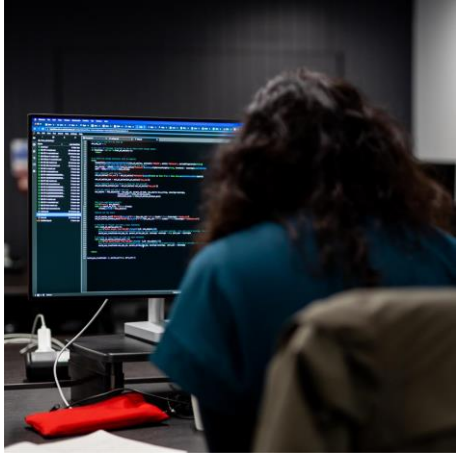
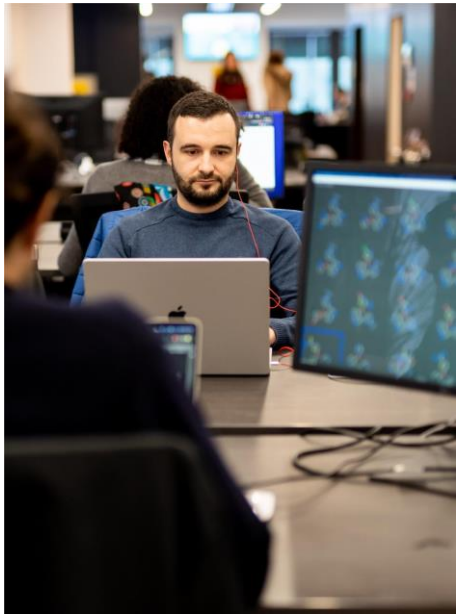
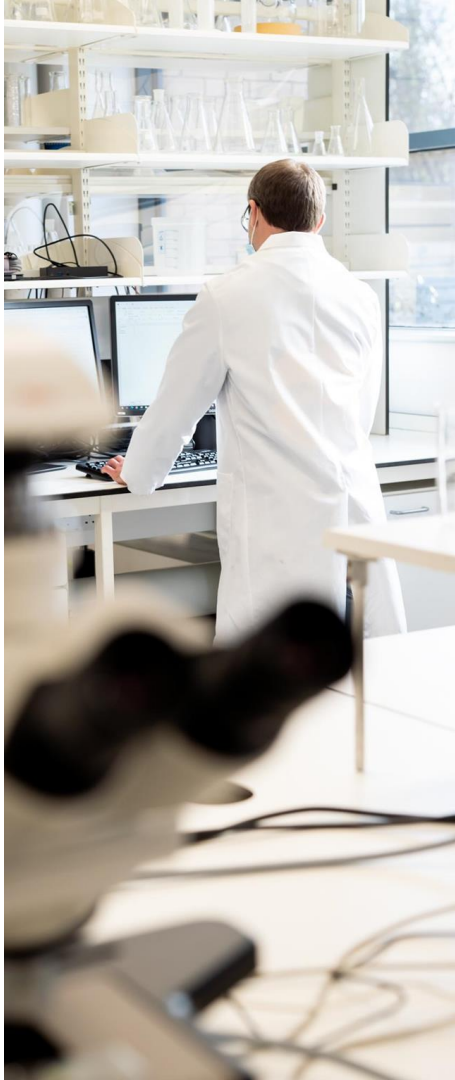
Benevolent[AI]

Building trust in an uncertain world

James Malone | 24th April 2024







What if AI does not solve all our problems?

Amazon AI's Just Walk Out

- Amazon began trialling a system to allow customers to simply pick up items and leave
- The system uses AI image recognition to detect what they have picked up
- They use human annotators to review 70% of the transactions that occur to build new data to continue to improve the models
- But they are also now closing a number of the Just Walk Out driven transactions and moving to a virtual shopping cart type model
- Why?
    - Because customers want to know what they're buying as they go
    - They want knowledge over the goods that Amazon *thinks* they have in their basket

The Verge / Tech / Reviews / Science / Entertainment

AMAZON / TECH / BUSINESS

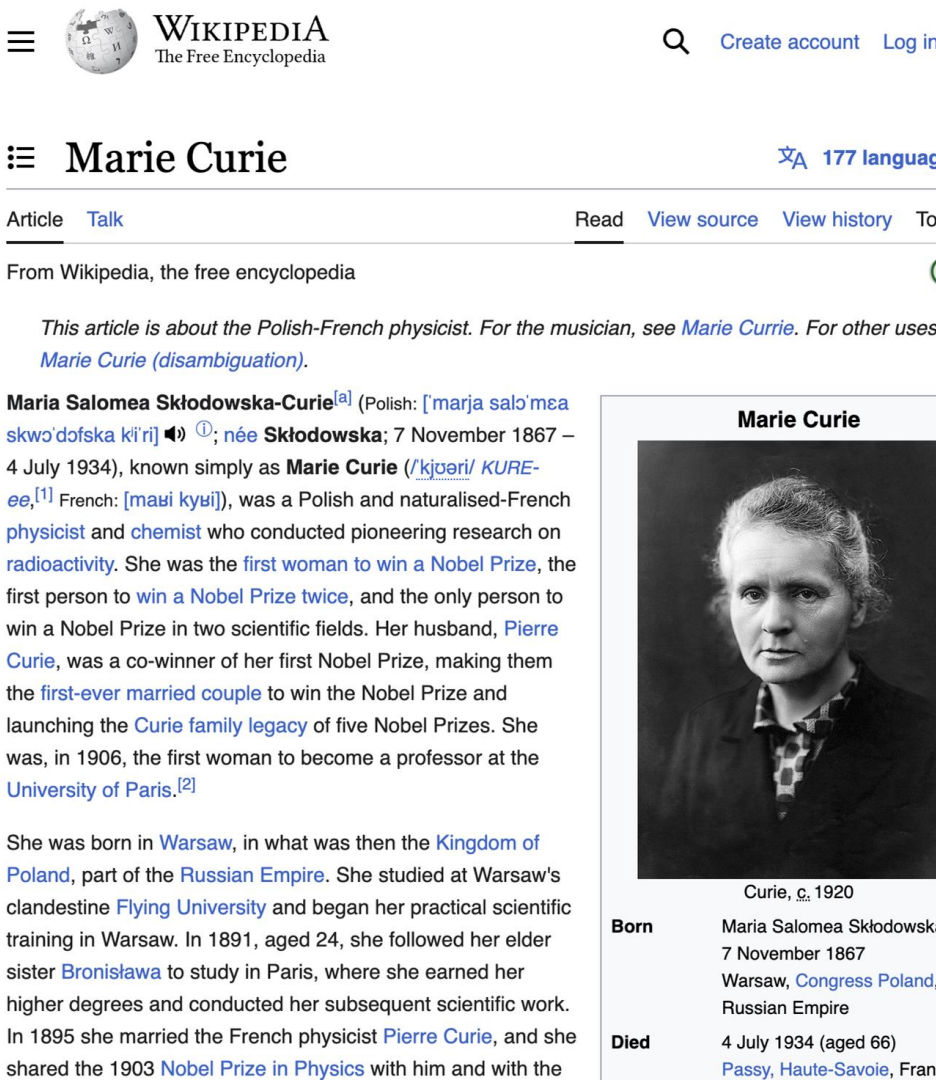
Amazon gives up on no-checkout shopping in its large grocery stores / The company will try letting customers scan while they shop, instead.

What will prevent AI from solving some of our problems?

wikipedia globe vector [no layers]

Wikipedia

- Goal - to extract life science relevant Wikipedia articles for an ML model and graph building
- A lot of the effort involved in machine learning work is in generating relevant and representative training and test sets
- Manually sifting through the 5.8 million articles in English Wikipedia to extract the life sciences subset, which is our goal, is totally unfeasible.
- But Wikipedia has an inbuilt ‘ontology’ from the World’s biggest crowd sourced website ever built



WIKIPEDIA
The Free Encyclopedia

Create account Log in

Marie Curie

177 languag

Article Talk Read View source View history To

From Wikipedia, the free encyclopedia

This article is about the Polish-French physicist. For the musician, see Marie Currie. For other uses Marie Curie (disambiguation).

Maria Salomea Skłodowska-Curie[a] (Polish: [ˈmarja salɔˈmɛa skwɔˈdɔfska kʲiˈri] ⓘ; née **Skłodowska**; 7 November 1867 – 4 July 1934), known simply as **Marie Curie** (/ˈkjʊəri/ *KURE-ee*,[1] French: [maʁi kyʁi]), was a Polish and naturalised-French physicist and chemist who conducted pioneering research on radioactivity. She was the first woman to win a Nobel Prize, the first person to win a Nobel Prize twice, and the only person to win a Nobel Prize in two scientific fields. Her husband, Pierre Curie, was a co-winner of her first Nobel Prize, making them the first-ever married couple to win the Nobel Prize and launching the Curie family legacy of five Nobel Prizes. She was, in 1906, the first woman to become a professor at the University of Paris.[2]

She was born in Warsaw, in what was then the Kingdom of Poland, part of the Russian Empire. She studied at Warsaw's clandestine Flying University and began her practical scientific training in Warsaw. In 1891, aged 24, she followed her elder sister Bronisława to study in Paris, where she earned her higher degrees and conducted her subsequent scientific work. In 1895 she married the French physicist Pierre Curie, and she shared the 1903 Nobel Prize in Physics with him and with the

Marie Curie

Curie, c. 1920

Born Maria Salomea Skłodowsk
7 November 1867
Warsaw, Congress Poland,
Russian Empire

Died 4 July 1934 (aged 66)
Passy, Haute-Savoie, Fran

Wikipedia’s ‘ontology’

- It uses **‘categories’** on every page to classify the content - which is not really an ontology, semantically fairly weak
- But seems a reasonable approach to crawling for a bio subset of Wikipedia, rich in content and well referenced in literature and contains many references for its source material!

Subcategories

This category has the following 36 subcategories, out of 36 total.

- ▶ Biology by city (1 C)
- ▶ Biology by country subdivision (1 C)
- ▶ Biology by dependent territory (5 C)
- ▶ Biology by continent (3 C)
- ▶ Biology by country (136 C)

;

- ▶ Branches of biology (46 C, 42 P)

:

- ▶ Organisms (36 C, 4 P)

*

- ▶ Biologists (18 C, 3 P)

+

- ▶ Biology-related lists (16 C, 113 P)
- ▶ Works about biology (10 C, 1 P)

C

- ▶ Biological censuses (1 C, 14 P)
- ▶ Biological classification (9 C, 49 P)
- ▶ Biological concepts (8 C, 23 P)
- ▶ Biology and culture (13 C, 30 P)

D

- ▶ Biological descriptions (3 C, 2 P)

E

- ▶ Biology education (6 C, 24 P)
- ▶ Eponyms in biology (2 C, 4 P)
- ▶ Eukaryote biology (4 C, 6 P)

F

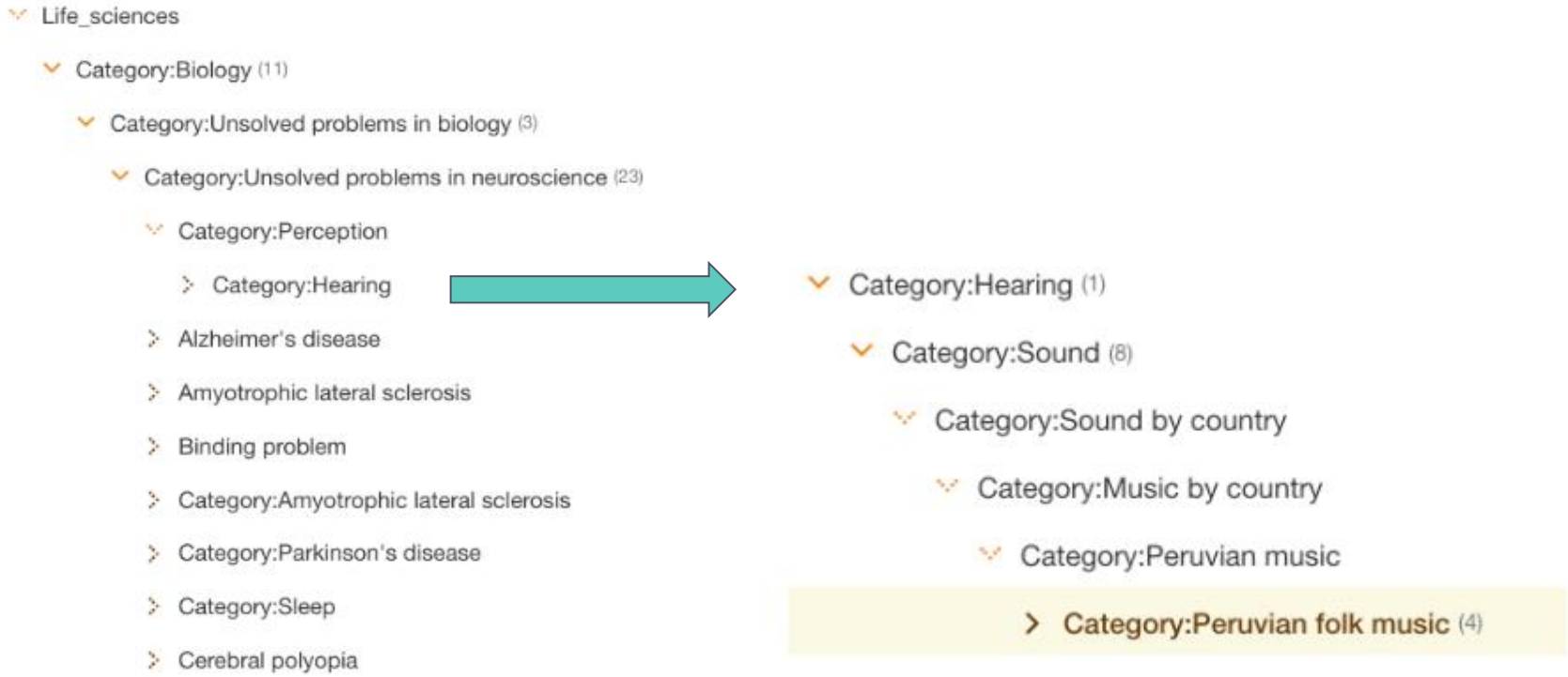
- ▶ Food science (16 C, 92 P)

H

- ▶ History of biology (14 C, 48 P)

What could possibly go wrong?

Wikipedia's 'ontology'



Wikipedia's 'ontology'

- Life_sciences
  - Category:Biology
    - Category:Unsolved problems in biology
      - Category:Ailments of unknown cause
        - Category:Senescence
          - Category:Anti-aging substances
            - Category:Antioxidants
              - Category:Dietary antioxidants
                - Category:Phytochemicals
                  - Category:Phytochemicals by taxon
                    - Category:Alkaloids found in plants
                      - Category:Alkaloids found in Malpighiales
                        - Category:Alkaloids found in Erythroxylaceae
                          - Category:Alkaloids found in Erythroxylum coca
                            - Category:Tropane alkaloids found in Erythroxylum coca

- Category:Alkaloids found in Erythroxylum coca
  - Category:Tropane alkaloids found in Erythroxylum coca
    - Category:Cocaine
      - Category:Fictional cocaine users (38)
        - → **Sherlock Holmes**
        - → Krusty the Clown

Training on Literature

- So is this a real problem?
- Arguably I've misused the category system
- So let's consider a problem closer to home
- How confident are we in the data sets we all use to train our models?

RESEARCH ARTICLE | BIOLOGICAL SCIENCES |

Meta-assessment of bias in science

Daniele Fanelli, Rodrigo Costas, and John P. A. Ioannidis Authors Info & Affiliations

Significance

Science is said to be suffering a reproducibility crisis caused by many biases. How common are these problems, across the wide diversity of research fields? We probed for multiple bias-related patterns in a large random sample of meta-analyses taken from all disciplines. The magnitude of these biases varied widely across fields and was on average relatively small. However, we consistently observed that small, early, highly cited studies published in peer-reviewed journals were likely to overestimate effects. We found little

The Lack of Good Datasets is a Major Issue for AI in Drug Discovery

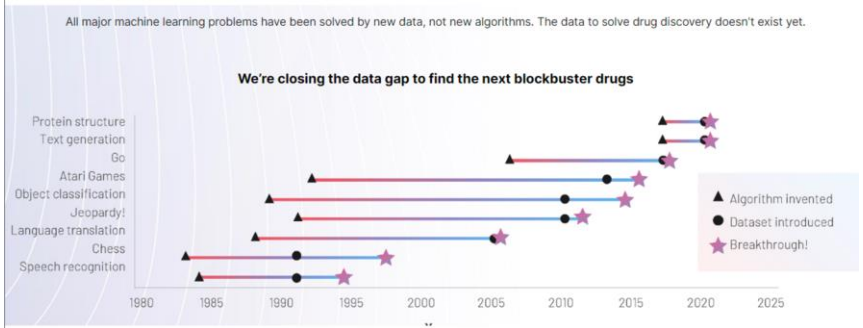
Blanco-González A, Cabezón A, Seco-González A, Conde-Torres D, Antelo-Riveiro P, Piñeiro Á, Garcia-Fandino R. The Role of AI in Drug Discovery: Challenges, Opportunities, and Strategies. *Pharmaceuticals* (Basel). 2023 Jun 18;16(6):891.

Despite the potential benefits of AI in drug discovery, there are several challenges and limitations that must be considered. **One of the key challenges is the availability of suitable data. AI-based approaches typically require a large volume of information for training purposes.** In many cases, the amount of data that is accessible may be limited, or the data may be of low quality or inconsistent, which can affect the accuracy and reliability of the results.

Source: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10302890/

89

It's The Data that Really Matters



Source: https://www.leash.bio/#data

87

So if our data is imperfect today, how do we build trust and confidence?

EASL

Evidence Attribution and Synthesis Layer

See paper for full details

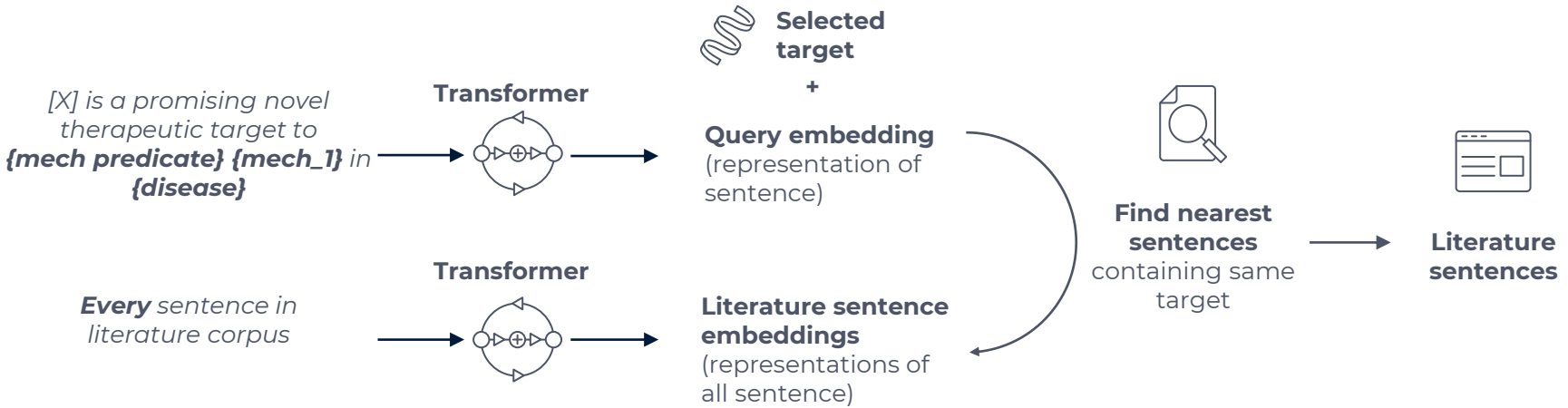
https://arxiv.org/abs/2402.04068

Scientists like to know why as well as what

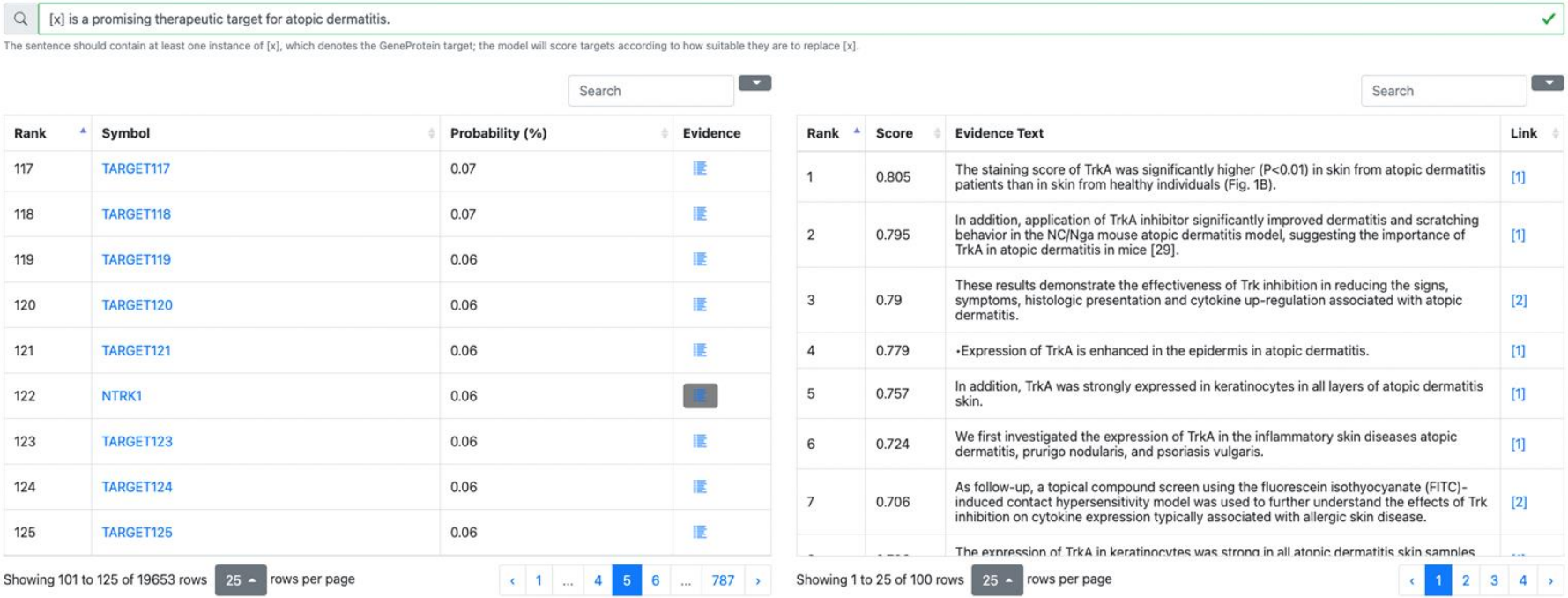
Issue: LLMs are black-box!

Evidence Surfacing with the Masked Language Model

Surfaces best literature evidence for each ranked target with FAISS indices



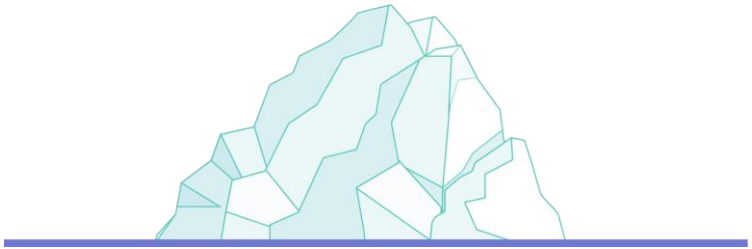
What the Drug Discovery Scientist Sees (Debug UX)



Generate predictions and evidence using a query sentence.

[x] is a promising therapeutic target for glioblastoma.

The sentence should contain at least one wildcard ([x] or [t] or [d]), where [x] or [t] denotes a target entity and [d] denotes a disease entity. The model will score targets according to how suitable they are to replace the wildcard. You cannot mix different wildcards together.

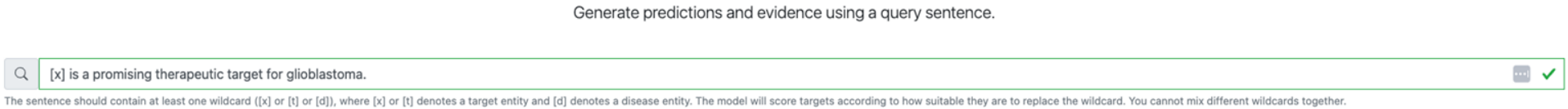


Rank 1: Therapeutically, we confirmed that inhibition of CHEK1 via either shRNAs or small molecule inhibitor could **reduce tumor growth** and **enhance radio sensitivity** in **GBM**.

Rank 7: Taken together, these data indicated that CHEK1 was preferentially expressed in **GBM**.

Rank 10: These findings indicated that CHEK1 was required for **GBM** proliferation both in vitro and in vivo.

Rank 17: Indeed, **GBM** cells are resistant to radiotherapy, and in this sense, CHK1 reduction **improves their radio-sensitivity**.



Rank 19: Moreover, by using mice xenograft model, we found that inhibition of CHEK1 by shRNA significantly reduced **tumorigenesis** and **radioresistance of U87 cells** (Figure 3G).

Rank 51: CHK1 inhibition also increased the cytotoxicity of **TMZ** in a p53 independent manner [126].

Rank 65: For example, targeting CHK1 significantly enhances cell killing effect by chemotherapy or radiation therapy in **ovarian, triple negative breast, and brain cancers**.

Rank 66: Most importantly, our analysis revealed that elevated expression of CHK1 correlated with poor long-term survival in patients with Group 3 subgroup **medulloblastoma**.

Generate predictions and evidence using a query sentence.

Rank 16: Moreover, by using mice xenograft model, we found that inhibition of CHEK1 by shRNA significantly reduced **tumorigenesis** and **radioresistance of U87 cells** (Figure 3G).

relevant mechanisms and cell lines

Rank 51: CHK1 inhibition also increased the cytotoxicity of **TMZ** in a p53 independent manner [126].

complements existing GBM drug

Rank 65: For example, targeting CHK1 significantly enhances cell killing effect by chemotherapy or radiation therapy in **ovarian, triple negative breast, and brain cancers**.

related diseases

Rank 66: Most importantly, our analysis revealed that elevated expression of CHK1 correlated with poor long-term survival in patients with Group 3 subgroup **medulloblastoma**.

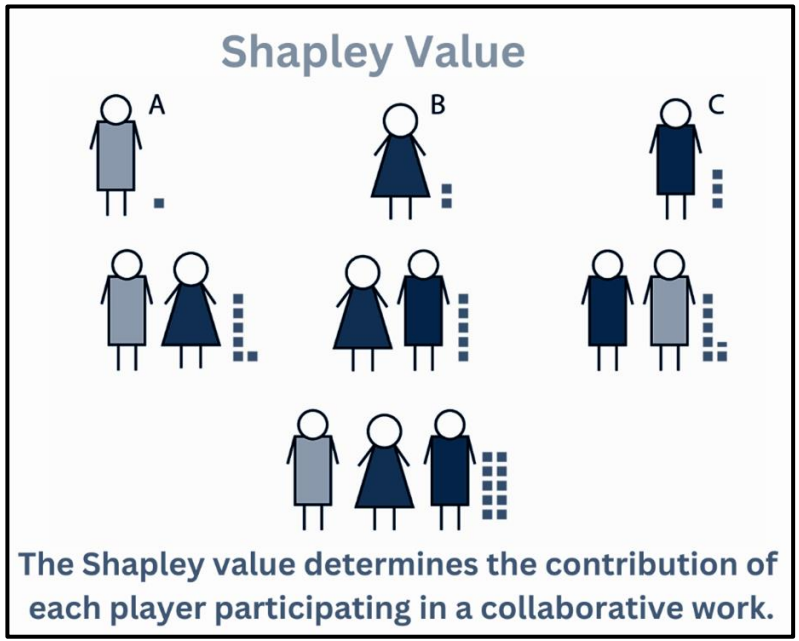
related diseases

Issue: Target prediction and evidence surfacing are decoupled

How else could we explain a prediction?

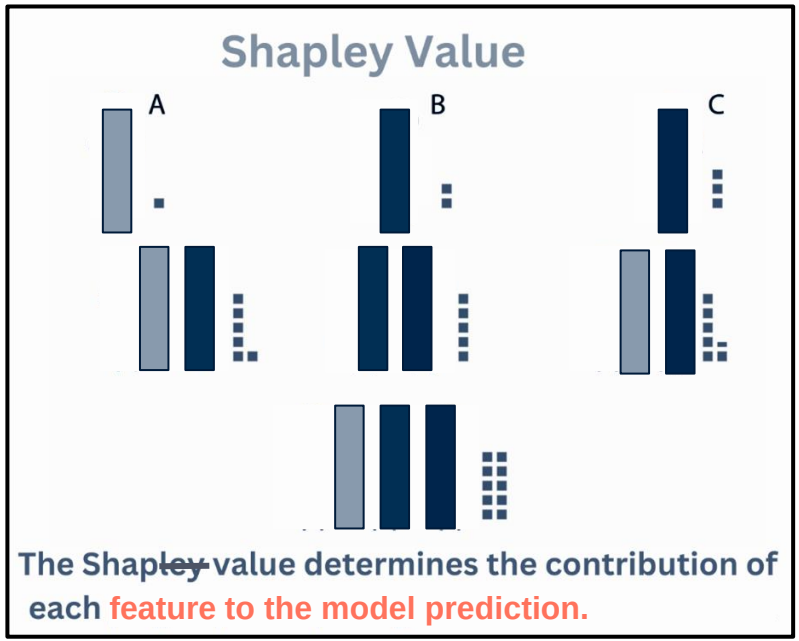
Can we use SHAP?

(***SHAP is an approximation to Shapley***)



Can we use SHAP?

(***SHAP is an approximation to Shapley***)



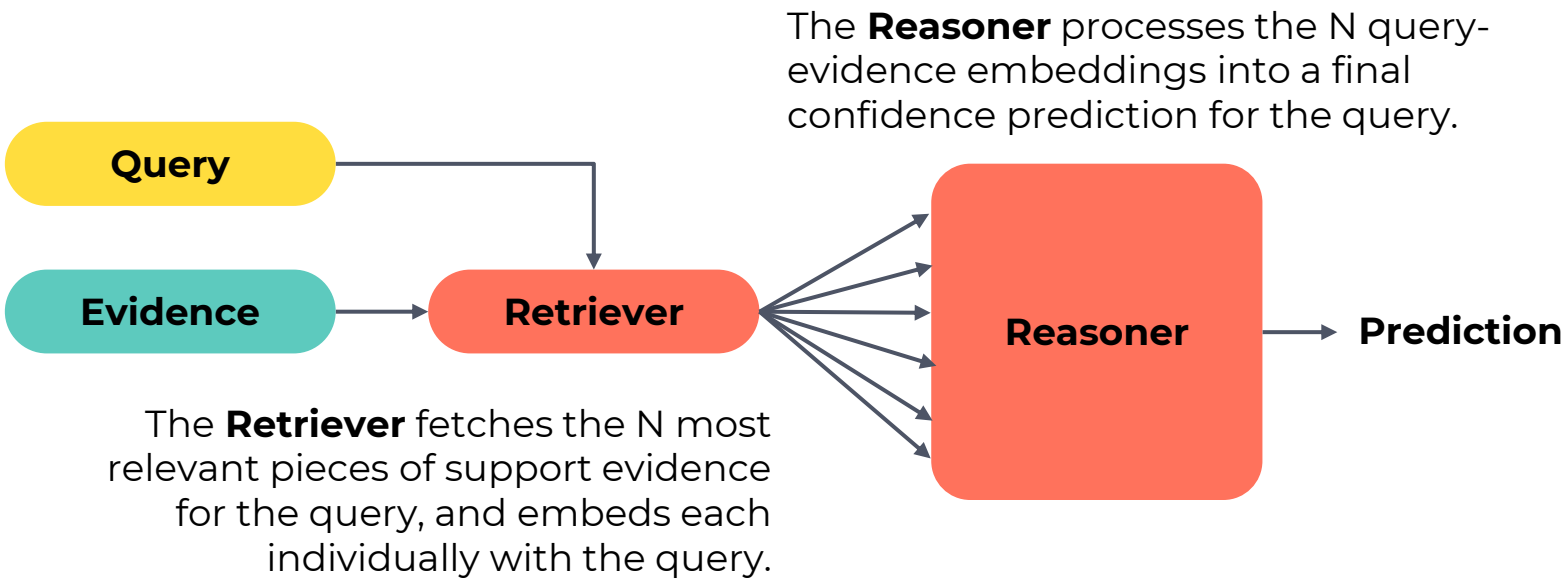
No because SHAP attributes contribution to input features

But we could IF prediction was based on the evidence

We need a new model architecture!

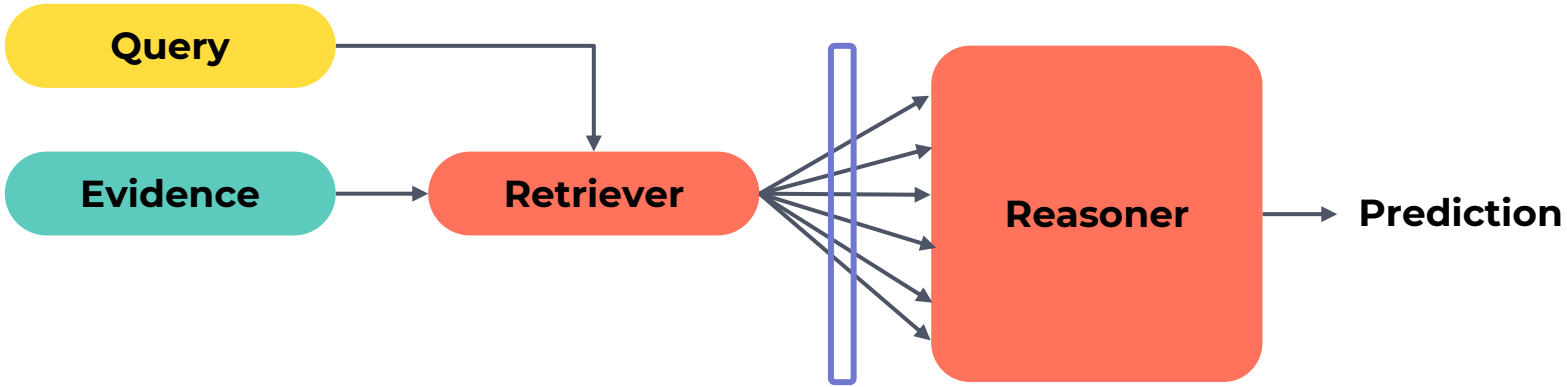
Every MLOps Engineer’s favourite phrase

Retrieval-Driven Reasoning



Retrieval-Driven Reasoning

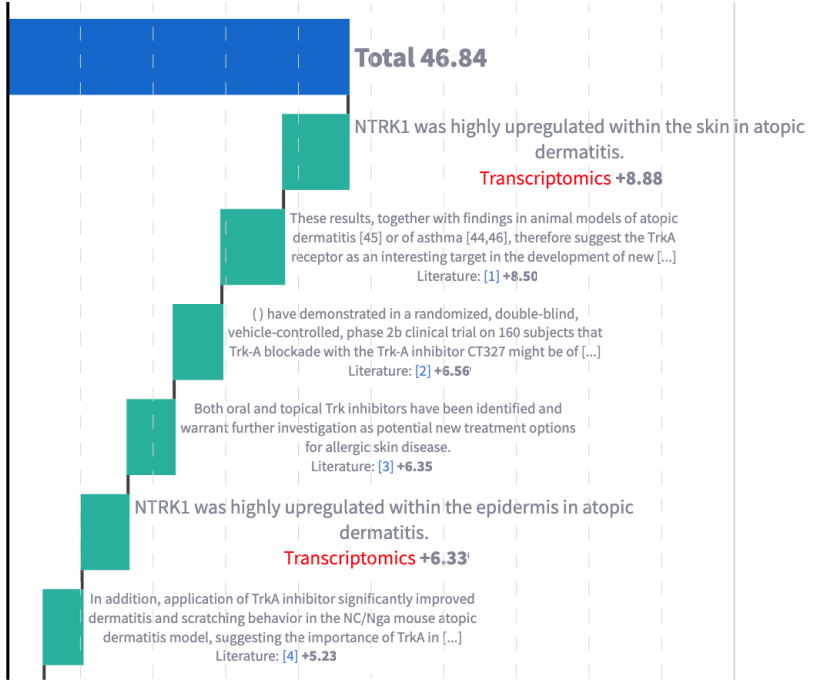
We use **SHAP** to estimate the **contribution** of each piece of evidence to the final prediction



User workflow: Example

Query
NTRK1 is a novel promising therapeutic target for atopic dermatitis.

Summarised explanation
*There is **moderate evidence** that NTRK1 is a promising therapeutic target for atopic dermatitis. NTRK1 is highly upregulated in the skin in atopic dermatitis, and TrkA blockade has previously been identified as a promising treatment for psoriasis [2]...*



•••

User workflow: Example

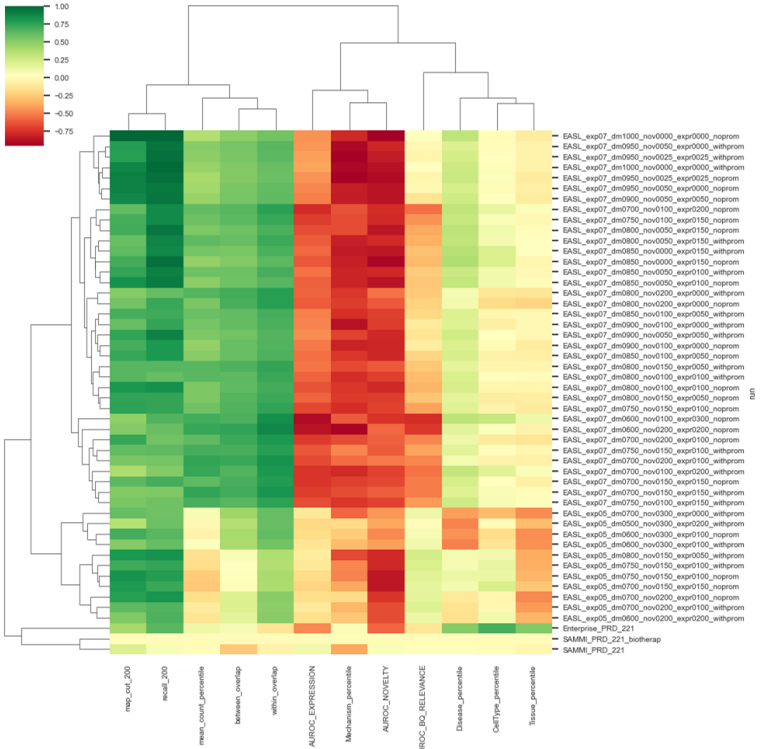
Query
NTRK1 is a novel promising therapeutic target for atopic dermatitis.

New evidence (including **proprietary customer data**) can be added **without retraining** the model. Also **textualised structured data** can also be added.

•••

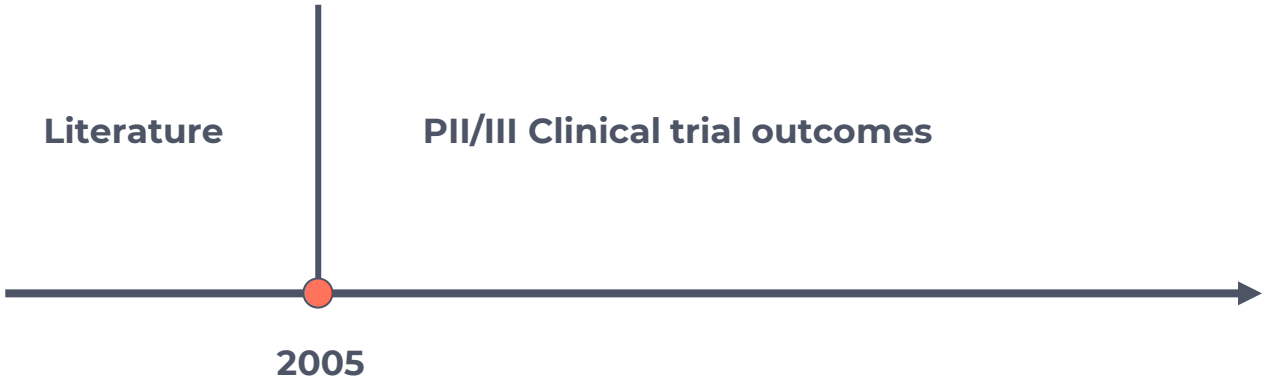
How does EASL perform?

- Comparable metrics-wise to our vanilla LLM
- But with
  - explainability
  - multimodal reasoning
  - dynamic corpora

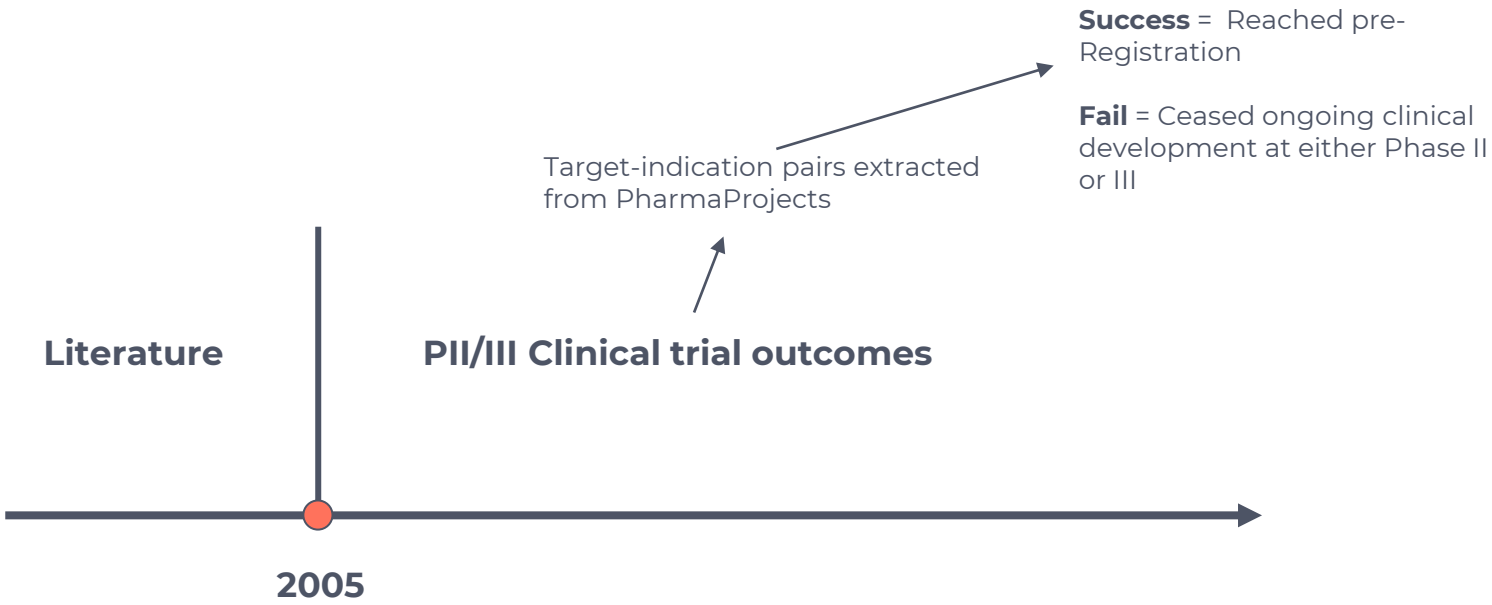


Can EASL predict clinical trial outcomes?

EASL: Evaluated on time-split clinical trial outcomes



EASL: Evaluated on time-split clinical trial outcomes

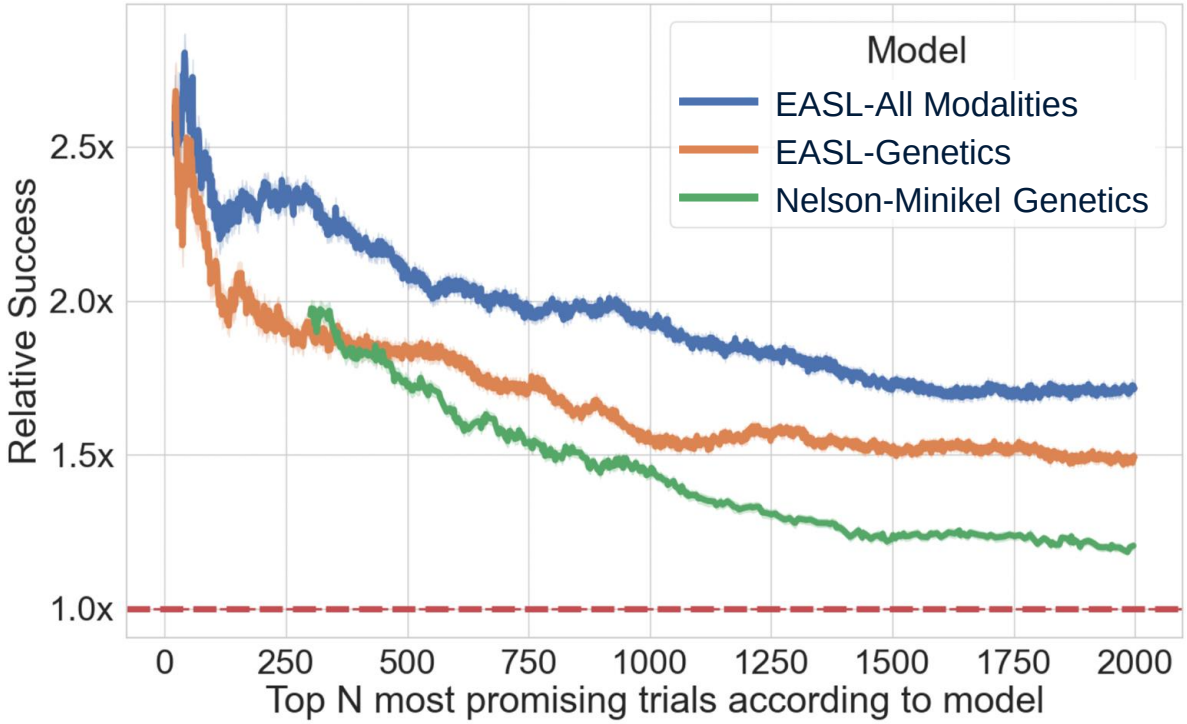


EASL outperforms industry standard indicators to predict clinical success of drug targets

Retrospective prediction of post-2005 PhII/III **clinical trial efficacy** outcomes, using EASL trained only on evidence published before 2005.

We compare how much more likely the top N trials are to succeed as predicted by the

- **industry-leading genetics approach**
- **EASL using the same genetics data**
- **EASL using all modalities**



We can develop successful AI by considering the user, the problem, and the data

The Clever Kids - BenevolentAI Research

Thank you

 benevolent.com

 @benevolent_ai

 benevolentai

 hello@benevolent.ai