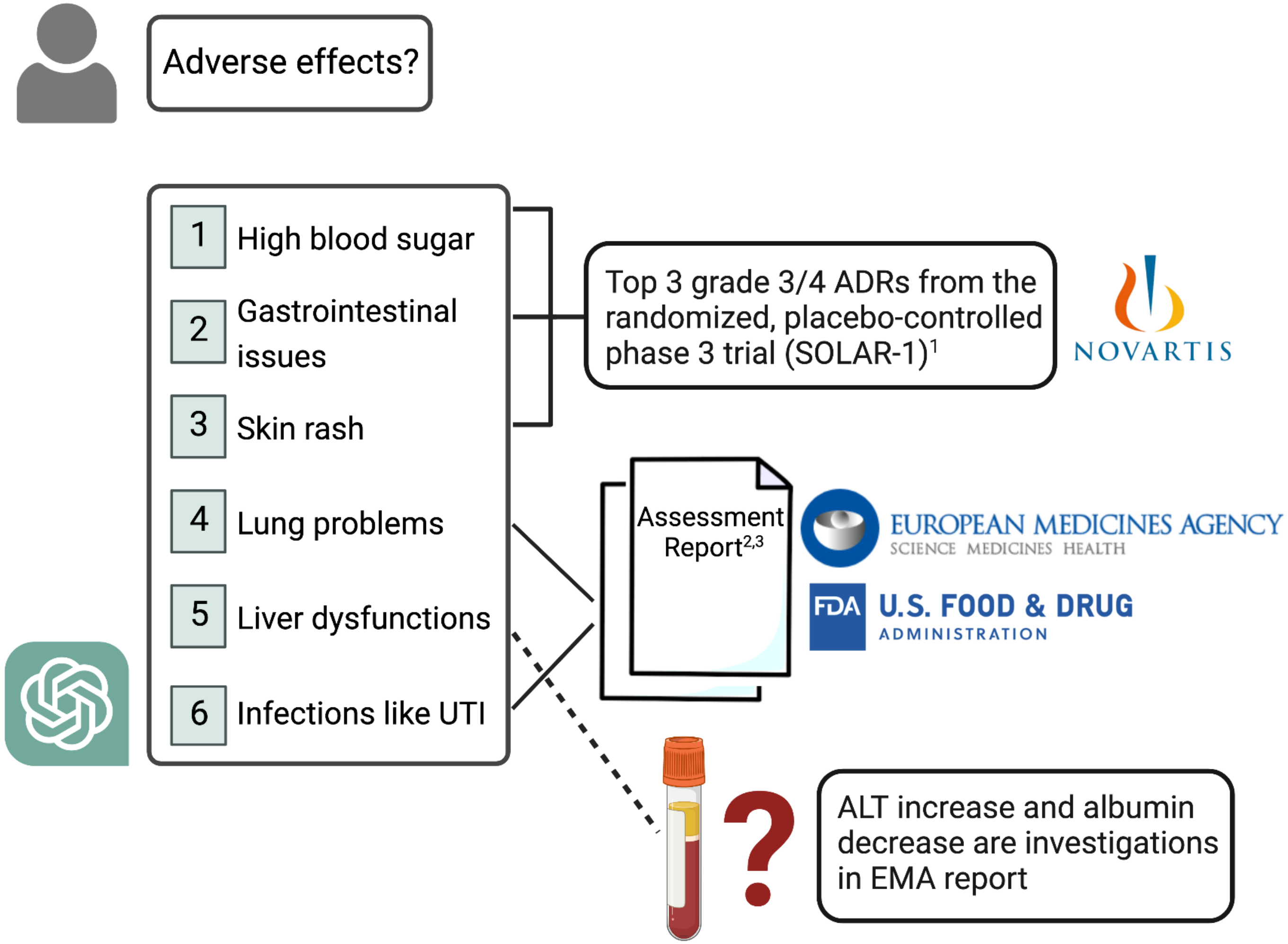
1. Carpenter et al., 2023

ADR ranking and signal detection capabilities of ChatGPT



PIQRAY®
(alpelisib) tablets
50 mg · 150 mg · 200 mg





Adverse effects?

1. High blood sugar
2. Gastrointestinal issues
3. Skin rash
4. Lung problems
5. Liver dysfunctions
6. Infections like UTI

Top 3 grade 3/4 ADRs from the randomized, placebo-controlled phase 3 trial (SOLAR-1)[1]

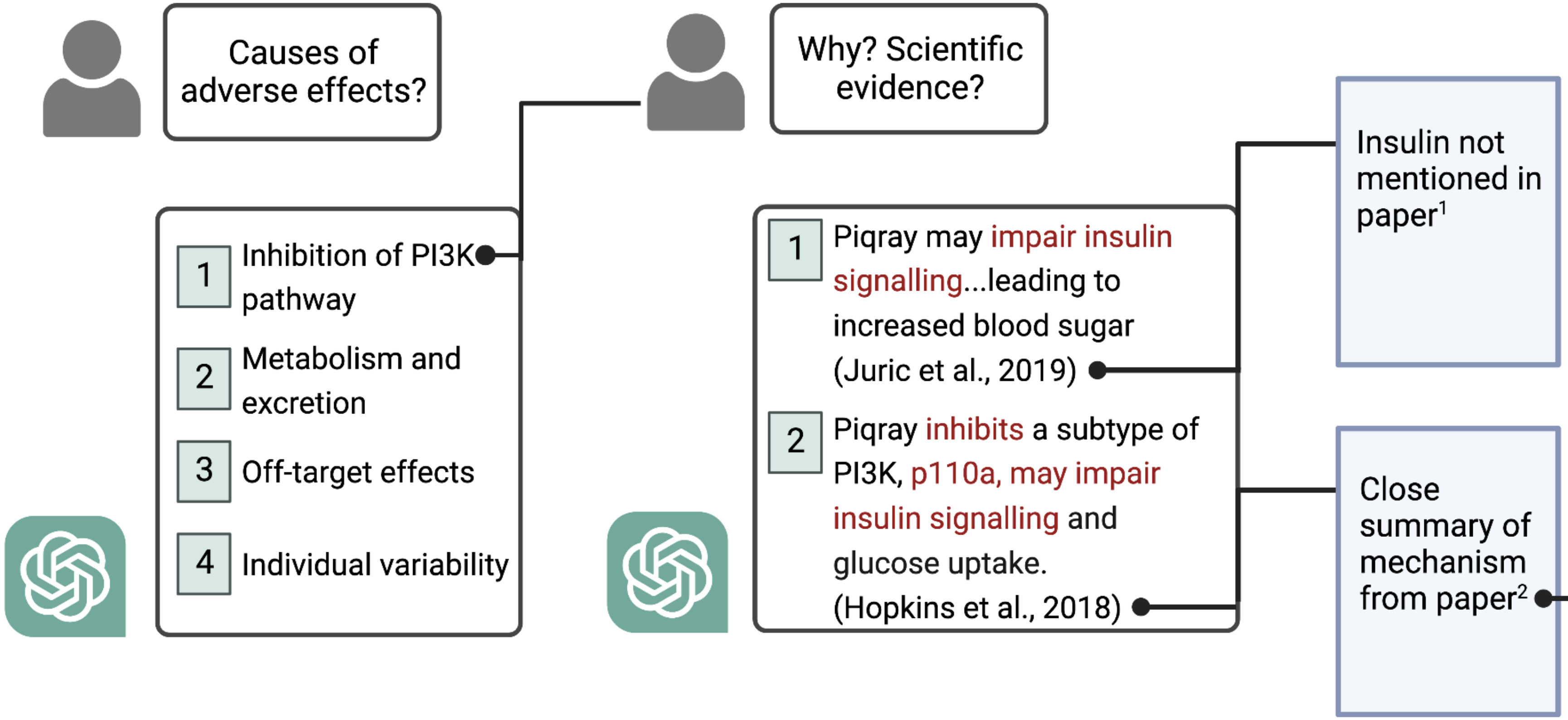
Assessment Report[2,3]

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH
FDA U.S. FOOD & DRUG ADMINISTRATION
?

ALT increase and albumin decrease are investigations in EMA report

1. Rugo et al., 2020
2. PIQRAY® (alpelisib) tablets, for oral use. 2019
3. Agency EM. Assessment report - Piqray. 2020

Using ChatGPT as a starting point for scientific evidence retrieval



Causes of adverse effects?

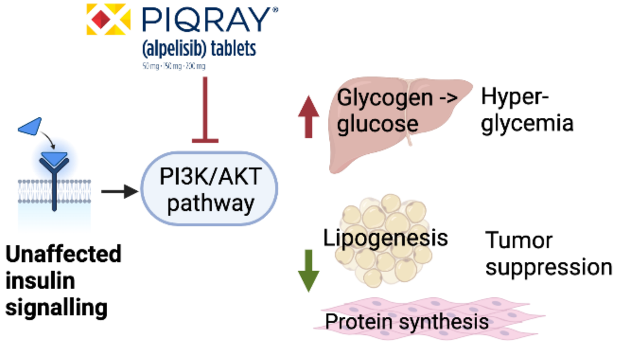
1. Inhibition of PI3K pathway
2. Metabolism and excretion
3. Off-target effects
4. Individual variability

Why? Scientific evidence?

1. Piqray may impair insulin signalling...leading to increased blood sugar (Juric et al., 2019)
2. Piqray inhibits a subtype of PI3K, p110a, may impair insulin signalling and glucose uptake. (Hopkins et al., 2018)

Insulin not mentioned in paper[1]

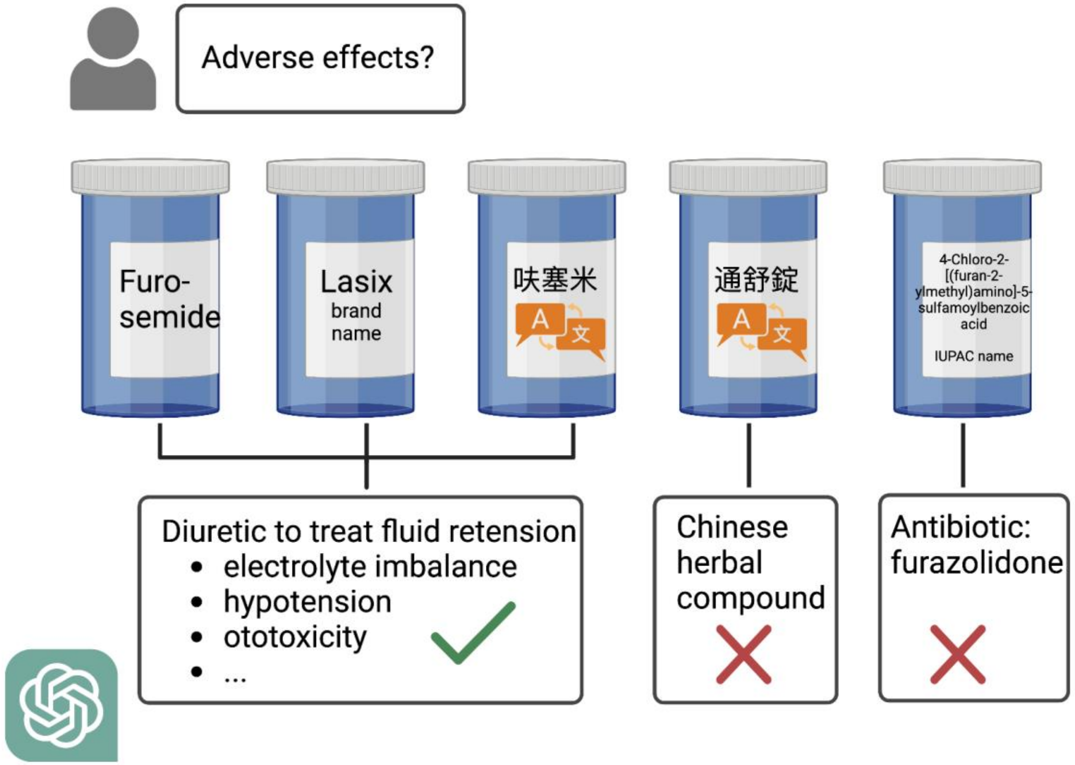
Close summary of mechanism from paper[2]



1. Jain et al., 2018
2. Hopkins et al., 2018

Potential of ChatGPT as a pharmacovigilance knowledge database

Choice of query language matters



Adverse effects?

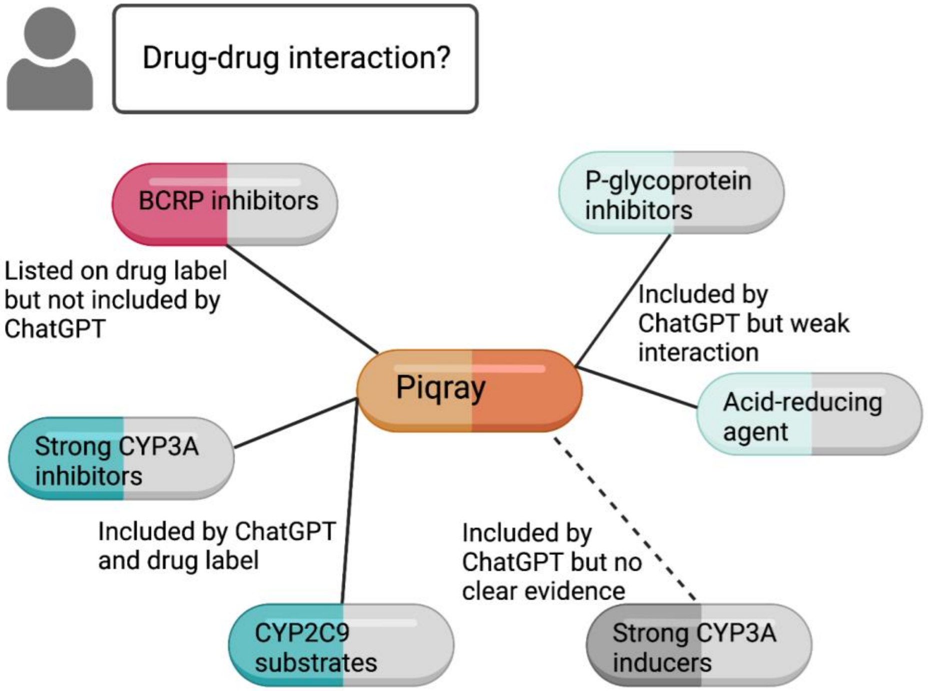
Diuretic to treat fluid retension
- electrolyte imbalance
- hypotension
- ototoxicity
- ...

✓

Chinese herbal compound ✗

Antibiotic: furazolidone ✗

Selective coverage in response



Drug-drug interaction?

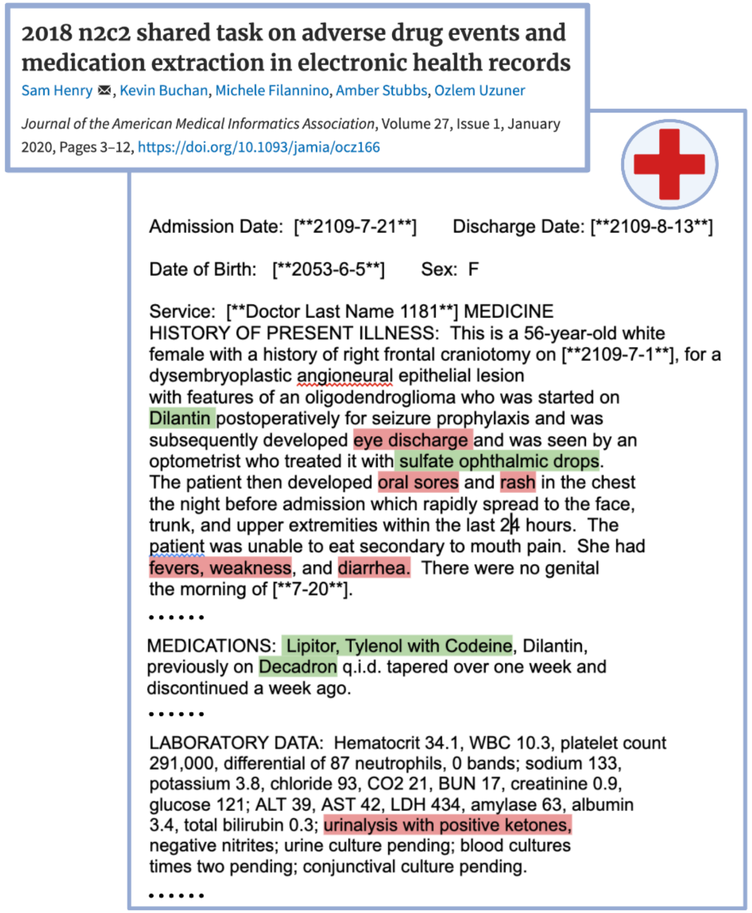
Listed on drug label but not included by ChatGPT

Included by ChatGPT but weak interaction

Included by ChatGPT and drug label

Included by ChatGPT but no clear evidence

ADR case review and text summarization capabilities of ChatGPT



2018 n2c2 shared task on adverse drug events and medication extraction in electronic health records

Sam Henry, Kevin Buchan, Michele Filannino, Amber Stubbs, Ozlem Uzuner

Journal of the American Medical Informatics Association, Volume 27, Issue 1, January 2020, Pages 3–12, https://doi.org/10.1093/jamia/ocz166

Admission Date: [**2109-7-21**] Discharge Date: [**2109-8-13**]

Date of Birth: [**2053-6-5**] Sex: F

Service: [**Doctor Last Name 1181**] MEDICINE
HISTORY OF PRESENT ILLNESS: This is a 56-year-old white female with a history of right frontal craniotomy on [**2109-7-1**], for a dysembryoplastic angioneural epithelial lesion with features of an oligodendroglioma who was started on Dilantin postoperatively for seizure prophylaxis and was subsequently developed eye discharge and was seen by an optometrist who treated it with sulfate ophthalmic drops. The patient then developed oral sores and rash in the chest the night before admission which rapidly spread to the face, trunk, and upper extremities within the last 24 hours. The patient was unable to eat secondary to mouth pain. She had fevers, weakness, and diarrhea. There were no genital the morning of [**7-20**].

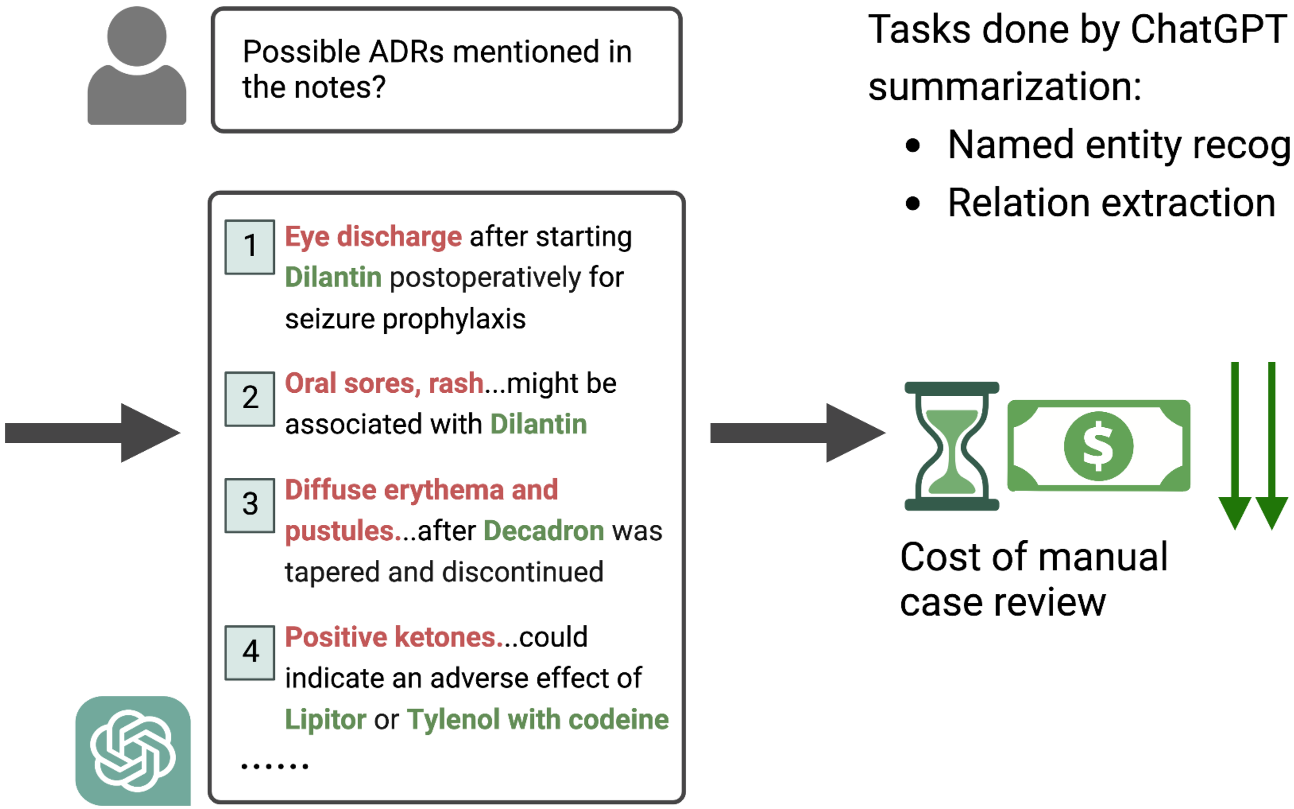
……

MEDICATIONS: Lipitor, Tylenol with Codeine, Dilantin, previously on Decadron q.i.d. tapered over one week and discontinued a week ago.

……

LABORATORY DATA: Hematocrit 34.1, WBC 10.3, platelet count 291,000, differential of 87 neutrophils, 0 bands; sodium 133, potassium 3.8, chloride 93, CO2 21, BUN 17, creatinine 0.9, glucose 121; ALT 39, AST 42, LDH 434, amylase 63, albumin 3.4, total bilirubin 0.3; urinalysis with positive ketones, negative nitrites; urine culture pending; blood cultures times two pending; conjunctival culture pending.

……



Possible ADRs mentioned in the notes?

1 **Eye discharge** after starting **Dilantin** postoperatively for seizure prophylaxis

2 **Oral sores, rash**...might be associated with **Dilantin**

3 **Diffuse erythema and pustules**...after **Decadron** was tapered and discontinued

4 **Positive ketones**...could indicate an adverse effect of **Lipitor** or **Tylenol with codeine**

……

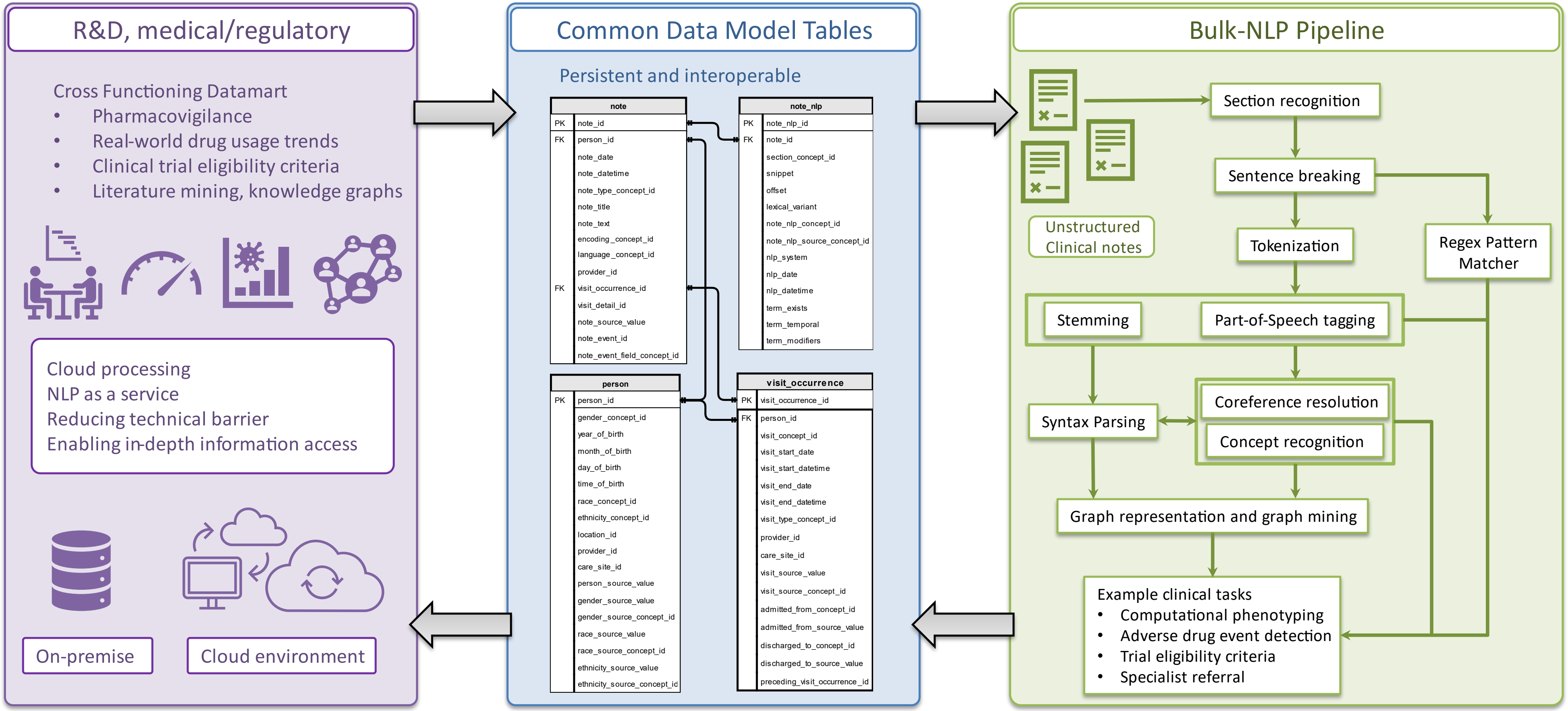
Cost of manual case review

Tasks done by ChatGPT for summarization:

- Named entity recognition
- Relation extraction

H Wang, J Ding, Y Luo. Future of ChatGPT in Pharmacovigilance. *Drug Safety, 2023 accepted*

Streamlined, interoperable and scalable biomedical NLP



R&D, medical/regulatory

Cross Functioning Datamart
- Pharmacovigilance
- Real-world drug usage trends
- Clinical trial eligibility criteria
- Literature mining, knowledge graphs

Cloud processing
NLP as a service
Reducing technical barrier
Enabling in-depth information access

On-premise

Cloud environment

Common Data Model Tables

Persistent and interoperable

	note
PK	note_id
FK	person_id
	note_date
	note_datetime
	note_type_concept_id
	note_title
	note_text
	encoding_concept_id
	language_concept_id
	provider_id
FK	visit_occurrence_id
	visit_detail_id
	note_source_value
	note_event_id
	note_event_field_concept_id

	note_nlp
PK	note_nlp_id
FK	note_id
	section_concept_id
	snippet
	offset
	lexical_variant
	note_nlp_concept_id
	note_nlp_source_concept_id
	nlp_system
	nlp_date
	nlp_datetime
	term_exists
	term_temporal
	term_modifiers

	person
PK	person_id
	gender_concept_id
	year_of_birth
	month_of_birth
	day_of_birth
	time_of_birth
	race_concept_id
	ethnicity_concept_id
	location_id
	provider_id
	care_site_id
	person_source_value
	gender_source_value
	gender_source_concept_id
	race_source_value
	race_source_concept_id
	ethnicity_source_value
	ethnicity_source_concept_id

	visit_occurrence
PK	visit_occurrence_id
FK	person_id
	visit_concept_id
	visit_start_date
	visit_start_datetime
	visit_end_date
	visit_end_datetime
	visit_type_concept_id
	provider_id
	care_site_id
	visit_source_value
	visit_source_concept_id
	admitted_from_concept_id
	admitted_from_source_value
	discharged_to_concept_id
	discharged_to_source_value
	preceding_visit_occurrence_id

Bulk-NLP Pipeline

Unstructured Clinical notes

Section recognition

Sentence breaking

Tokenization

Regex Pattern Matcher

Stemming

Part-of-Speech tagging

Syntax Parsing

Coreference resolution

Concept recognition

Graph representation and graph mining

Example clinical tasks
- Computational phenotyping
- Adverse drug event detection
- Trial eligibility criteria
- Specialist referral

Scalability

Interoperability

Streamlining

PMID: 30591037

PMID: 29854260

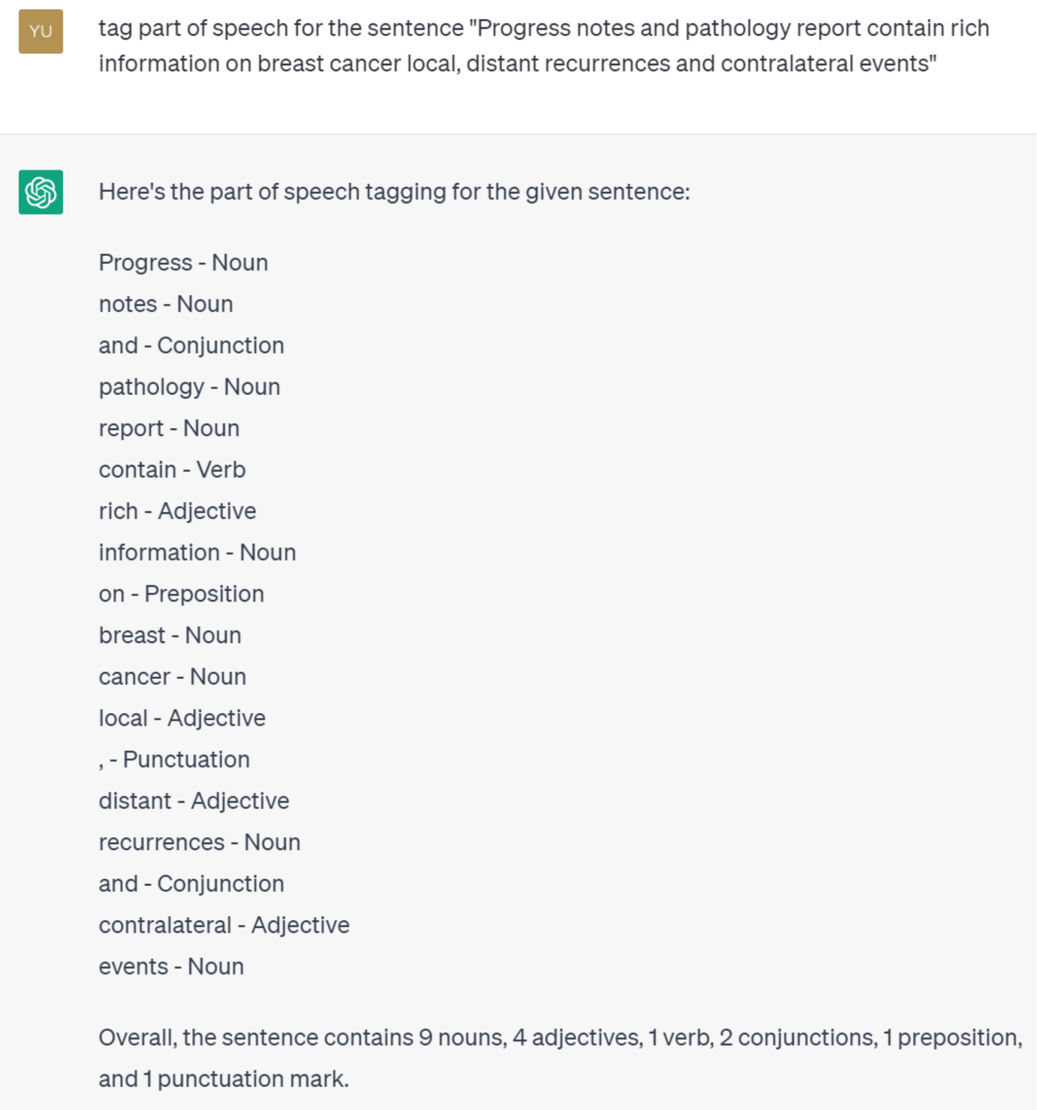
PMID: 32308812

PMID: 32308871

PMID: 30943974

Democratize AI/ML literacy, tooling and results to lower access barriers for partners

How much can LLM help?

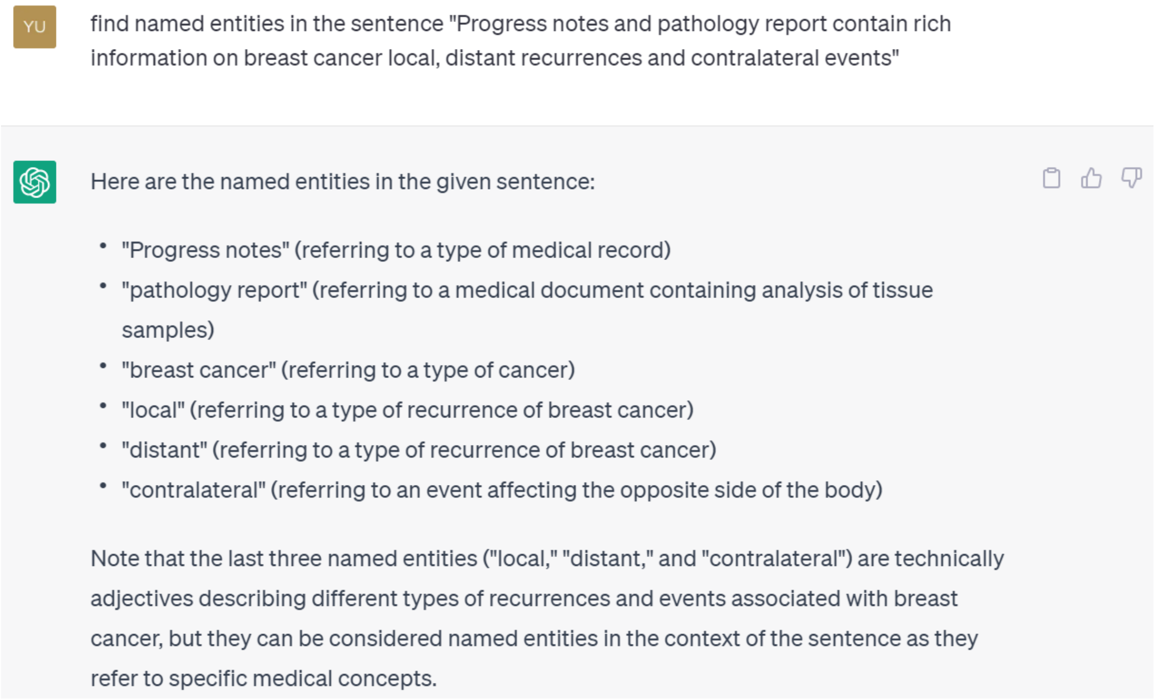


tag part of speech for the sentence "Progress notes and pathology report contain rich information on breast cancer local, distant recurrences and contralateral events"

Here's the part of speech tagging for the given sentence:

Progress - Noun
notes - Noun
and - Conjunction
pathology - Noun
report - Noun
contain - Verb
rich - Adjective
information - Noun
on - Preposition
breast - Noun
cancer - Noun
local - Adjective
, - Punctuation
distant - Adjective
recurrences - Noun
and - Conjunction
contralateral - Adjective
events - Noun

Overall, the sentence contains 9 nouns, 4 adjectives, 1 verb, 2 conjunctions, 1 preposition, and 1 punctuation mark.



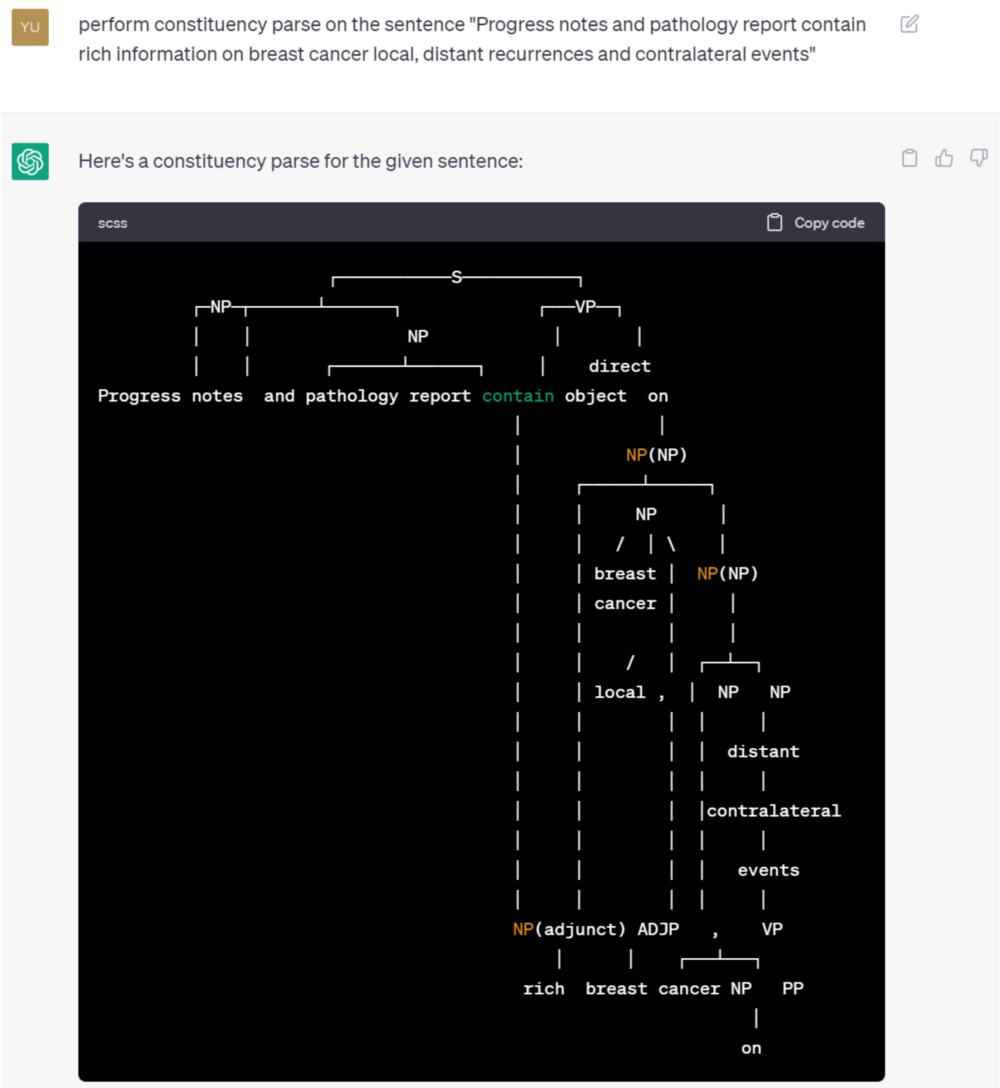
find named entities in the sentence "Progress notes and pathology report contain rich information on breast cancer local, distant recurrences and contralateral events"

Here are the named entities in the given sentence:

- "Progress notes" (referring to a type of medical record)
- "pathology report" (referring to a medical document containing analysis of tissue samples)
- "breast cancer" (referring to a type of cancer)
- "local" (referring to a type of recurrence of breast cancer)
- "distant" (referring to a type of recurrence of breast cancer)
- "contralateral" (referring to an event affecting the opposite side of the body)

Note that the last three named entities ("local," "distant," and "contralateral") are technically adjectives describing different types of recurrences and events associated with breast cancer, but they can be considered named entities in the context of the sentence as they refer to specific medical concepts.

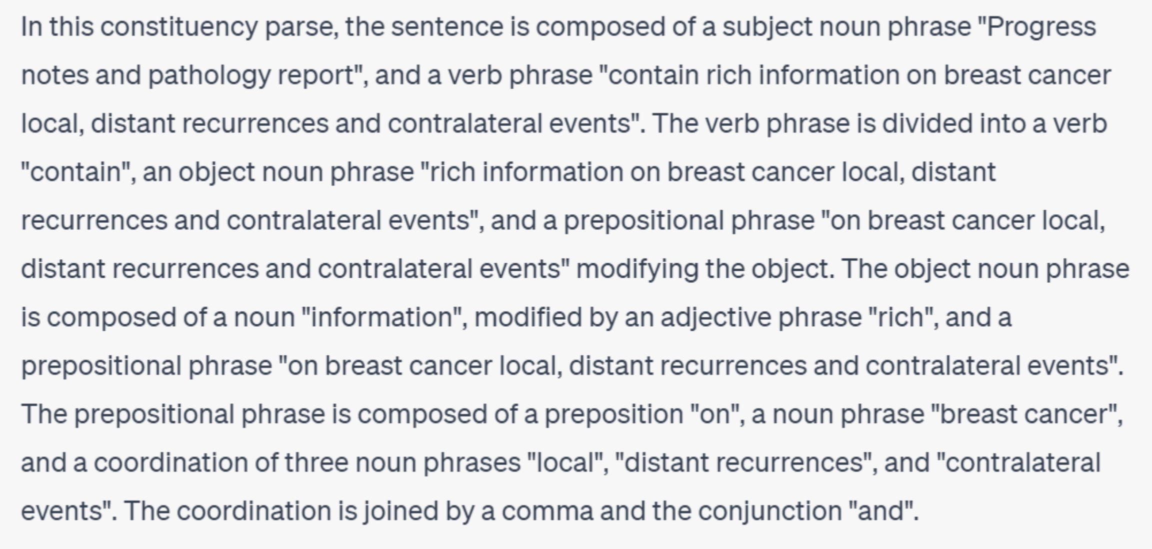
How much can LLM help?



perform constituency parse on the sentence "Progress notes and pathology report contain rich information on breast cancer local, distant recurrences and contralateral events"

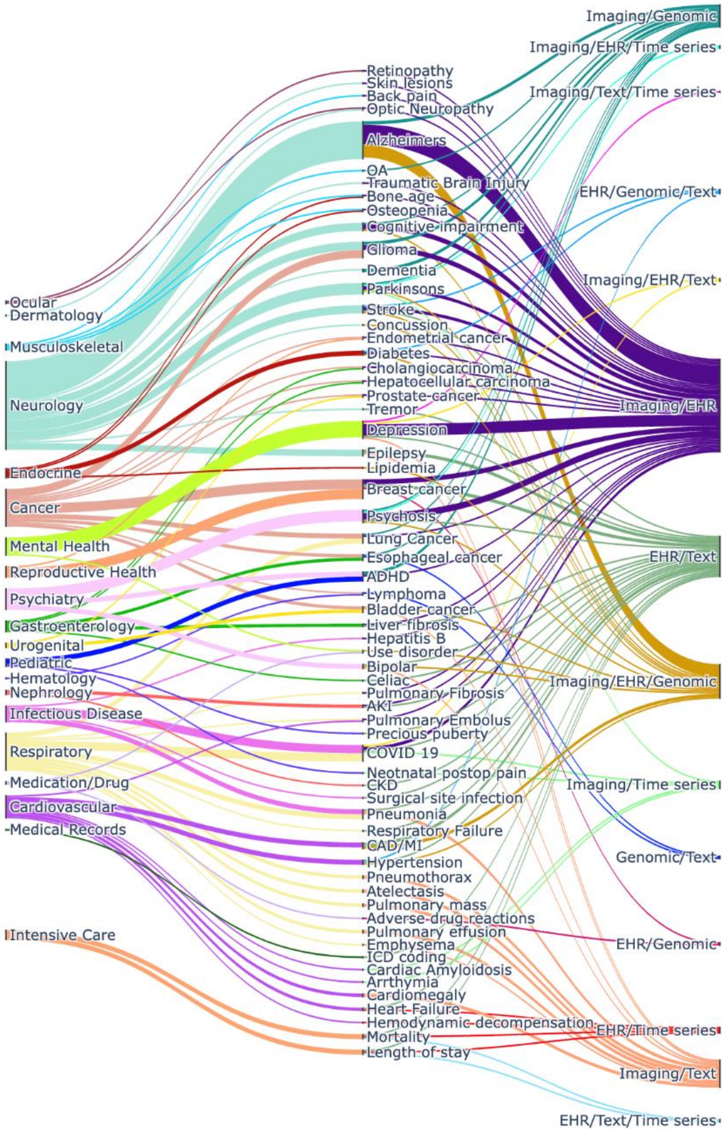
Here's a constituency parse for the given sentence:

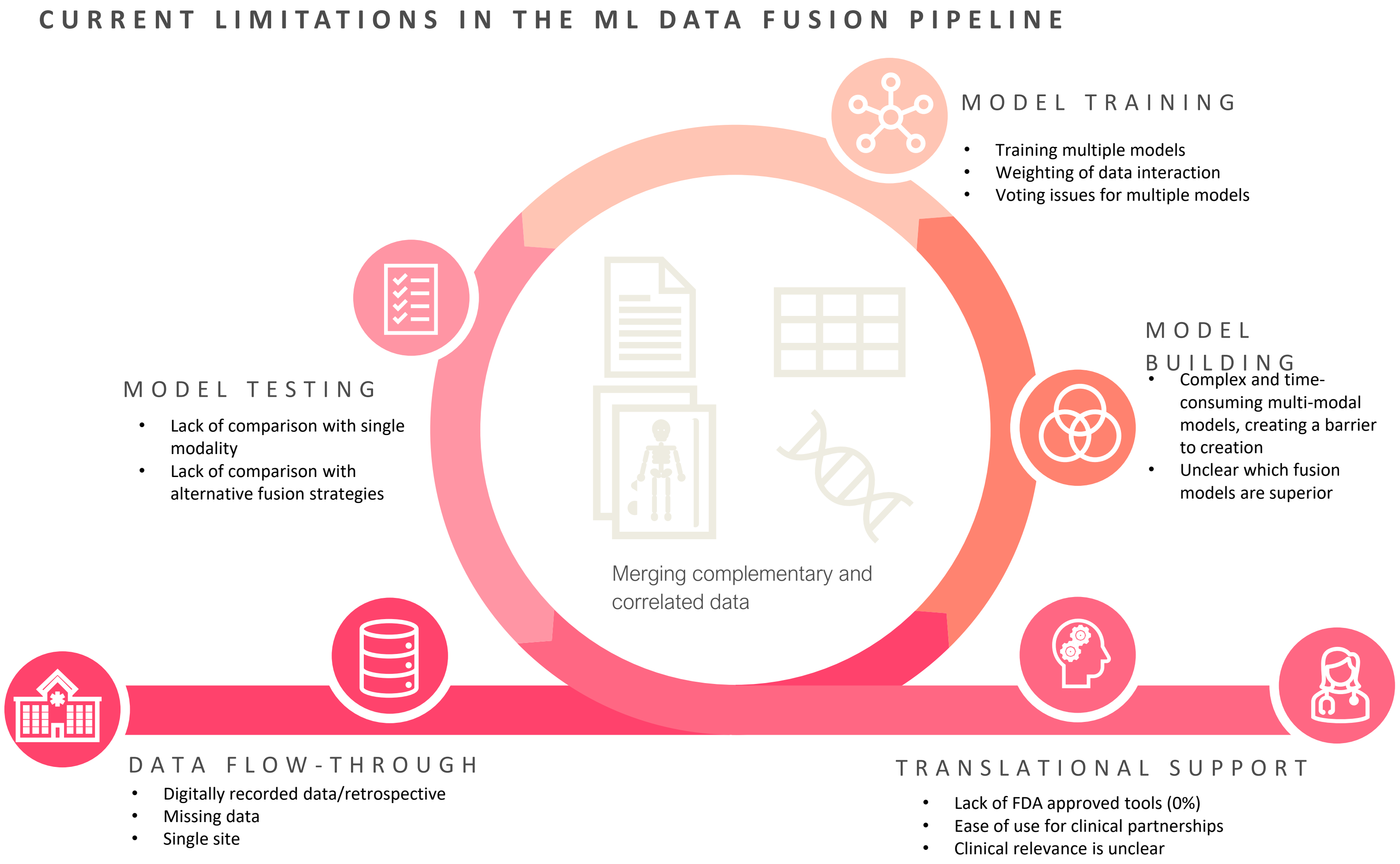
```
                       ┌───────────S───────────┐
          ┌NP─┬───────┴──────┐             ┌──VP──┐
          |   |               NP            |       |
          |   |       ┌───────┴──────┐     |   direct
Progress notes  and pathology report contain object  on
                                         |            |
                                         |         NP(NP)
                                         |     ┌─────┴──────┐
                                         |     |     NP     |
                                         |     |   /  | \   |
                                         |     | breast |  NP(NP)
                                         |     | cancer |    |
                                         |     |        |    |
                                         |     |    /   | ┌──┴──┐
                                         |     | local , |  NP   NP
                                         |     |        | |     |
                                         |     |        | |  distant
                                         |     |        | |     |
                                         |     |        | |contralateral
                                         |     |        | |     |
                                         |     |        | |   events
                                         |     |        | |     |
                                        NP(adjunct) ADJP    ,   VP
                                             |      |    ┌──┴───┐
                                         rich  breast cancer NP   PP
                                                               |
                                                              on
```



In this constituency parse, the sentence is composed of a subject noun phrase "Progress notes and pathology report", and a verb phrase "contain rich information on breast cancer local, distant recurrences and contralateral events". The verb phrase is divided into a verb "contain", an object noun phrase "rich information on breast cancer local, distant recurrences and contralateral events", and a prepositional phrase "on breast cancer local, distant recurrences and contralateral events" modifying the object. The object noun phrase is composed of a noun "information", modified by an adjective phrase "rich", and a prepositional phrase "on breast cancer local, distant recurrences and contralateral events". The prepositional phrase is composed of a preposition "on", a noun phrase "breast cancer", and a coordination of three noun phrases "local", "distant recurrences", and "contralateral events". The coordination is joined by a comma and the conjunction "and".

LLM to integrate multi-modal healthcare data





CURRENT LIMITATIONS IN THE ML DATA FUSION PIPELINE

MODEL TRAINING

- Training multiple models
- Weighting of data interaction
- Voting issues for multiple models

MODEL BUILDING

- Complex and time-consuming multi-modal models, creating a barrier to creation
- Unclear which fusion models are superior

MODEL TESTING

- Lack of comparison with single modality
- Lack of comparison with alternative fusion strategies

Merging complementary and correlated data

DATA FLOW-THROUGH

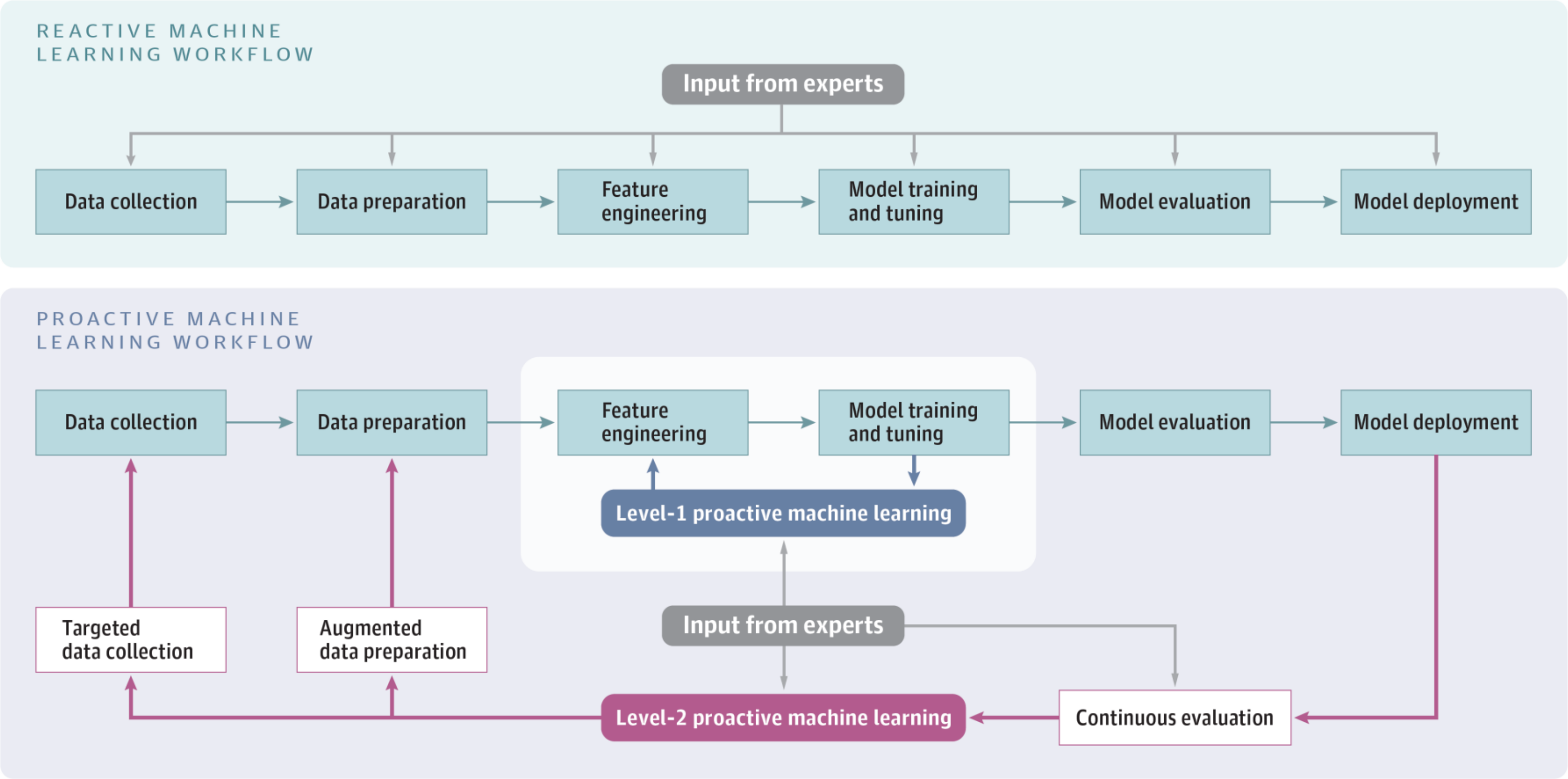
- Digitally recorded data/retrospective
- Missing data
- Single site

TRANSLATIONAL SUPPORT

- Lack of FDA approved tools (0%)
- Ease of use for clinical partnerships
- Clinical relevance is unclear

Kline A, Wang H, Li Y, Dennis S, Hutch M, Xu Z, Wang F, Cheng F, Luo Y. Multimodal Machine Learning in Precision Health: a Scoping Review. *npj Digital Medicine 2022 Nov 7;5(1):1-4*

LLMs can help us move from reactive to proactive machine learning



Dialogue LLMs such as ChatGPT fits perfectly in the proactive ML paradigm

Luo Y, Wunderink RG, Lloyd-Jones D. Proactive vs Reactive Machine Learning in Health Care: Lessons From the COVID-19 Pandemic. *JAMA. 2022.*

Let us work together and bring it to a whole new level

- Collaboration welcome
- yuan.luo@northwestern.edu
- @yuanhypnosluo
- We are hiring, multiple postdoc positions available
- https://labs.feinberg.northwestern.edu/luolab/
- Main funding support acknowledgement
    - U01TR003528
    - U54HL160273
    - R01LM013337
    - R01GM105688
    - R21LM012618
    - UL1TR001422
    - U01HG011169

NUCATS Clinical and Translational Sciences Institute

Northwestern

Northwestern Medicine

eMERGE Network electronic medical records & genomics