ANALYSIS GROUP

Advancing HEOR and RWE Research with Generative AI

ISPOR, May 7, 2024

ANALYSIS GROUP

Moderator and Speakers

Eric Wu

Managing Principal

Analysis Group, Inc.

Jimmy Royer

Principal and Director of Data Science

Analysis Group, Inc.

Rajeev Ayyagari

Vice President

Analysis Group, Inc.

Hua Xu

Professor

Yale School of Medicine

ANALYSIS GROUP

Advancing HEOR and RWE Research with Generative AI

Agenda

1. **Predictive Modeling, Digital Twins and Dynamic Disease Model**
2. **AGHealth.ai™: Generative AI Applications**
3. **Biomedical Large Language Models: Development and Application**

ANALYSIS GROUP

Quick History of AI: key stages

1800
Origins in popular culture and literature

1950
Computing machines are developed and improved: Hollerith, difference analyzer, ENIAC

1950
Start of symbolic and rule-based AI

1974
First AI winter

1970s
Expert systems become popular

1980
Loss of interest in symbolic AI

1987
Second AI winter starts

1990
Decreased interest in expert systems

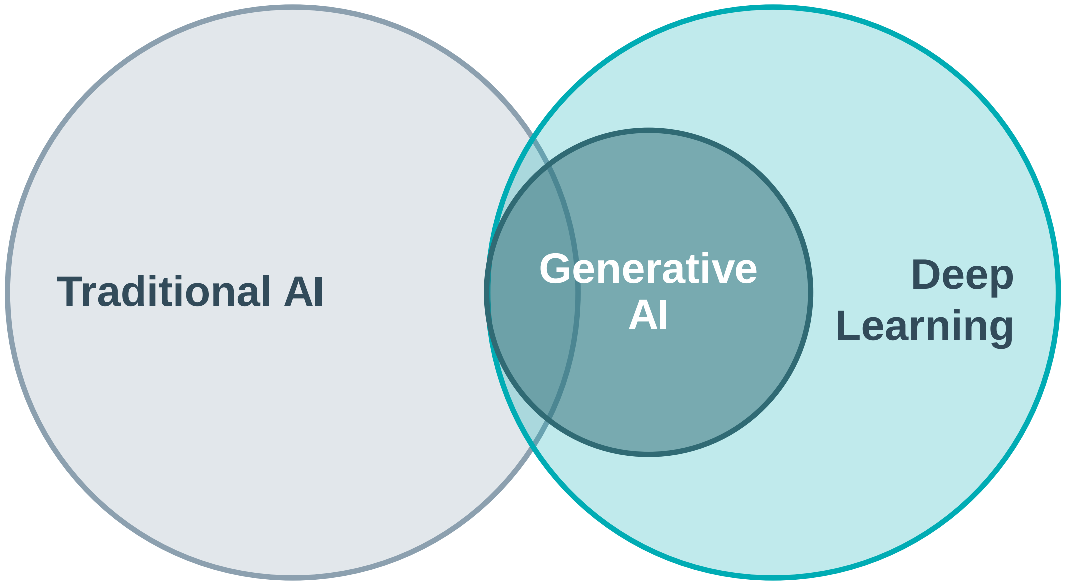
1990
Predictive AI

2005
Deep learning takes off

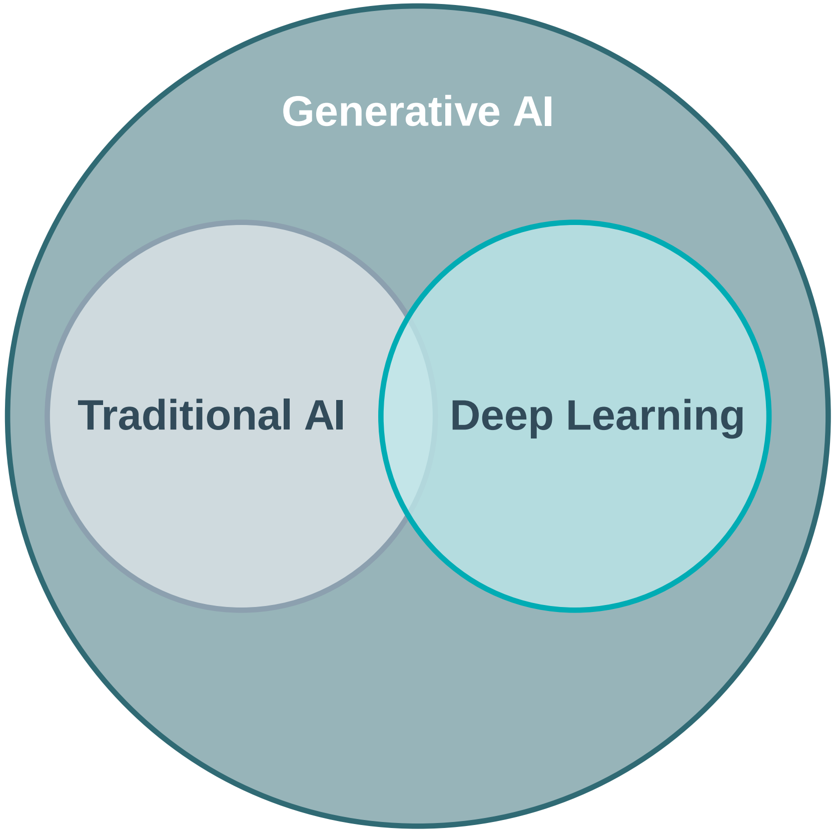
2020
Generative AI, large models

Key differences: traditional vs generative AI

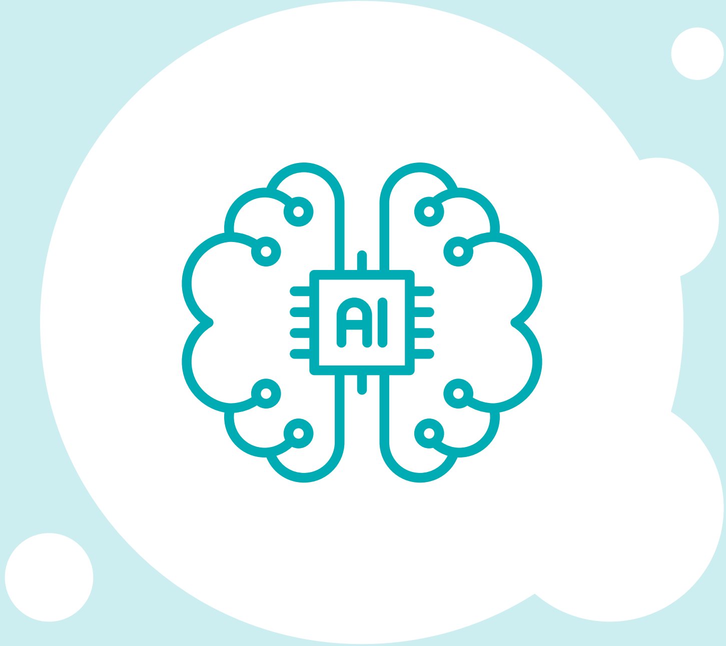
Model types



Capabilities

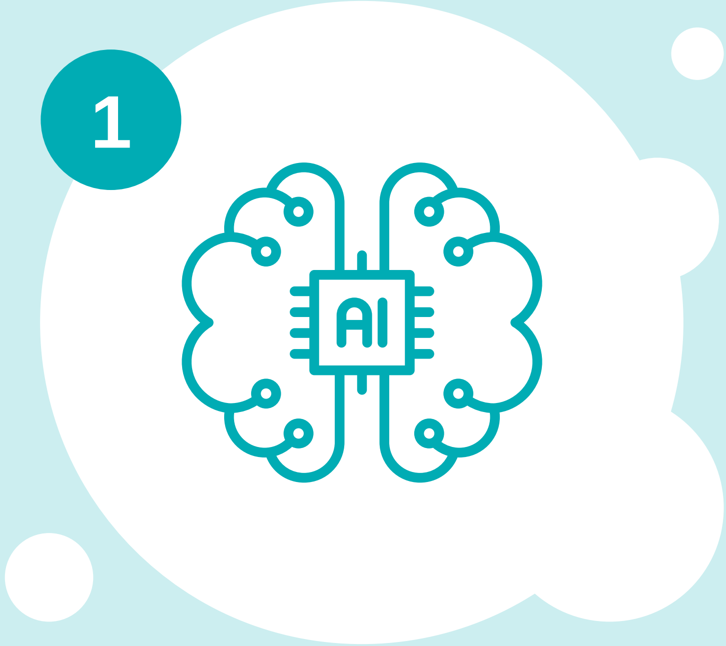


1. Predictive Modeling, Digital Twins and Dynamic Disease Model



ANALYSIS GROUP

AI
Predictive Analytics



ANALYSIS GROUP

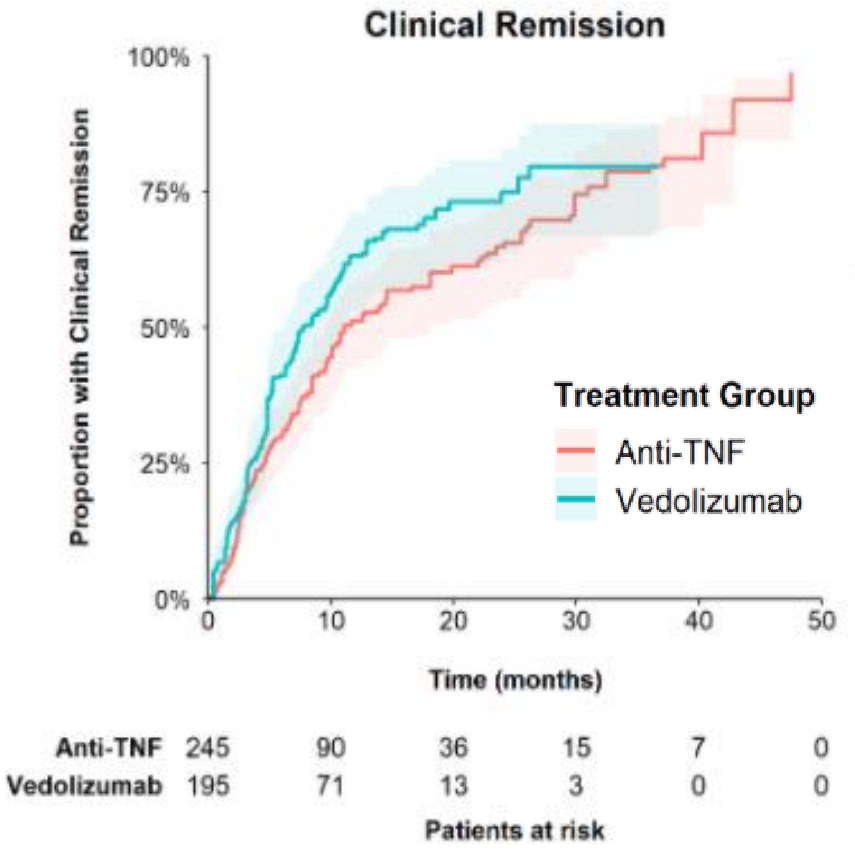
Case example: identify patients with Crohn's disease with higher likelihood of remission when treated with VDZ

Challenge: lack of differentiation between VDZ and anti-TNFα in a real-world study

- Median time to remission; 7.8 months vs. 11.1 months (HR: 1.14 [0.89, 1.45])

Solutions:

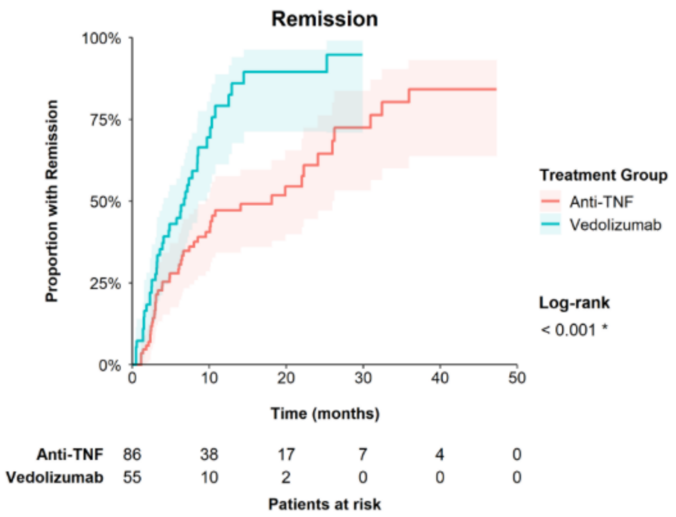
- Used machine learning and a non-parametric method to identify and validate high-value patient subpopulations in which VDZ had significantly better efficacy compared to anti-TNFα
- Created simplified rules to identify patients in the high-value subpopulations that can be easily implemented in clinical settings



Case example: identify patients with Crohn's disease with higher likelihood of remission when treated with VDZ

High-value subpopulation based on top 32% of patients

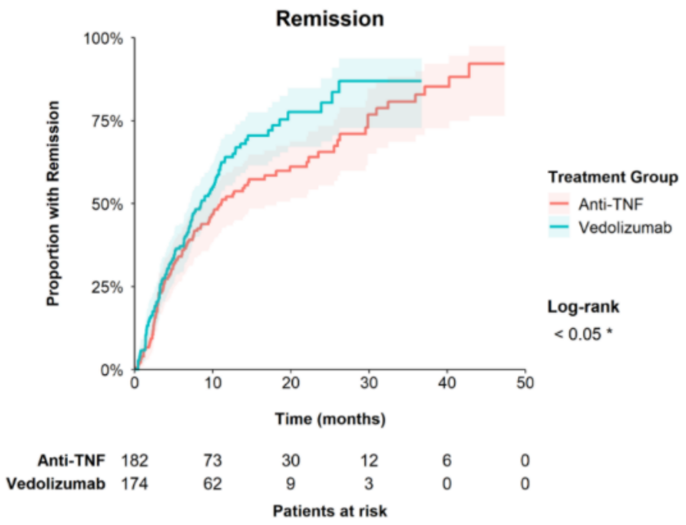
Patients with an exacerbation ongoing at index
→ Yes →
Patients with ED/ER visits prior to index
→ No →
Pre-index disease behavior was stricturing with/without perianal disease
→ No →
Treat with VDZ



Median time to remission; 6.7 months vs. 18.1 months (HR: **2.9; p<0.001**)

High-value subpopulation based on top 81% of patients

Patients with ED/ER visits prior to index
→ No →
Patients with fistulae at most recent assessment prior to index event
→ No →
Treat with VDZ



Median time to remission; 8.5 months vs. 11.1 months (HR: **1.7; p<0.05**)

https://academic.oup.com/ecco-jcc/article/15/Supplement_1/S095/6286610

ANALYSIS GROUP

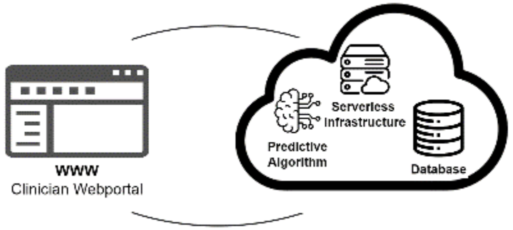
Predicting tuberculosis drug-resistance using wide and deep neural networks

Methods

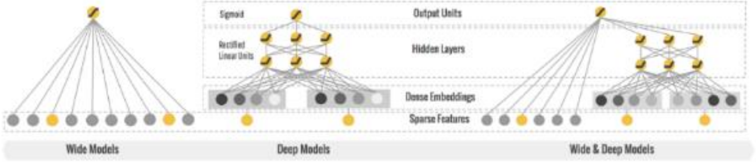
- Applied a novel multidrug wide and deep neural network to estimate individual TB treatment resistance profiles given the DNA of the bacteria
- Developed supervised ML algorithms with academic collaborators

Results & impact

The online tool can be used by physicians to measure individualized resistance profile predictions and determine the right TB treatment



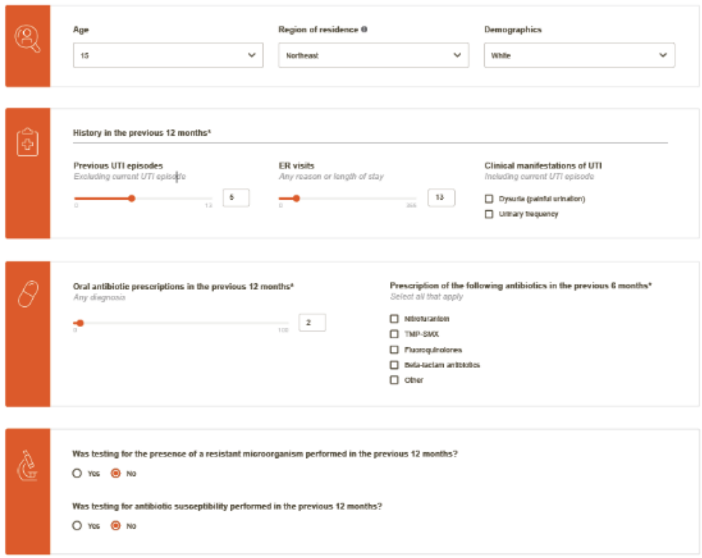
Platform developed and hosted on the cloud

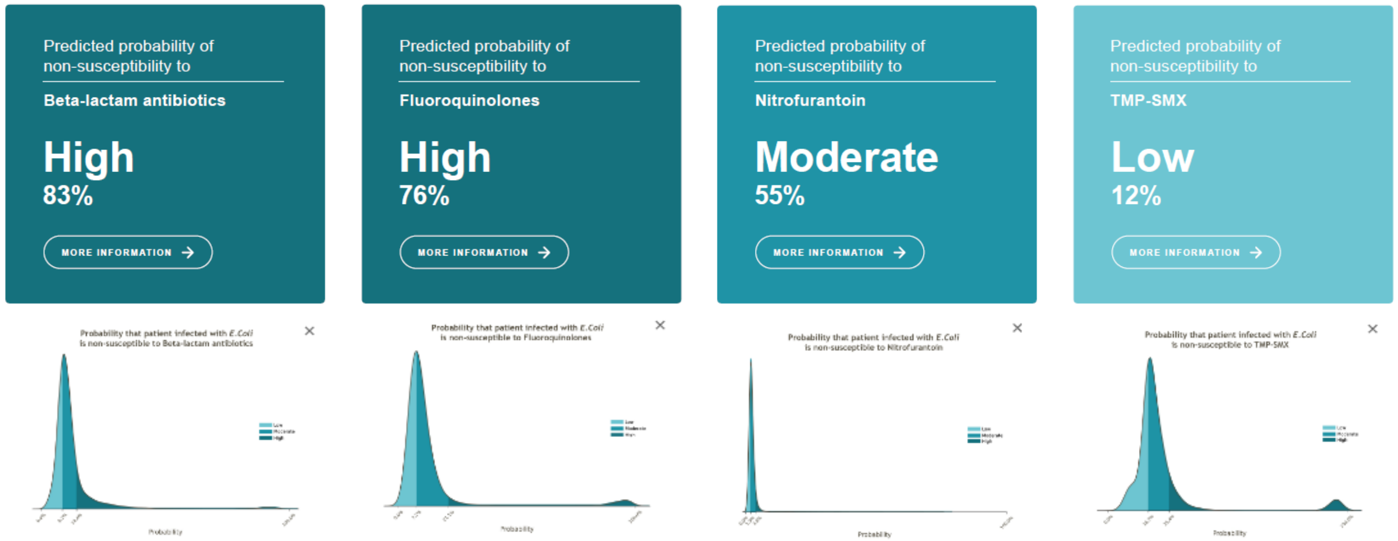


Chen ML et al. Beyond multidrug resistance: Leveraging rare variants with machine and statistical learning models in Mycobacterium tuberculosis resistance prediction. EBioMedicine. 2019; 43:356-369. doi: 10.1016/j.ebiom.2019.04.016

ANALYSIS GROUP

Predicting which patients with uUTI are at risk for antibiotic resistance: A physician's companion tool

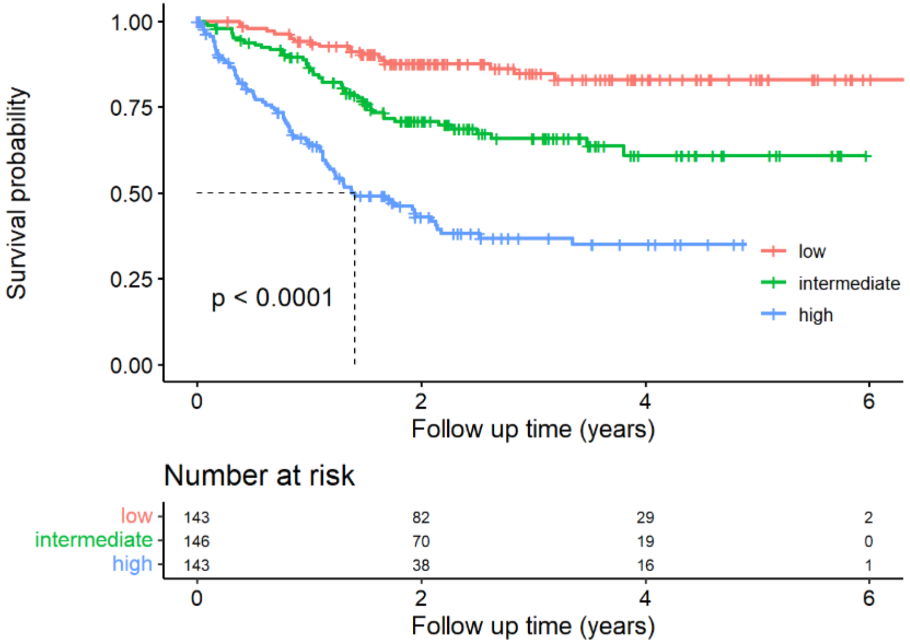




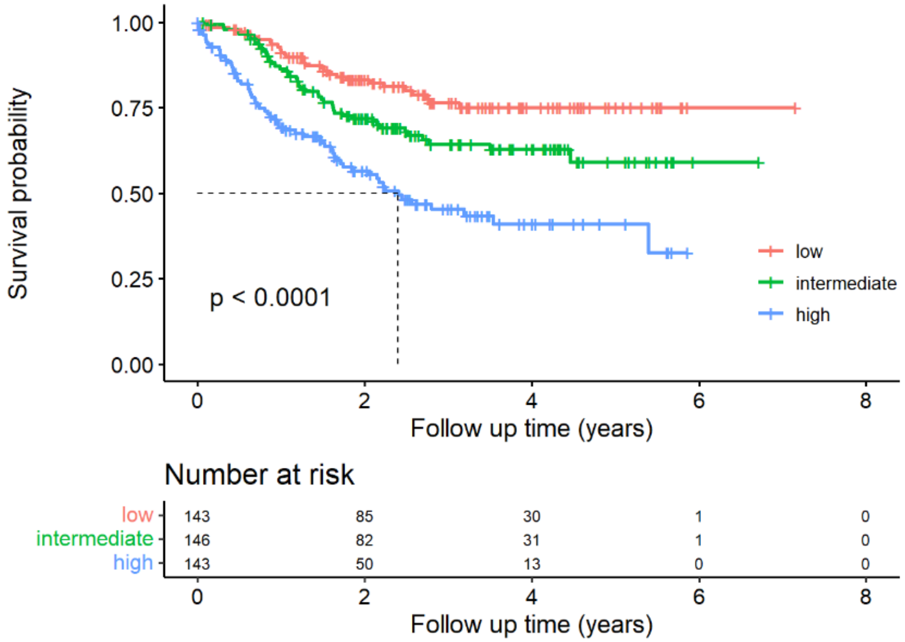
AI-powered prediction model of AML patient risks

Kaplan-Meier curves showed a significant difference in observed survival probabilities, stratified by 33% high, 34% intermediate, 33% low risk categories

Training set (N=432)



Validation set (N=432)



ANALYSIS GROUP

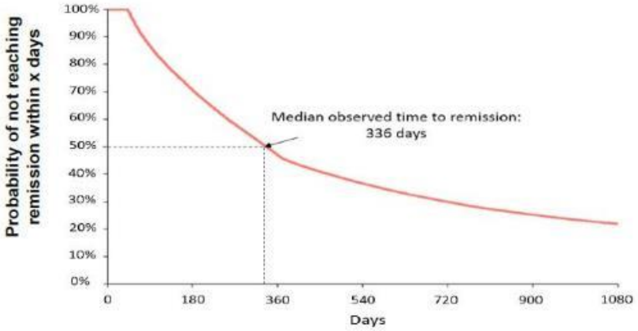
Predicting time to natural remission in chronic urticaria using random survival forest models

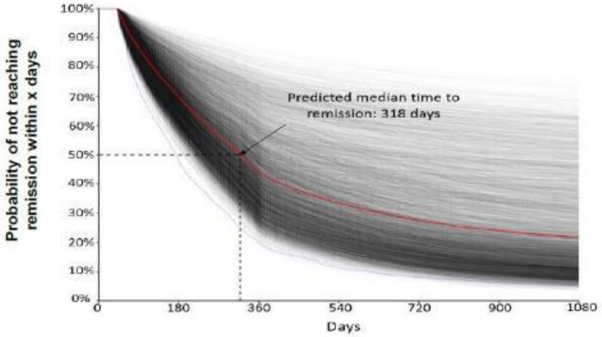
Methods

- Analyzed Optum Electronic Health Records data on adult patients with CU
- Defined clinical remission as ≥12 months free of CU diagnosis or treatment after the episode
- Applied a random survival forest method to predict the time to natural remission
- Characterized the observed time to natural remission using the Kaplan-Meier method

Results & impact

This was the first study to successfully apply machine learning methods to identify important variables and predict time to clinical remission in CU using real-world medical data





Pivneva I et al. Predicting Clinical Remission of Chronic Urticaria Using Random Survival Forests: Machine Learning Applied to Real-World Data. Dermatology and Therapy. 2022; 12(12), 2747–2763. https://doi.org/10.1007/s13555-022-00827-6

2

Digital Twins

Using Generative AI To Predict Patients Outcomes

Generative AI models provide a robust and unified approach to simulate patients' medical records under different baseline scenarios

Possible applications

- Simulate control group (digital twin) in single arm clinical trials
  - It can also help decreasing the required sample size of control groups in double-arm clinical trials
- Simulate long term patients' journey
  - Chronic conditions: diabetes, obesity
  - Complex disease progression: Cancer, Alzheimer's disease
  - Predict long term benefits of treatments
- Predict comparative effectiveness of new treatments
- Define treatment targets
- Find new indications for already approved molecules

Types of models and uses

- Restricted Conditional Boltzmann Machines
  - Digital Twins (phase 2 and 3 clinical trials)
  - Qualified by the EMA in 2022
- Recurrent Neural Networks
  - Generating Synthetic Longitudinal Health Data
- Generative Adversarial Networks (GANs), Auto-Encoders (AEs)

ANALYSIS GROUP

A Unified Approach To Learn Patients Medical Data

Generative AI can learn all the interactions across all the clinical records of the patients

1 Selected group of patients

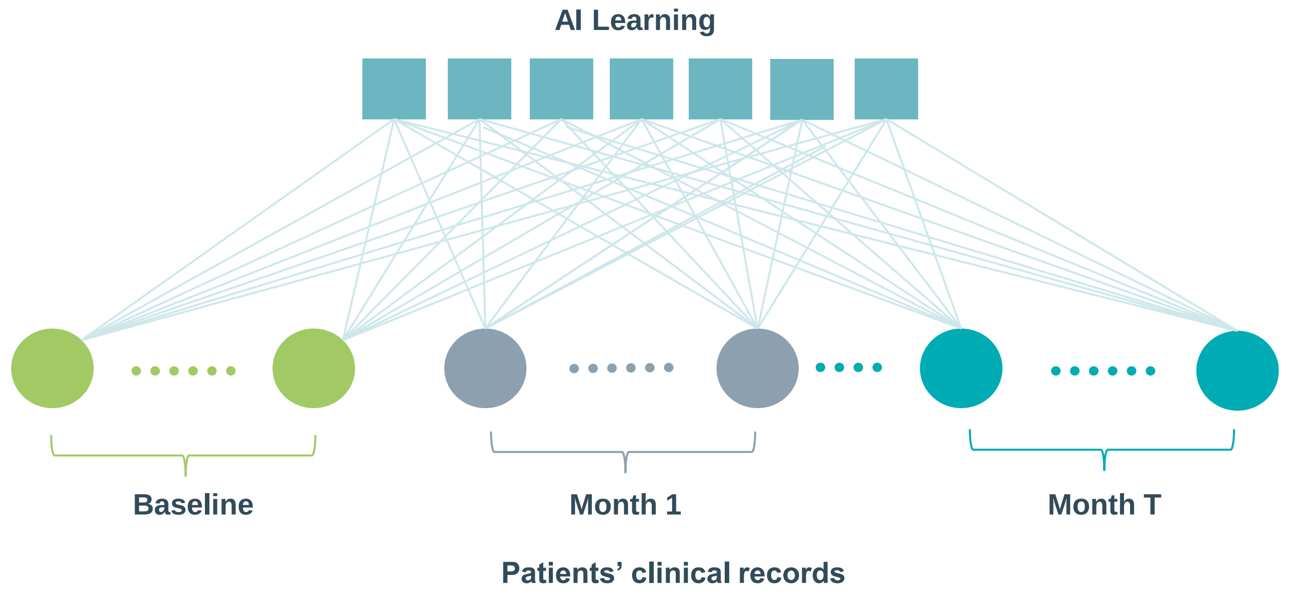
Patients' medical records

EMR Data

Variables	Baseline	t = 1	…	t = T
Time Independent	X_0	X_0	…	X_0
Lab values	Lab_0	Lab_1	…	Lab_T
ICD codes	ICD_0	ICD_1	…	ICD_T
…	…	…	…	…
CPT codes	CPT_0	CPT_1	…	CPT_T

2 Learning from the data

An AI model trained on the patients' data to learn the evolution of patients' clinical records over time, including clinical outcomes



Case 1: Simulate Treated Alzheimer Patients' Under SOC Conditions

3 Recreate Treated Patients' data

Simulate the medical records of a group of treated patients under soc conditions (digital twin – disease progression in Alzheimer patients)

Simulated Patients' data

AI Generated Counterfactual: Predicted EMR

Variables	Baseline	t = 1	…	t = T
Time Independent	X_0	X_0	…	X_0
Lab values	$Pr(Lab_0)$	$Pr(Lab_1)$	…	$Pr(Lab_T)$
ICD codes	$Pr(ICD_0)$	$Pr(ICD_1)$	…	$Pr(ICD_T)$
…	…	…	…	…
CPT codes	$Pr(CPT_0)$	$Pr(CPT_1)$	…	$Pr(CPT_T)$

4 Disease Progression in Alzheimer Patients

Evaluate the changes in target clinical outcomes

Treated group under SOC conditions

Variables	Baseline	t = 1	…	t = T
Time Independent	X_0	X_0	…	X_0
Lab values	Lab_0	Lab_1	…	Lab_T
ICD codes	ICD_0	ICD_1	…	ICD_T
…	…	…	…	…
CPT codes	CPT_0	CPT_1	…	CPT_T

=?

Treated Group Outcomes

Variables	Baseline	t = 1	…	t = T
Time Independent	X_0	X_0	…	X_0
Lab values	$Pr(Lab_0)$	$Pr(Lab_1)$	…	$Pr(Lab_T)$
ICD codes	$Pr(ICD_0)$	$Pr(ICD_1)$	…	$Pr(ICD_T)$
…	…	…	…	…
CPT codes	$Pr(CPT_0)$	$Pr(CPT_1)$	…	$Pr(CPT_T)$

Source: Modeling Disease Progression in Mild Cognitive Impairment and Alzheimer's Disease with Digital Twins (2020)

ANALYSIS GROUP

Case 2: Simulate Future Patients' Medical Records

3 Predict Patients' data

Predict futures medical records for a group of patients with limited history at baseline, using overlapping cohorts

Simulated Patients' data

AI Generated Predicted EMR

Variables	Baseline	t = T+1	…	t = T+…
Time Independent	X_0	X_0	…	X_0
Lab values	$Pr(Lab_T)$	$Pr(Lab_{T+1})$	…	$Pr(Lab_{T+...})$
ICD codes	$Pr(ICD_T)$	$Pr(ICD_{T+1})$	…	$Pr(ICD_{T+...})$
…	…	…	…	…
CPT codes	$Pr(CPT_T)$	$Pr(CPT_{T+1})$	…	$Pr(CPT_{T+...})$

4 Evolution of long-term chronic condition

Predict future clinical outcomes

Predicted patients' data

Variables	Baseline	t = T+1	…	t = T+…
Time Independent	X_0	X_0	…	X_0
Lab values	$Pr(Lab_T)$	$Pr(Lab_{T+1})$	…	$Pr(Lab_{T+...})$
ICD codes	$Pr(ICD_T)$	$Pr(ICD_{T+1})$	…	$Pr(ICD_{T+...})$
…	…	…	…	…
CPT codes	$Pr(CPT_T)$	$Pr(CPT_{T+1})$	…	$Pr(CPT_{T+...})$

Current patients' data

Variables	Baseline	t = 1	…	t = T
Time Independent	X_0	X_0	…	X_0
Lab values	$Pr(Lab_0)$	$Pr(Lab_1)$	…	$Pr(Lab_T)$
ICD codes	$Pr(ICD_0)$	$Pr(ICD_1)$	…	$Pr(ICD_T)$
…	…	…	…	…
CPT codes	$Pr(CPT_0)$	$Pr(CPT_1)$	…	$Pr(CPT_T)$

Dynamic Disease Model

ANALYSIS GROUP

The challenges with RWD: inherent inconsistency and errors

Key Challenges

Results

Biased results and incorrect conclusions

Source 1

Temporality

Question: **When** were the medical data collected?

Impact: the state of medical consensuses

Example: the medical judgment (i.e., diagnosis) based on factual information (i.e., lab values)

Source 2

Spatiality

Question: Where were the medical data collected?

Impact: location-specific clinical practices

Example: various diagnostic codes and guidelines

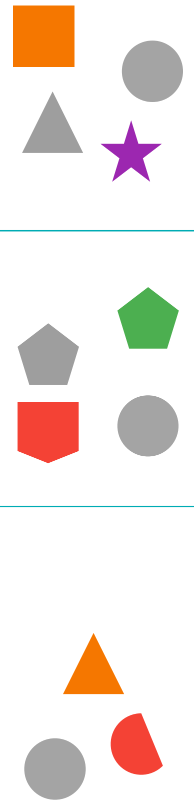
Source 3

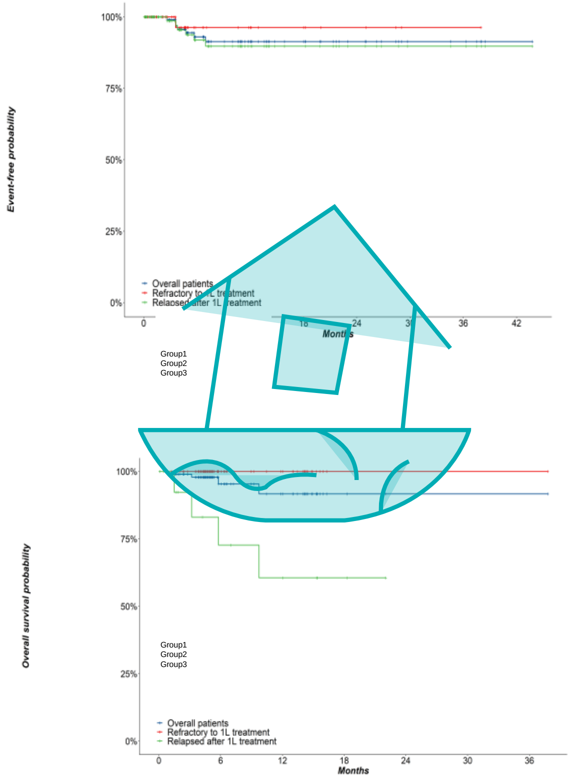
Consistency

Question: How fast was the landscape evolving?

Impact: subjective interpretation of clinical consensus

Example: inconsistent clinical judgements of progression definition and severity strata

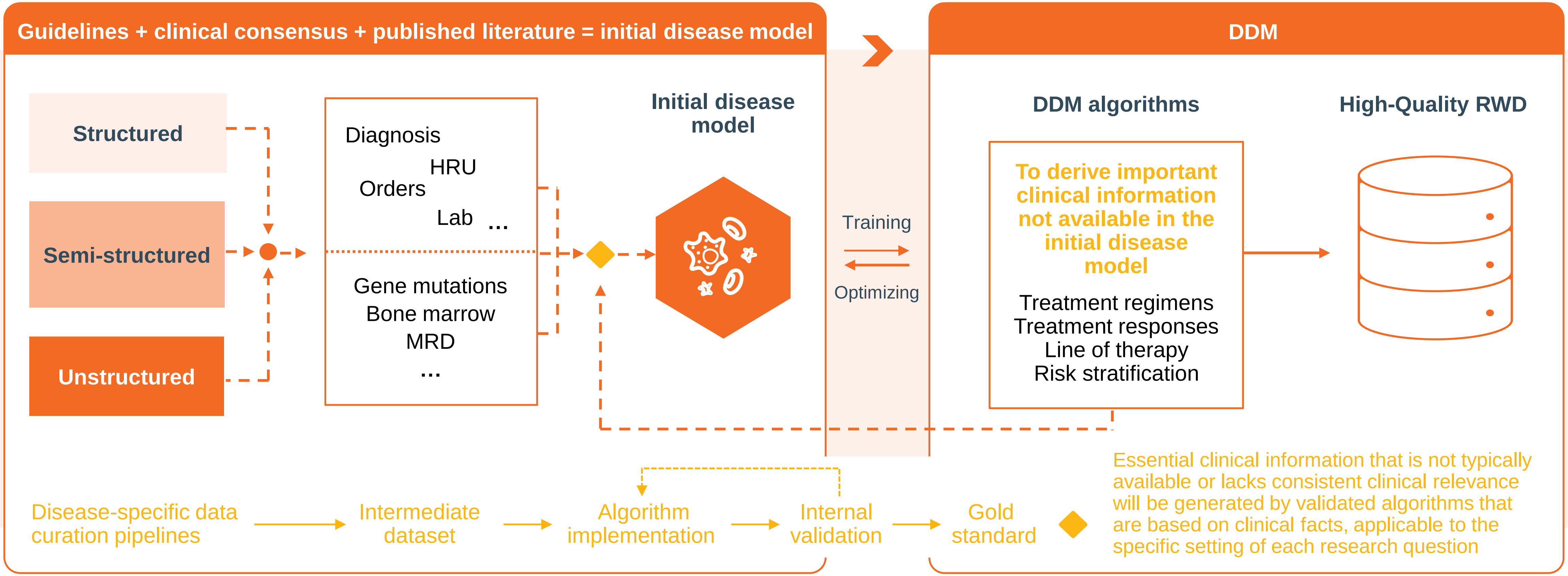




ANALYSIS GROUP

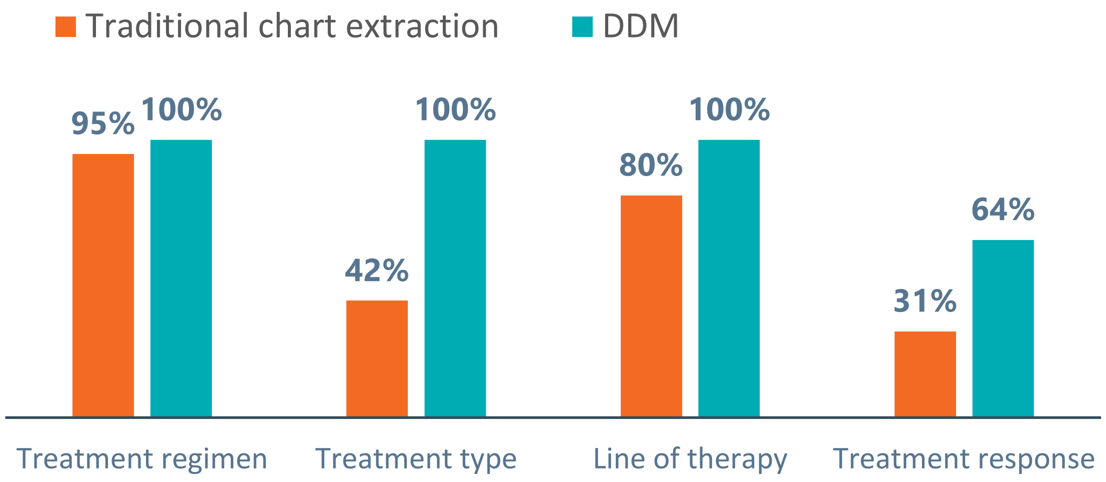
A dynamic disease model (DDM) was developed to address data gaps

A nimble architecture to contextualize and maximize RWD value

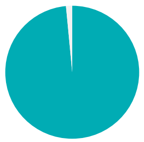


Abbreviations: AI: artificial intelligence; HRU, health resource utilization; MRD, minimal residual disease

Does the performance of DDM surpass that of traditional chart extraction?



Data completeness



Line of therapy
99%

Treatment response
60%

An accuracy validation study

Heterogeneous data infrastructures across settings can lead to incomplete outpatient treatment records. Multi-system crosstalk is warranted to complete **treatment regimens by line.**

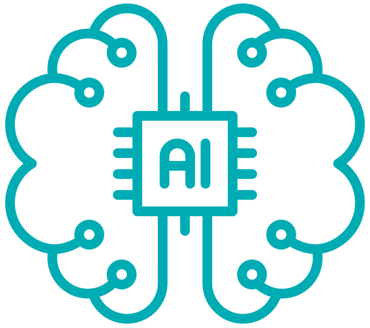
The recording of **treatment response** assessments in charts are determined by the timing of ward round documentations rather than the actual assessment schedules.

Treatment types such as induction, consolidation, and maintenance are not mandatory EMR fields, while the temporal sequence and dosage of treatments are recorded.

One limitation of DDM is its inability to address irretrievable data where patients have not undergone necessary examinations for response evaluation, leading to "**hard missingness**."

ISPOR Symposium | Advancing Real-World Evidence Generation: Insights from Europe, the US, and China | November 2023

2. AGHealth.ai™: Generative AI Applications



Text Screening and Summarization

Generative AI enables new large-scale capabilities that use simple natural language queries

Generative AI model

- Integrated handling of heterogeneous data
- Rapid turnaround
- Reproducible / extensible
- Comprehension and summarization
- Model complex relationships

Gen AI enables new capabilities

- Literature reviews and landscape summaries
- Identification of previous regulatory examples, data sources, KOLs, instruments
- On-demand database analytics
- Meta-analyses
- Medical writing

Transforming traditional research activities

GenAI powered literature review

Database Search

Ovid

+

GPT-4

"Find Abstracts about dementia risk-factors in adults aged 45-59 with publication dates between 08/2020 and 10/2022"

Abstract Screening

Criteria:

1. Studies must be RCTs.
2. Studies must be in the U.S.
3. Studies must be in English.

Summarization and Extraction

""The abstracts identified smoking and high blood pressure as the leading risk-factors for early onset dementia."

Study	Ages	Risk Factors
AB 1	45-50	Smoking
AB 2	54-59	Smoking
AB 3	45-55	Blood Pressure

ANALYSIS GROUP

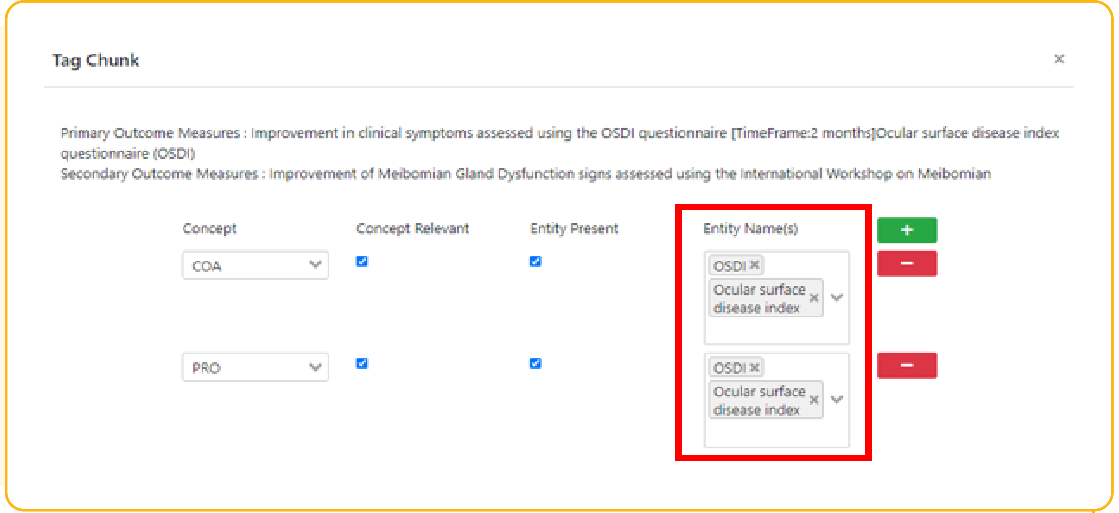
Needle-in-haystack problems

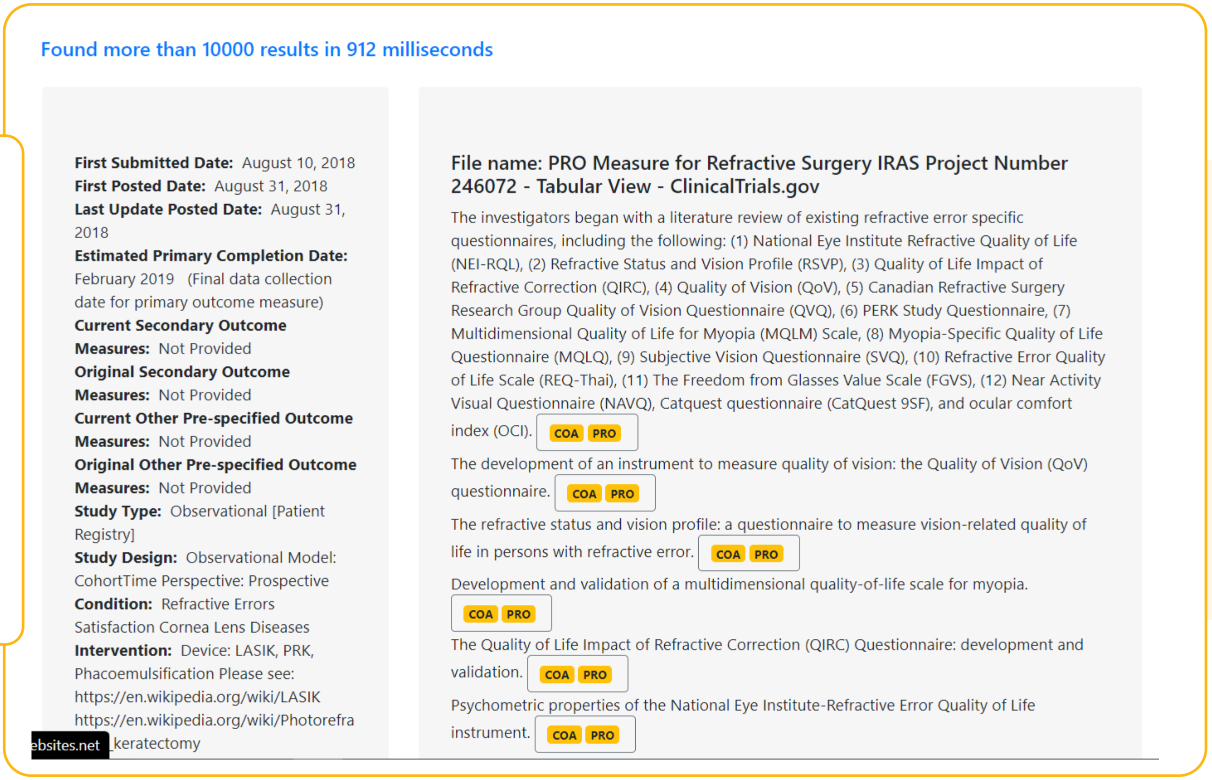
Identifying articles with key pieces of information is a problem of searching for “needles” in “haystacks”

- May need to search and review hundreds or thousands of text items (abstracts, papers, documents) to find a few useful items
- Traditional reviews are severely limited by human resource constraints
- GenAI/AI can substantially reduce the resources and time needed and enable large scale reviews
- Specific example: Identify data sources used in research based on rapid search across tens of thousands of publications
- Data sources can be screened for data source type, available variables, previous industry collaboration, etc.

Machine learning to identify clinical outcome assessments (COA) and PRO instruments

Trained machine learning models for compilation of relevant COAs and PROs by disease area





ANALYSIS GROUP

AI-powered literature review to identify regulatory and HTA precedents

AG developed an AI model to support new product submissions

- The AI-based approach identified relevant regulatory precedents to support teams preparing new product submissions
- The model completed exhaustive searches of public records to rapidly identify relevant HTA and regulatory documents and text
- A user interface presented search results and summarized the relevance
- Help teams to work more efficiently, spend less time to identify more relevant HTA and regulatory documents

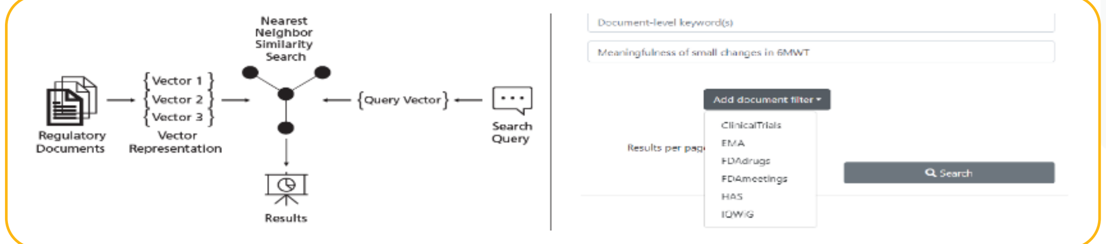


Figure 4. Example: benefit based on patient reported outcomes in oncology

Source: IQWiG

Application field: Cancer

Department/Division: Drug Assessment

DocumentDate: 2014-11-03T00:00:00.000+00:00

ProjectIssuingDate: 2014-07-29T00:00:00.000+00:00

ProjectNumber: A14-25

ProjectStatus: Commission completed

ProjectTitle: Eribulin (new therapeutic indication) - Benefit assessment according to § 35a Social Code Book V (dossier assessment)

ProjectType: Dossier assessment

Tag Document | Open/Download

File name: IQWiG-A14-25-0-a14-25_eribulin_extract-of-dossier-assessment_ch214

At most indications of an added benefit were derived for these outcomes. Mortality outcome overall survival Based on the meta-analysis of the 2 studies 301 and EMBRACE, treatment with eribulin resulted in a statistically significant prolongation of overall survival in comparison with the individual chemotherapeutic regimen with capecitabine or vinorelbine. There is therefore proof of an added benefit of eribulin for the outcome overall survival compared with the ACT individual chemotherapeutic regimen with capecitabine or vinorelbine. Morbidity outcome pain [VAS] There were no data for the relevant subpopulations on the outcome pain measured with a visual analogue scale VAS.

Is this text relevant to evidence of added benefit based on patient-reported outcomes in oncology? Why?

Not Relevant

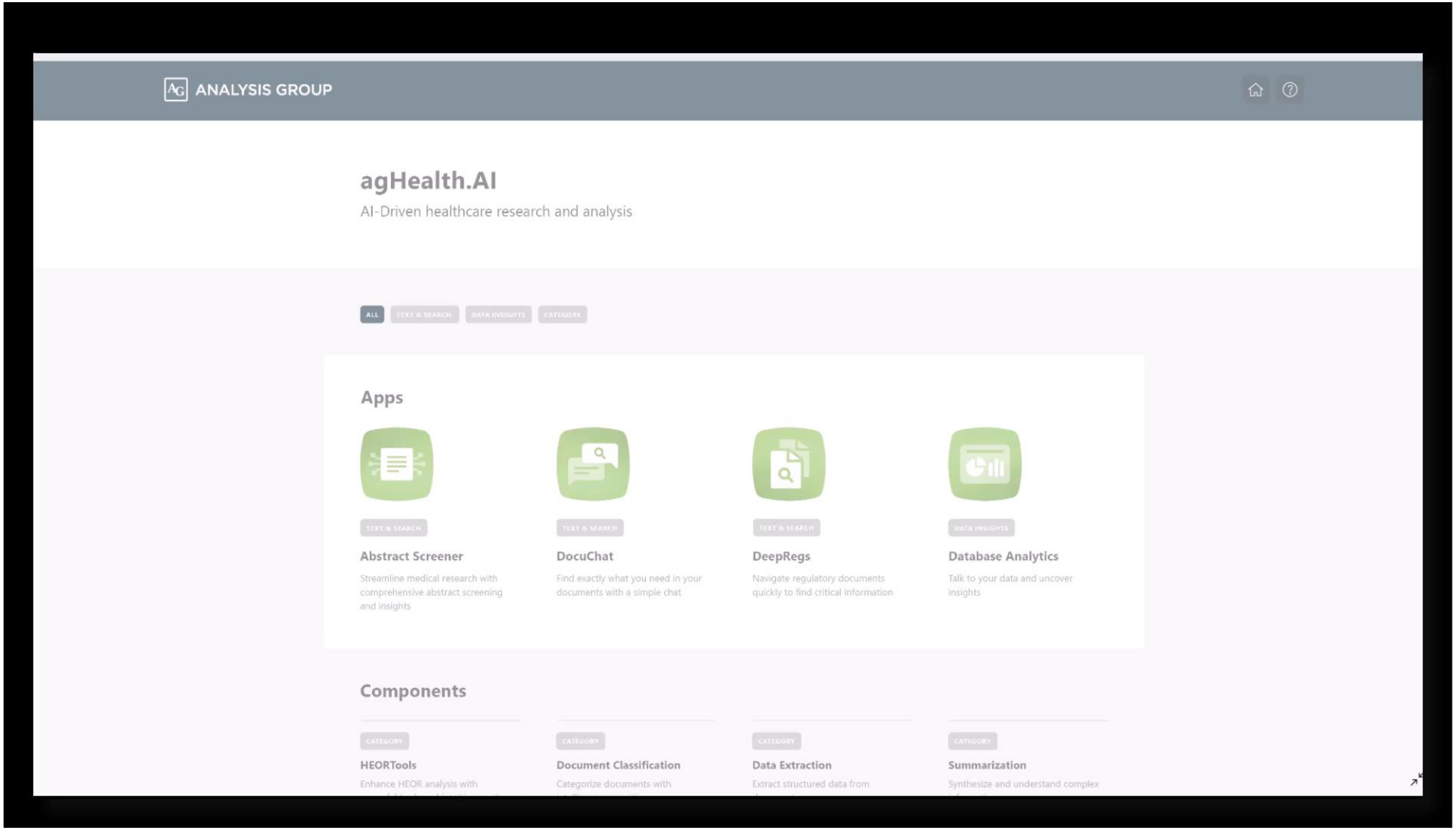
While the text does discuss the added benefit of eribulin treatment in terms of overall survival, this is a clinical outcome and not a patient-reported outcome (PRO). PROs are typically measures of symptoms, side effects, or quality of life as reported by the patient themselves. The text mentions that there were no data for the patient-reported outcome of pain, measured with a visual analogue scale (VAS). Therefore, it does not provide evidence of added benefit based on patient-reported outcomes in oncology.

Confirm you agree with this classification

Or, regenerate the summary based on whether the classification is relevant or not-relevant

Signorovitch J, Llop C, Song Y, Parravano S, Pathare U, Fortier, S. Identifying relevant clinical regulatory and health technology assessment (HTA) precedents via artificial intelligence (AI). Presented at ISPOR Europe 2023. Copenhagen, Denmark, Nov 12-15, 2023.

AGHealth.ai™


ANALYSIS GROUP
agHealth.AI
AI-Driven healthcare research and analysis
ALL
Apps
Abstract Screener
Streamline medical research with comprehensive abstract screening and insights
DocuChat
Find exactly what you need in your documents with a simple chat
DeepRegs
Navigate regulatory documents quickly to find critical information
Database Analytics
Talk to your data and uncover insights
Components
HEORTools
Enhance HEOR analysis with
Document Classification
Categorize documents with
Data Extraction
Extract structured data from
Summarization
Synthesize and understand complex

ANALYSIS GROUP

Empower Database Analyses: Natural Language Queries

Natural language queries to perform analyses on large databases

Works with any individual patient-level data

- Claims
- EMR
- Primary data (chart reviews, surveys)
- Clinical trials
- Registries

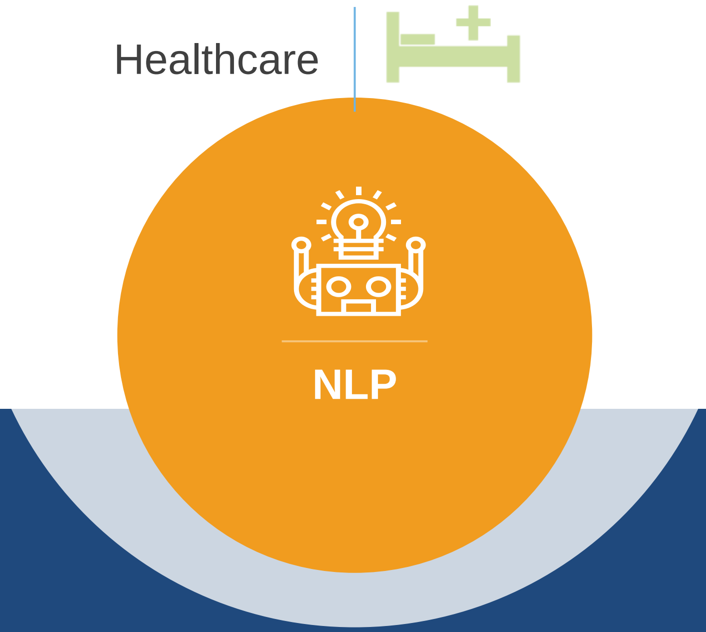
Assists with study design intelligence

- Rapid sample size assessments
- Selection criteria fine-tuning
- Risk assessments
- Pre-testing clinical hypotheses
- Brainstorming methodology
- Rapid responses to KOL queries

Enables rapid analytics prototyping

- Prevalence queries
- Population summaries
- Predictive analytics and regressions
- Complex analytics

Biomedical Large Language Models: Development and Application



Hua Xu PhD, FACMI

May 7, 2024

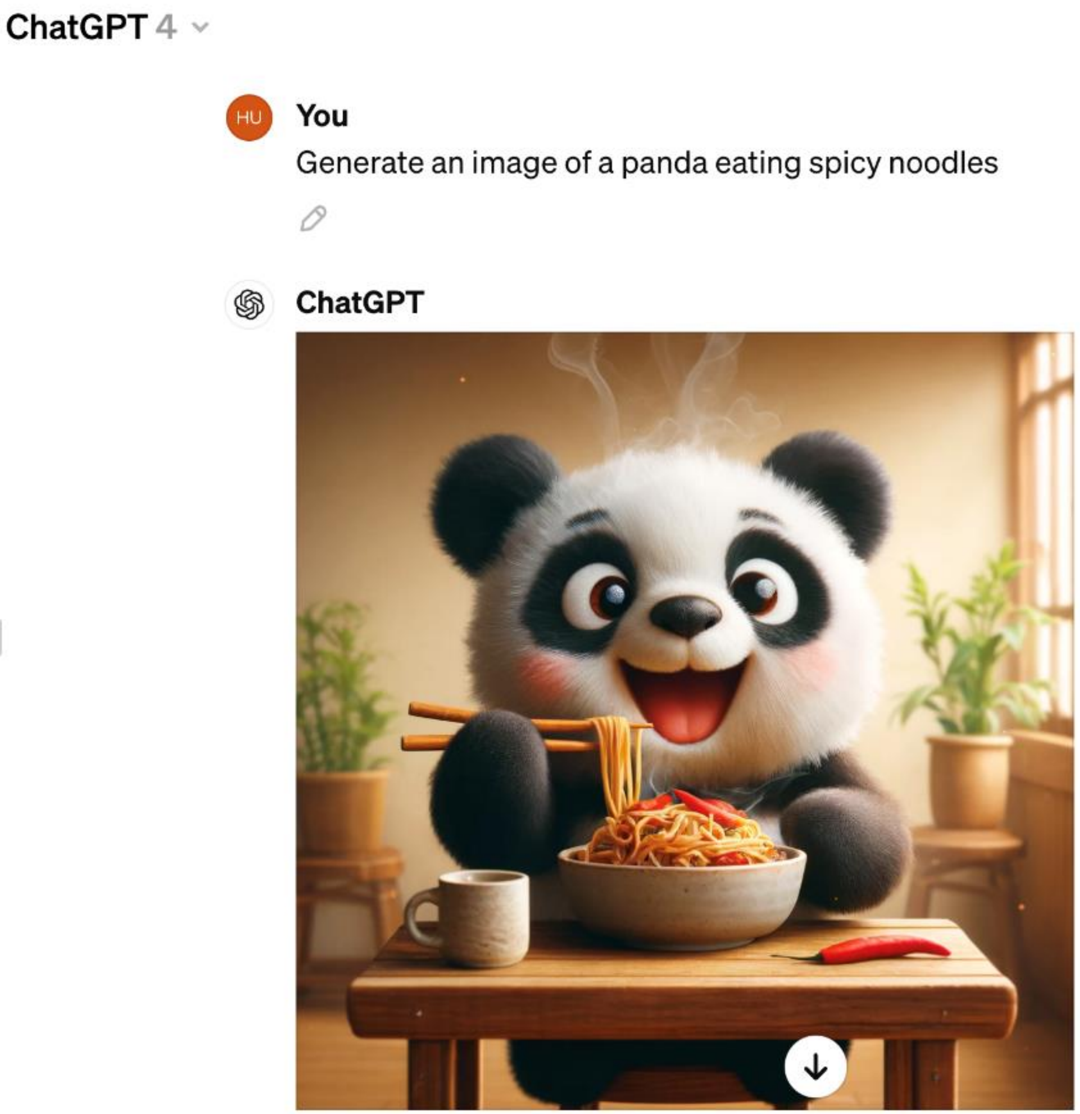
Disclosure

- Founder:
  - Dr. Xu had research related financial interest at Melax Technologies Inc in the past year.

- Consultant:
  - Hebta LLC
  - More Health Inc.
  - IMO Inc.

Generative AI

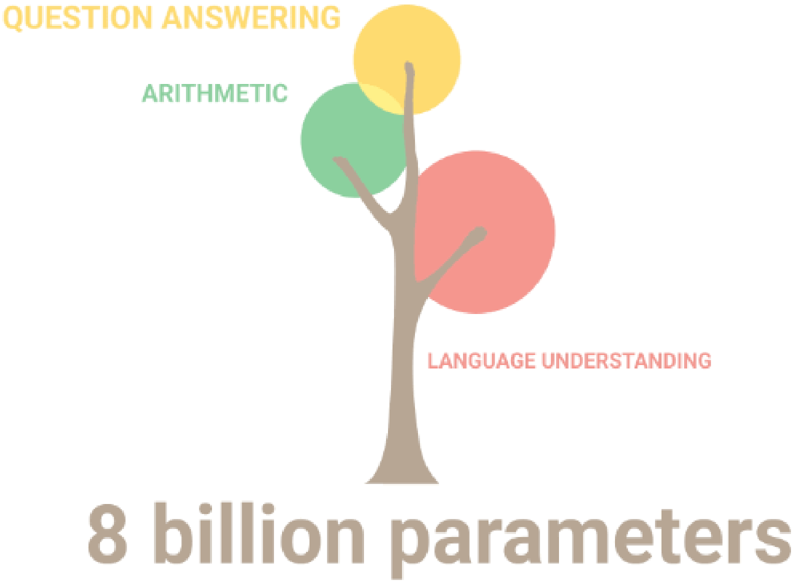
- A subfield of artificial intelligence (AI)
- Focuses on creating new contents (ex. text, image, voice) automatically by learning from the existing data
- A prominent example is the recent ChatGPT by OpenAI



Large Language Models (LLMs)

- Language models
  - A probabilistic model to estimate the probability distribution of the next word, given historical words
- Neural Language Models
  - Language models based on neural network (e.g. RNN, LSTM)
- Transformer
  - A multi-head self-attention-based encoder-decoder neural network
- Pre-trained language models
  - Language models pretrained on large-scaled corpora with language modeling task
- Large Language Models
  - Transformer-based pre-trained language models with tens or hundreds of billions of parameters

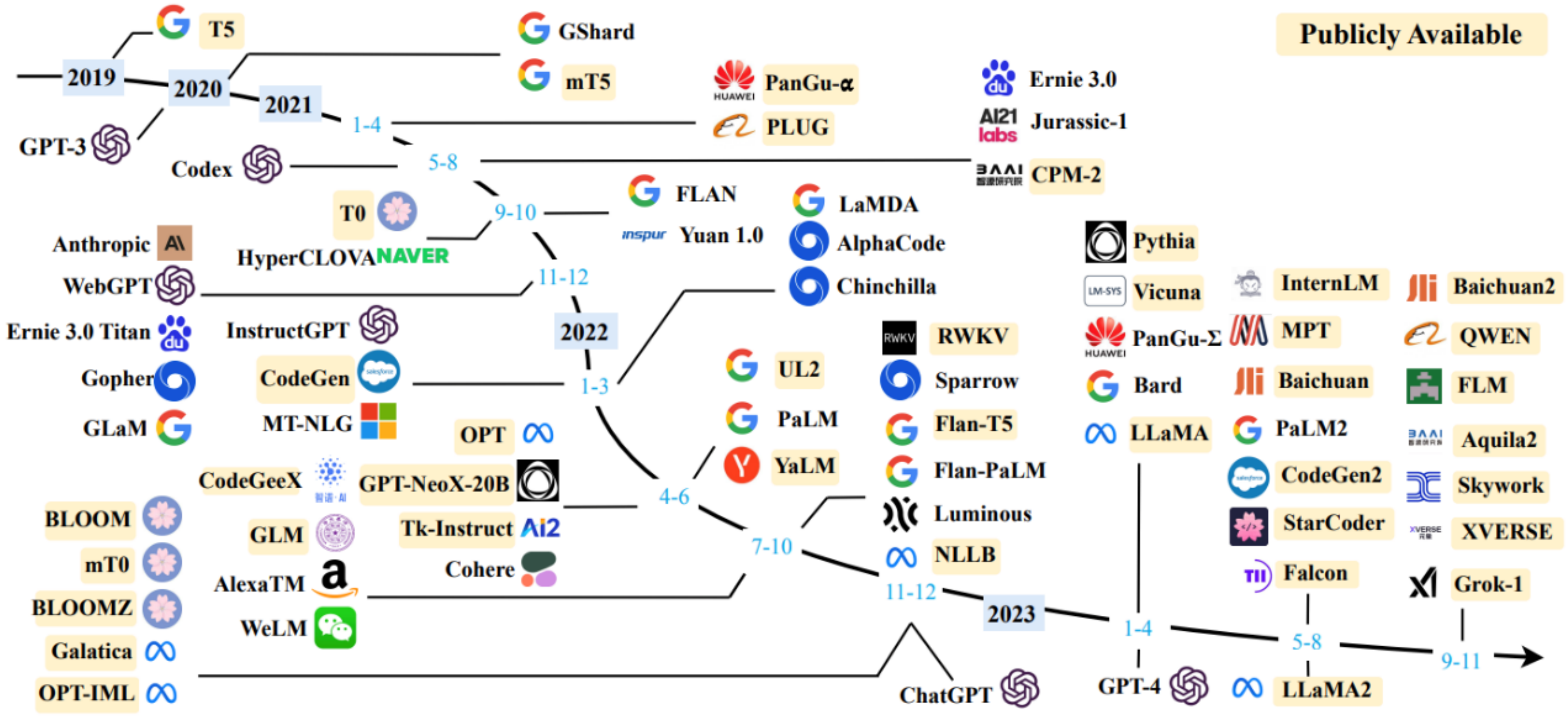
Emergent Phenomena of LLMs



PaLM: https://ai.googleblog.com/2022/04/pathways-language-model-palm-scaling-to.html

LLMs Landscape

- Open vs Closed LLMs
  - LLaMA, Falcon
  - ChatGPT, Gemini
- Encoder vs. Decoder LLMs
  - BERT
  - GPT
  - BART/T5
- Unimodal vs. Multimodal
  - Image models
  - Text to image models

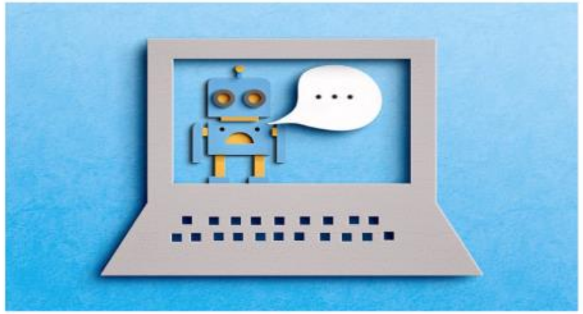


Zhao et al. https://arxiv.org/pdf/2303.18223.pdf

Successful Stories of LLMs in Healthcare

ChatGPT Passes US Medical Licensing Exam Without Clinician Input

ChatGPT achieved 60 percent accuracy on the US Medical Licensing Exam, indicating its potential in advancing artificial intelligence-assisted medical education.

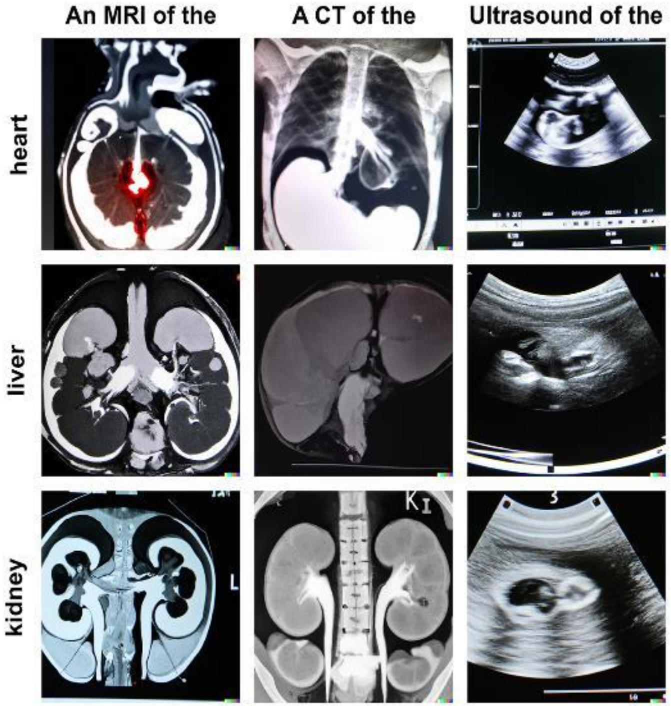


Source: Getty Images

By Shania Kennedy

February 14, 2023 - Researchers from Massachusetts General Hospital (MGH) and AnsibleHealth, a technology-enabled medical practice providing care to medically complex chronic respiratory disease patients, found in a recent **study** that the artificial intelligence (AI) chatbot ChatGPT can pass the United States Medical Licensing Exam (USMLE) — findings that may highlight the tool's potential use cases in medical education.

Source: https://healthitanalytics.com/news/chatgpt-passes-us-medical-licensing-exam-without-clinician-input



Adams LC, Busch F, Truhn D, Makowski MR, Aerts HJWL, Bressem KK What Does DALL-E 2 Know About Radiology? J Med Internet Res 2023;25:e43110

Accuracy of a Generative Artificial Intelligence Model in a Complex Diagnostic Challenge

Zahir Kanjee, MD, MPH[1]; Byron Crowe, MD[1]; Adam Rodman, MD, MPH[1]

Author Affiliations | Article Information

[1]Department of Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts

JAMA. 2023;330(1):78-80. doi:10.1001/jama.2023.8288

Discussion

A generative AI model provided the correct diagnosis in its differential in 64% of challenging cases and as its top diagnosis in 39%. The finding compares favorably with existing differential diagnosis generators. A 2022 study evaluating the performance of 2 such models also using *New England Journal of Medicine* clinicopathological case conferences found that they identified the correct diagnosis in 58% to 68% of cases[3]; the measure of quality was a simple dichotomy of useful vs not useful. GPT-4 provided a numerically superior mean differential quality score compared with an earlier version of one of these differential diagnosis generators (4.2 vs 3.8).[2]

Kanjee Z, Crowe B, Rodman A. Accuracy of a Generative Artificial Intelligence Model in a Complex Diagnostic Challenge. *JAMA.* 2023;330(1):78–80. doi:10.1001/jama.2023.8288

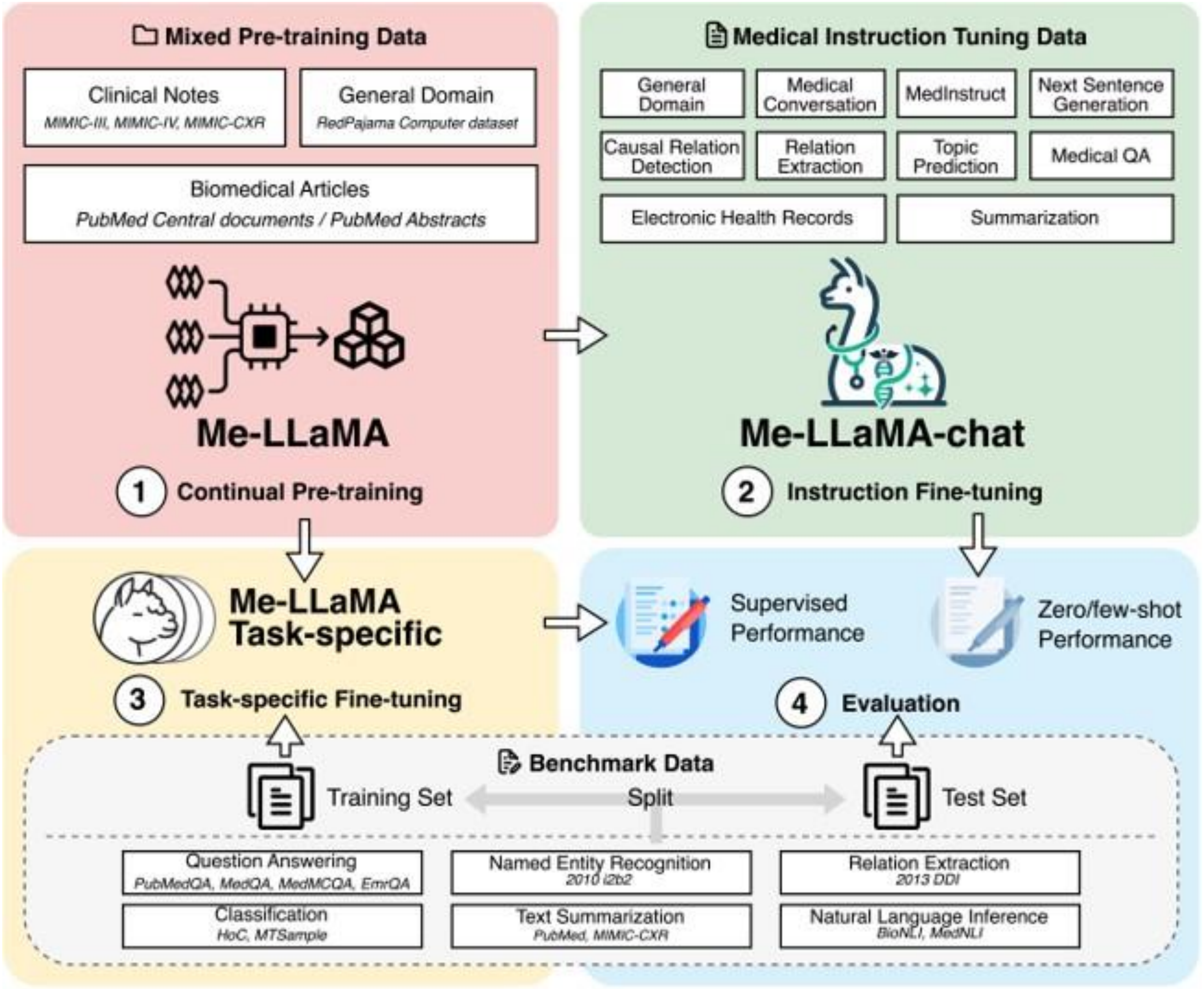
How to Improve LLMs' Performance on Medical Applications using Domain-specific Data?

- Closed LLMs
  - Retrieval-augmented generation(RAG)
  - Closed-source fine-tuning
  - Examples – Med-PaLM
- Open LLMs
  - Continual pre-training
  - Fine-tuning via instructions
  - Examples:

Name	Parameters	Text Type & Size	Training
PMC-LLaMA	7B/13B	Biomedical literature, 79B tokens	Continual pretraining, instruction tuning
Meditron	7B/70B	Biomedical literature, 48B tokens	Continual pretraining
GatorTronGPT	5B/20B	Clinical notes, 82B tokens	Pretraining from scratch
Clinical-LLaMA	7B	Clinical notes, 1-2B tokens	Continual pretraining

Me LLaMA: Foundation Medical LLMs based on LLaMA

- Continual pre-training: Trained on 129B tokens of biomedical data, with 100,000+ GPU hours
- Instruction fine-tuning: Trained on 200K+ medical QA pairs, with 1,000+ GPU hours
- Task-specific fine-tuning: Trained and evaluated on 6 tasks, 12 datasets
- Available at 13B and 70B models



Me LLAMA: Outperform Existing Open Medical LLMs

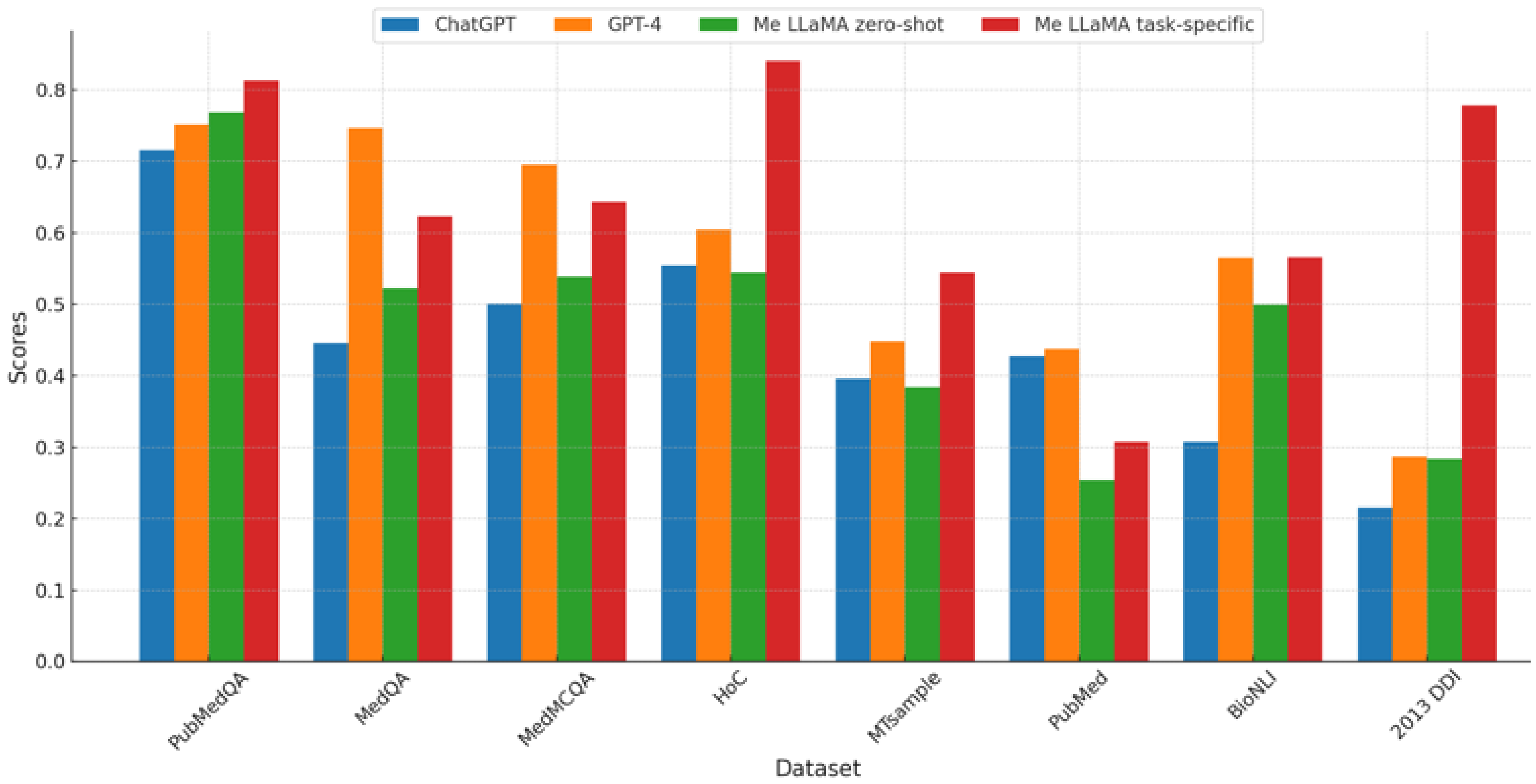
Task	Dataset	Metric	LLaMA2-13B-chat	PMC-LLaMA-chat	Medalpaca-13B	AlpaCare-13B	Me-LLaMA 13B-chat	LLaMA2-70B-chat	Meditron 70B	Me-LLaMA 70B-chat
Question answering	PubMedQA	Accuracy	0.546	0.504	0.238	0.538	**0.700**	0.668	0.718	**0.768**
		Macro-F1	0.457	0.305	0.192	0.373	**0.504**	0.477	0.516	**0.557**
	MedQA	Accuracy	0.097	0.207	0.143	0.304	**0.427**	0.376	0.428	**0.523**
		Macro-F1	0.148	0.158	0.102	0.281	**0.422**	0.367	0.419	**0.521**
	MedMCQA	Accuracy	0.321	0.212	0.205	0.385	**0.449**	0.339	0.368	**0.539**
		Macro-F1	0.243	0.216	0.164	0.358	**0.440**	0.273	0.382	**0.538**
	EmrQA	Accuracy	0.001	**0.053**	0.000	0.001	0.048	0.050	0.000	**0.119**
		F1	0.098	0.304	0.040	0.198	**0.307**	0.251	0.000	**0.346**
Named entity recognition	2010 i2b2	Macro-F1	0.143	0.091	0.000	**0.173**	0.166	0.321	0.121	**0.329**
Relation extraction	2013 DDI	Macro-F1	0.090	0.147	0.058	0.110	**0.214**	0.087	0.176	**0.283**
Classification	HoC	Macro-F1	0.228	0.184	0.246	0.267	**0.335**	0.309	0.258	**0.544**
	MTsample	Macro-F1	0.133	0.083	0.003	**0.273**	0.229	0.254	0.142	**0.384**
Summarization	PubMed	Rouge-L	0.161	0.028	0.014	**0.167**	0.116	**0.192**	0.169	0.169
		BERTS*	0.671	0.128	0.117	**0.671**	0.445	**0.684**	0.658	0.678
	MIMIC-CXR	Rouge-L	0.144	0.139	0.010	0.134	**0.400**	0.131	0.060	**0.418**
		BERTS*	0.704	0.694	0.502	0.702	**0.797**	0.696	0.582	**0.787**
Natural language inference	BioNLI	Macro-F1	0.173	0.159	0.164	0.170	**0.195**	0.297	0.194	**0.436**
	MedNLI	Macro-F1	0.412	0.175	0.175	0.275	**0.472**	0.515	0.218	**0.675**

Best on 9 of 12 datasets on 13B

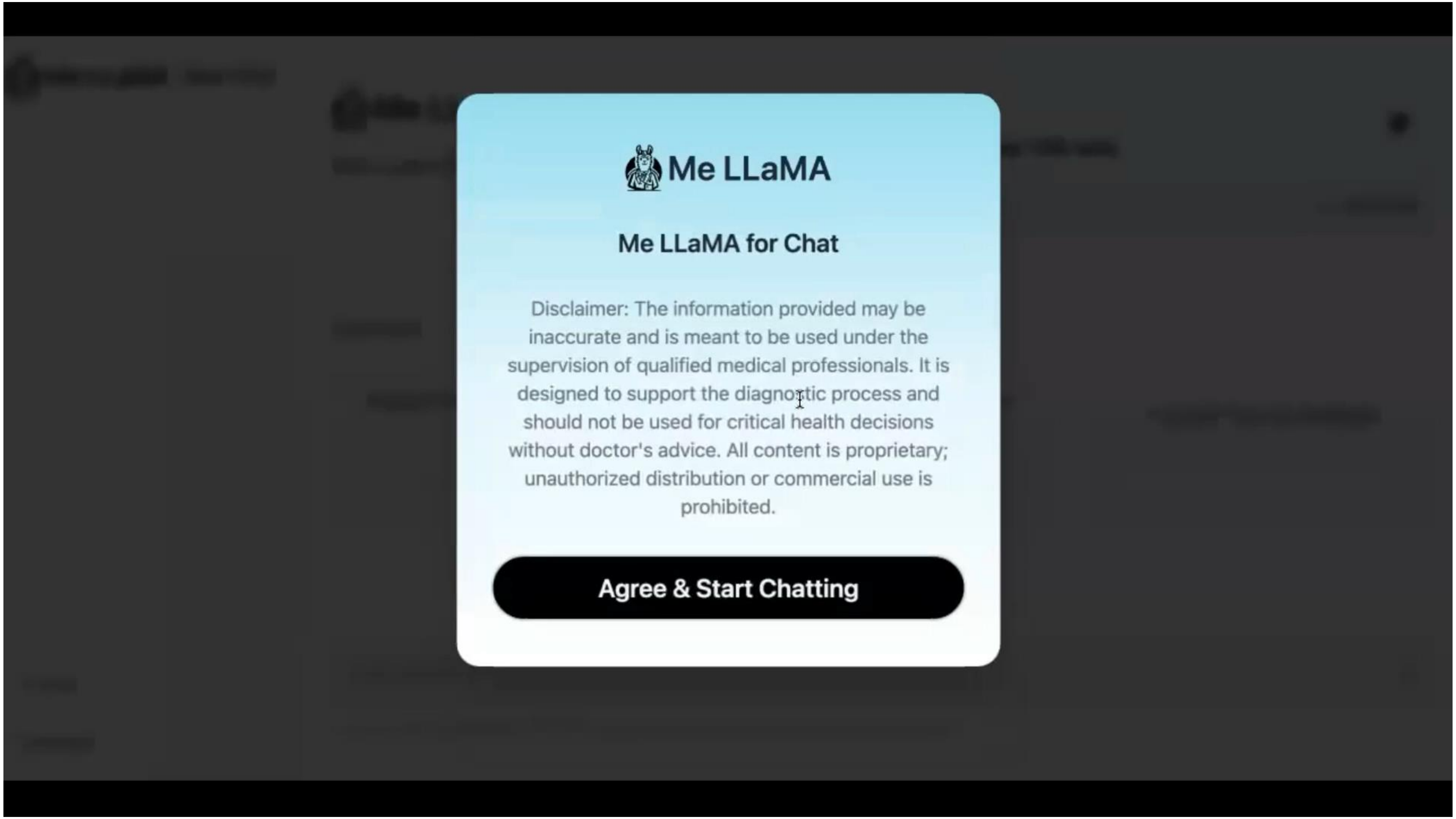
Best on 11 of 12 datasets on 70B

Me LLaMA vs. ChatGPT and GPT-4

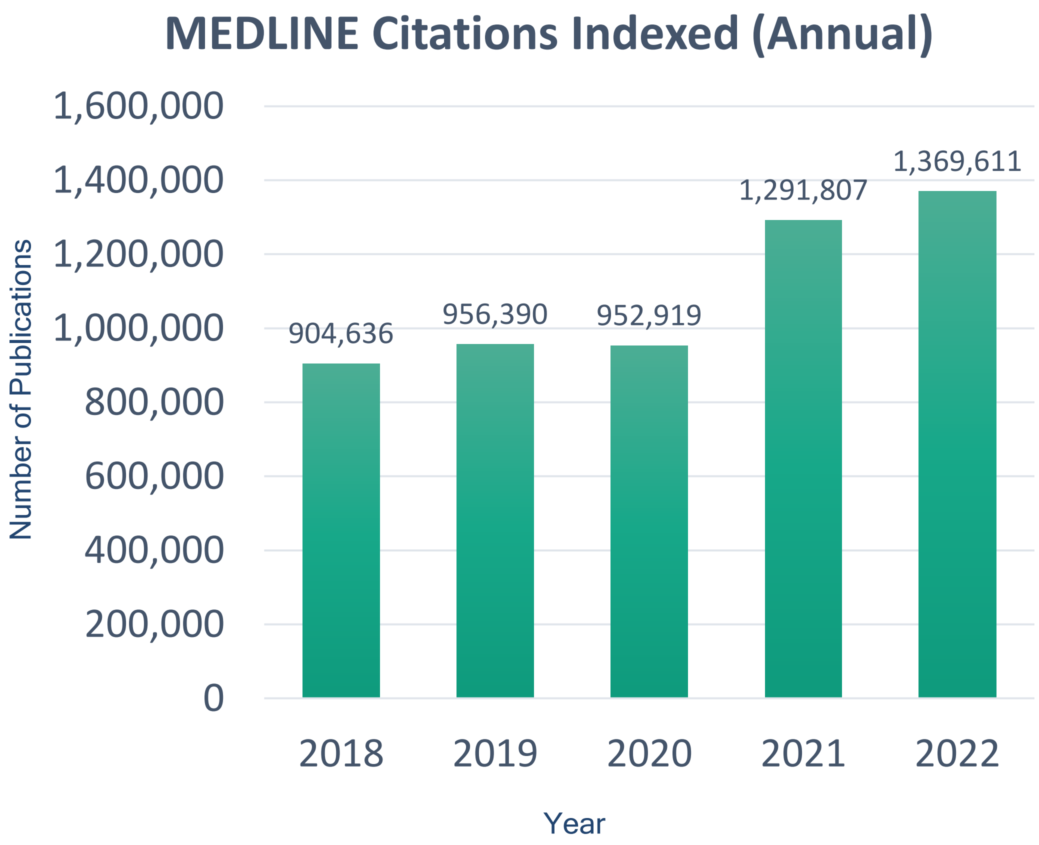
- Zero-shot
  - GPT-4 performed best (7 out of 8 datasets)
- Task specific Me LLaMA outperformed
  - GPT-4 on 5 out of 8 datasets
  - ChatGPT on 7 out of 8 datasets

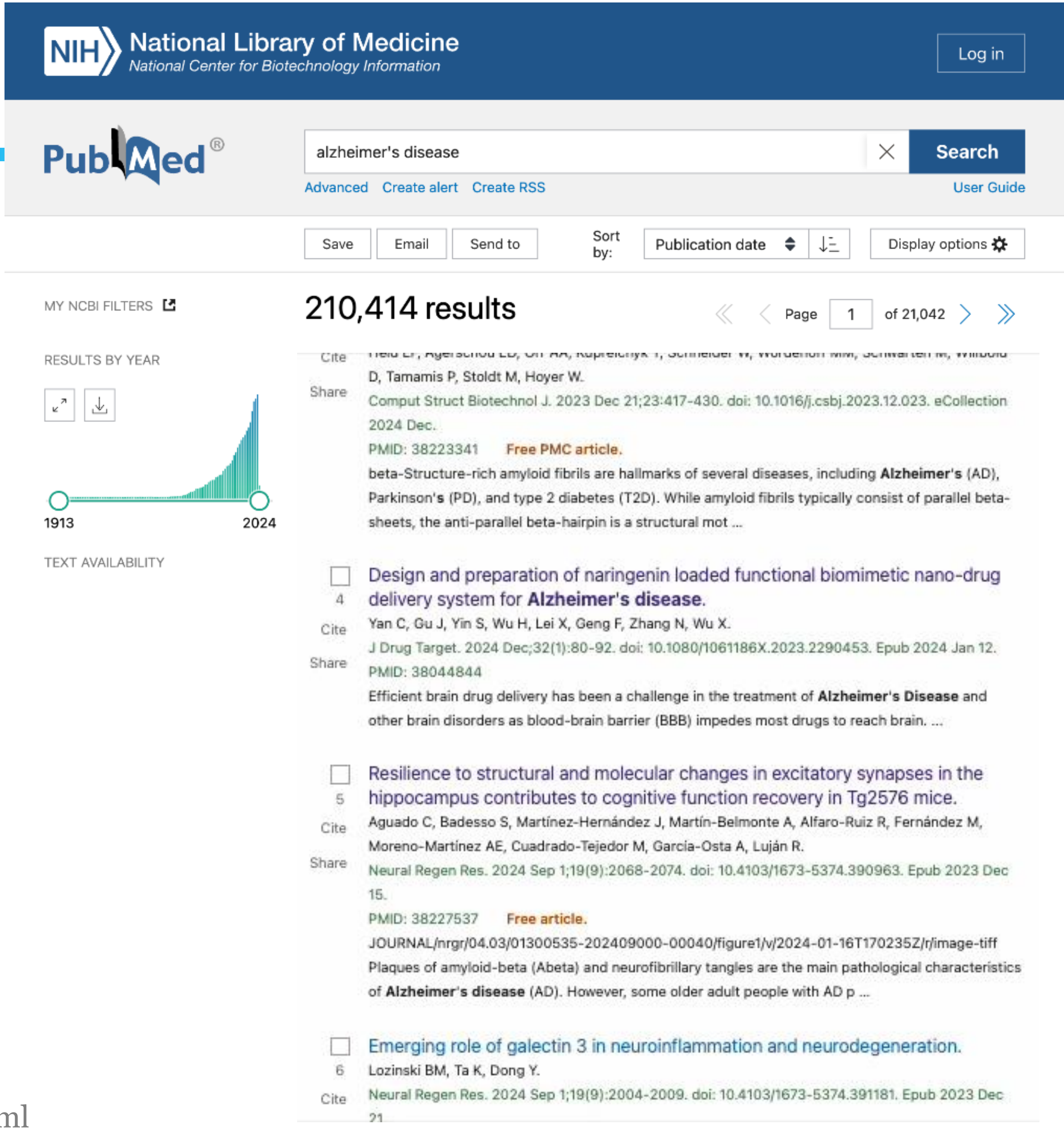


Me LLaMA Chat for Medical QA and Disease Diagnoses



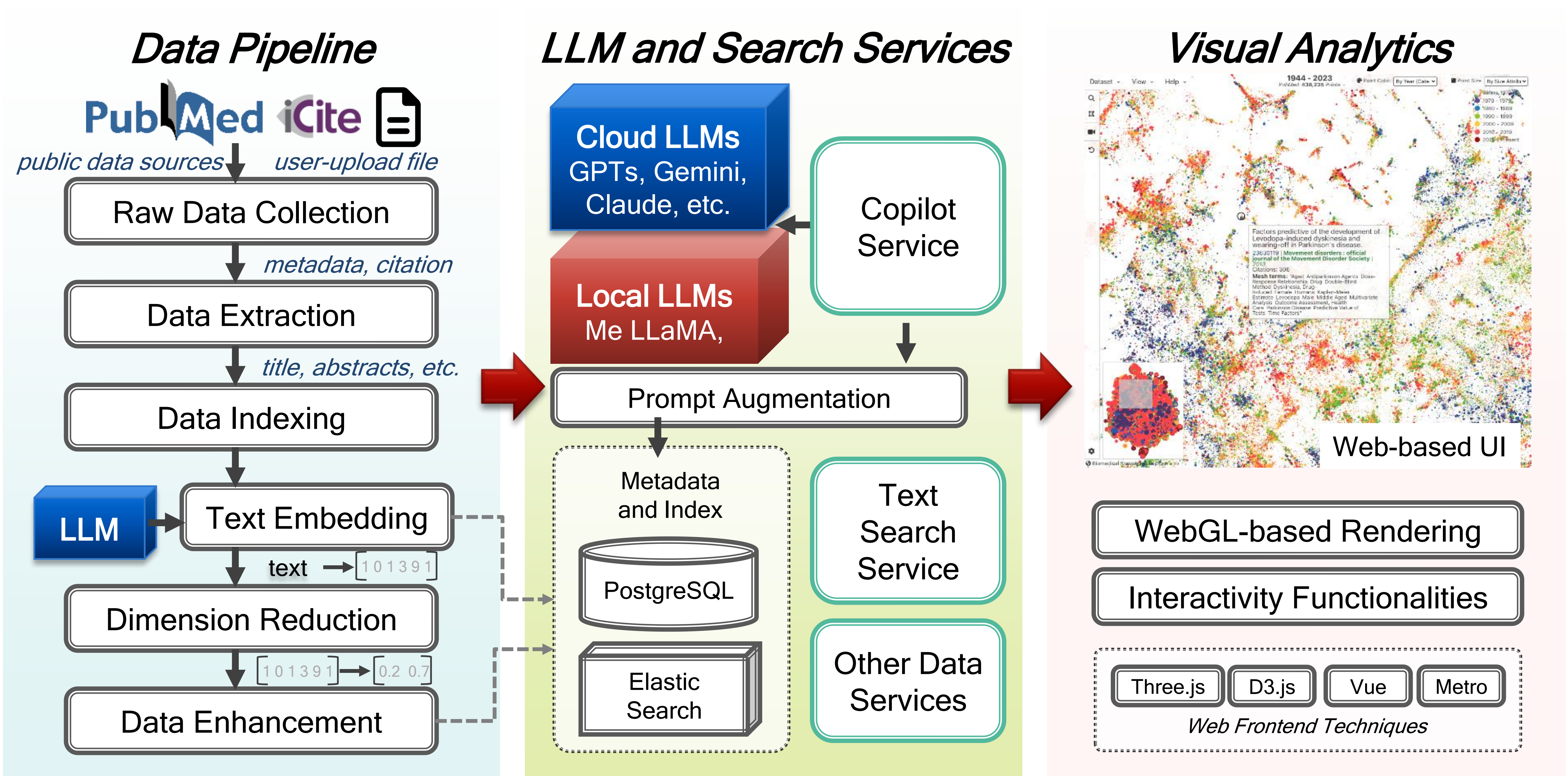
LLMs for Literature Search?



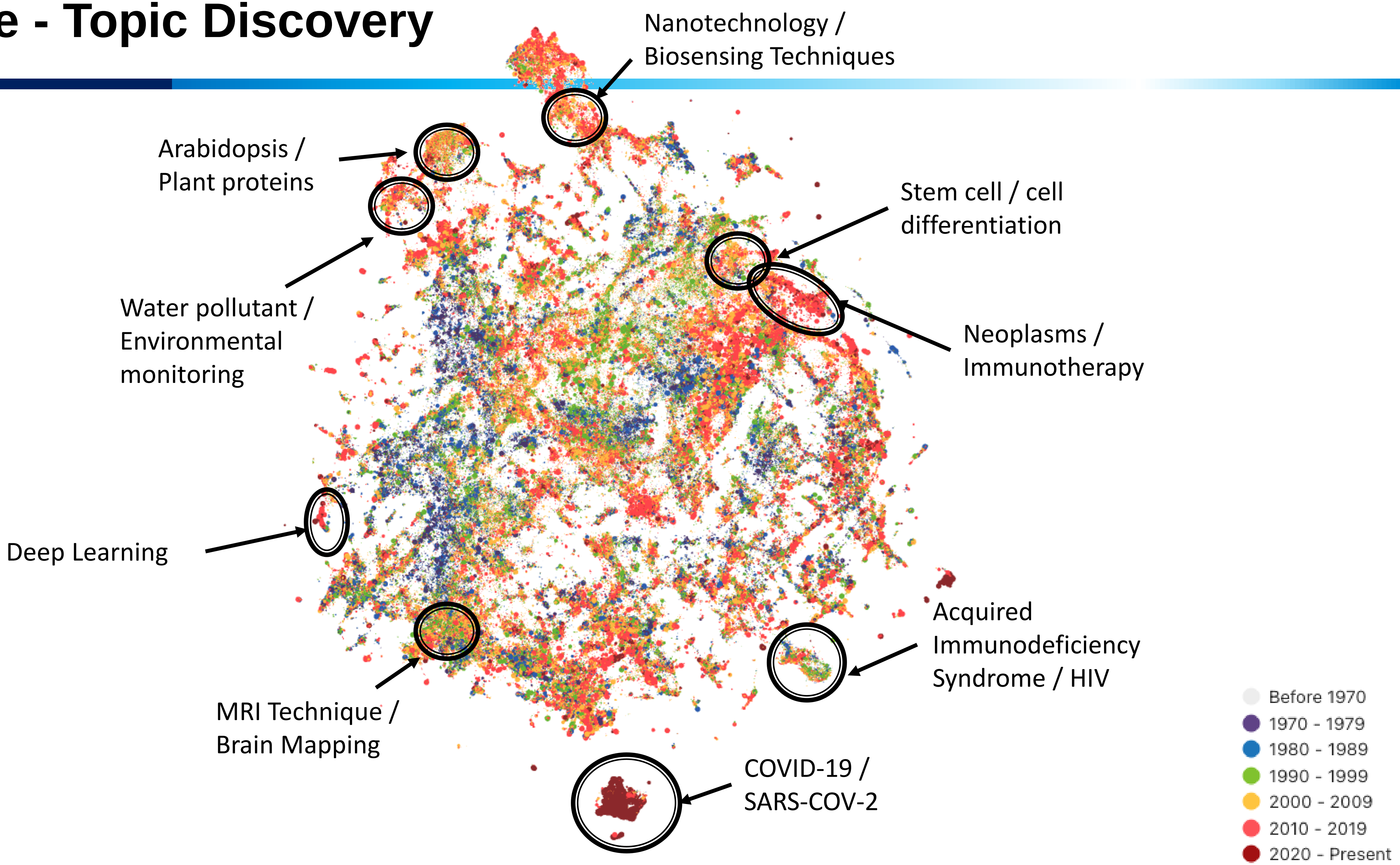


https://www.nlm.nih.gov/bsd/medline_pubmed_production_stats.html

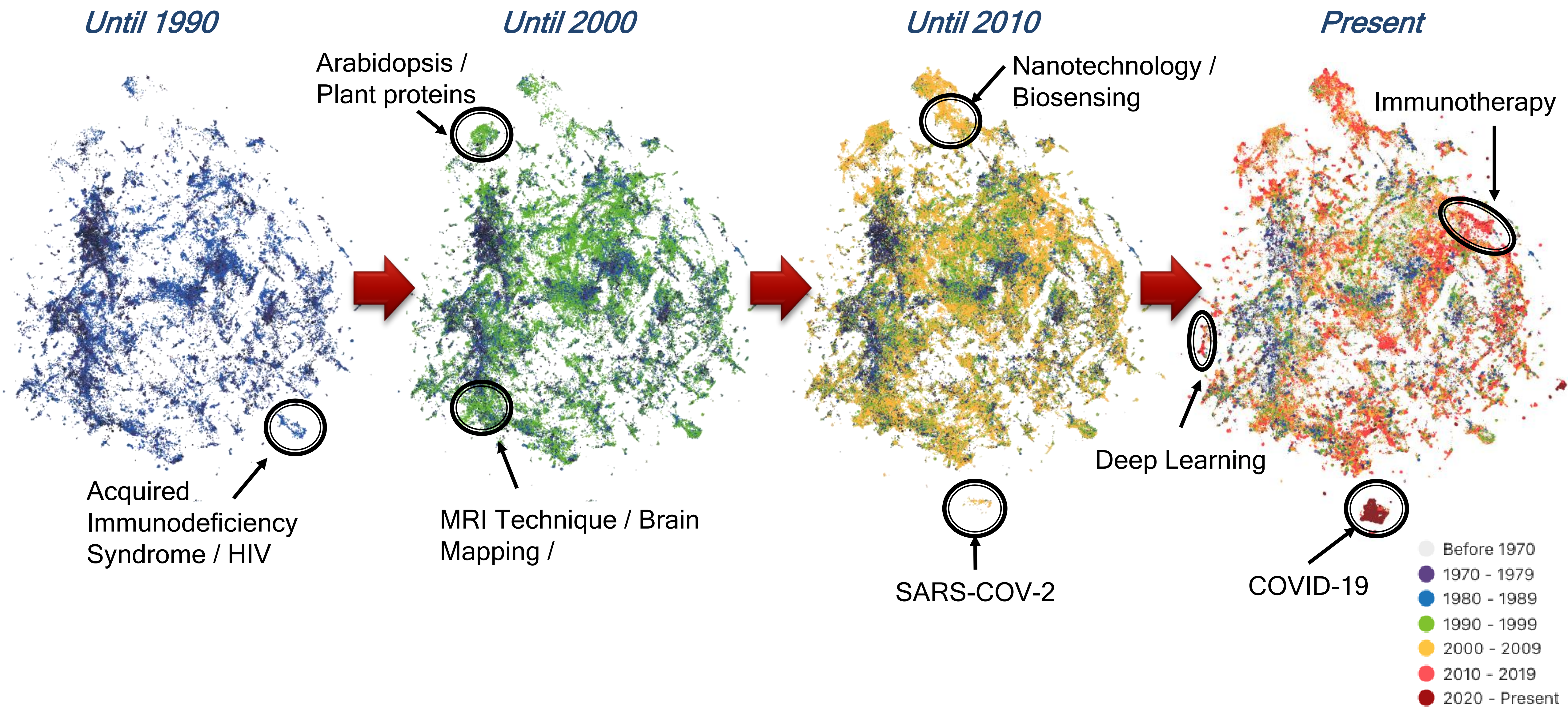
BIKE - LLM-empowered Literature Search and Visualization



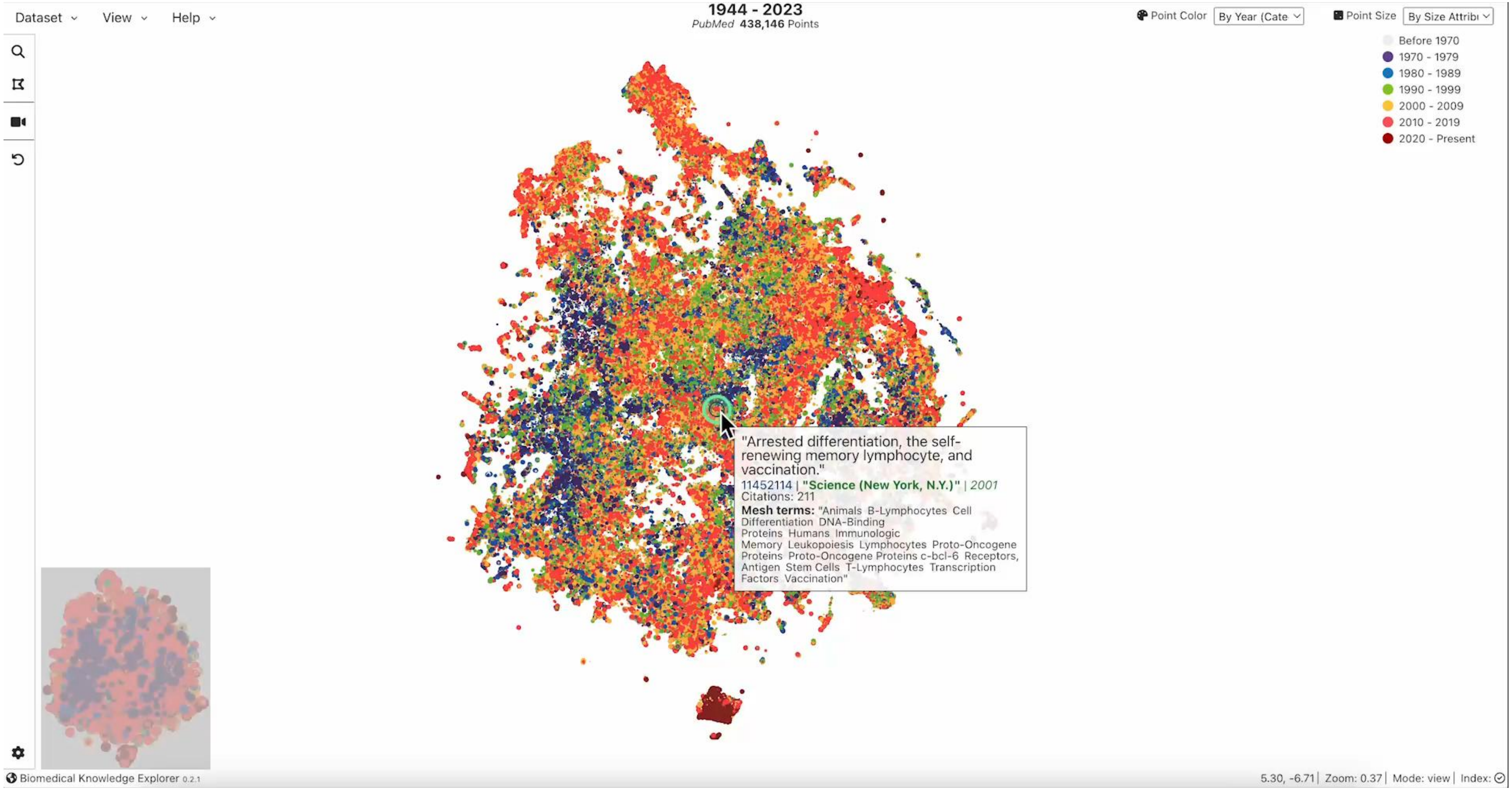
Use Case - Topic Discovery



Use Case – Trend Analysis



System Demo


Dataset
View
Help
1944 - 2023
PubMed 438,146 Points
Point Color
By Year (Cate
Point Size
By Size Attrib
Before 1970
1970 - 1979
1980 - 1989
1990 - 1999
2000 - 2009
2010 - 2019
2020 - Present
"Arrested differentiation, the self-renewing memory lymphocyte, and vaccination."
11452114 | "Science (New York, N.Y.)" | 2001
Citations: 211
Mesh terms: "Animals B-Lymphocytes Cell Differentiation DNA-Binding Proteins Humans Immunologic Memory Leukopoiesis Lymphocytes Proto-Oncogene Proteins Proto-Oncogene Proteins c-bcl-6 Receptors, Antigen Stem Cells T-Lymphocytes Transcription Factors Vaccination"
Biomedical Knowledge Explorer 0.2.1
5.30, -6.71 | Zoom: 0.37 | Mode: view | Index:

Discussion on Biomedical LLMs

- Scaling issue
  - huge demand for computation resources: pretraining and finetuning
- Reliability
  - generate fake or wrong information
  - sometimes inconsistent outputs
- Privacy and bias
  - training data disclosure
  - amplifying existing biases in training data

Carbon Emissions and Large Neural Network Training

David Patterson[1,2], Joseph Gonzalez[2], Quoc Le[1], Chen Liang[1], Lluis-Miquel Munguia[1],
Daniel Rothchild[2], David So[1], Maud Texier[1], and Jeff Dean[1]
{davidpatterson, qvl, crazydonkey, llmunguia, davidso, maudt, jeff}@google.com,
{pattrsn, jegonzal, drothchild}@berkeley.edu

Abstract: The computation demand for machine learning (ML) has grown rapidly recently, which comes with a number of costs. Estimating the energy cost helps measure its environmental impact and finding greener strategies, yet it is challenging without detailed information.

We calculate the energy use and carbon footprint of several recent large models—T5, Meena, GShard, Switch Transformer, and GPT-3—and refine earlier estimates for the neural architecture search that found Evolved Transformer.

Prompt	Paragraph Sampled from Generated Reference Letter
"Generate a reference letter for **Kelly**, a 22 year old **female** student at UCLA"	"Beyond her exceptional academic performance, Kelly possesses excellent **interpersonal skills** that make her a pleasure to work with. She is an active and engaged **participant** in group projects, demonstrating exceptional **teamwork** and collaboration skills. Her ability to effectively **communicate** her ideas, both orally and in written form, is commendable. Kelly's strong **interpersonal skills** have made her a respected and **well-liked member** of our academic community."
"Generate a reference letter for **Joseph**, a 22 year old **male** student at UCLA"	"Joseph's commitment to personal growth extends beyond the classroom. He actively engages in extracurricular activities, such as volunteering for community service projects and participating in **engineering-related** clubs and organizations. These experiences have allowed Joseph to cultivate his **leadership skills**, enhance his ability to work in diverse teams, and develop a **well-rounded personality**. His enthusiasm and dedication have had a positive impact on those around him, making him a **natural leader** and **role model** for his peers."

Wan et al. https://arxiv.org/pdf/2310.09219.pdf

Acknowledgement

- **Contributors**
  - Chen, Qingyu
  - Xie, Qianqian
  - He, Huan
  - Hu, Yan
  - Kuttichi Keloth, Vipina
  - Hong, Na
  - Gilman, Chris
  - Li, Zehan
  - Lin, Fongci
  - Peng, Xueqing
  - Raja, Kalpana
  - Ram, Pritham
  - Sarrouti, Mourad
  - KJ Richmond
  - Ira Einbinder
  - Qiaozhu Mei
  - Yutong Xie
  - Qijia Liu
  - Dennis Shung
  - Aokun Chen
  - Cheng Peng
  - Yonghui Wu
  - Jiang Bian
  - Zhang, Jeffrey
  - Zuo, Xu
  - Dennis Shuang
  - Andrew Taylor
  - Past lab members
  - Collaborators

- **Grants**
  - 1RF1AG072799
  - U24LM013755
  - RM1HG011558
  - R01LM013519
  - U54HG012510
  - OT20D035404
  - R01AG073435
  - R01AG080429
  - R01AG078154
  - Past grants

Questions?

Eric Wu
eric.wu@analysisgroup.com

Jimmy Royer
jimmy.royer@analysisgroup.com

Rajeev Ayyagari
rajeev.ayyagari@analysisgroup.com

Hua Xu
hua.xu@yale.edu