Benevolent^AI

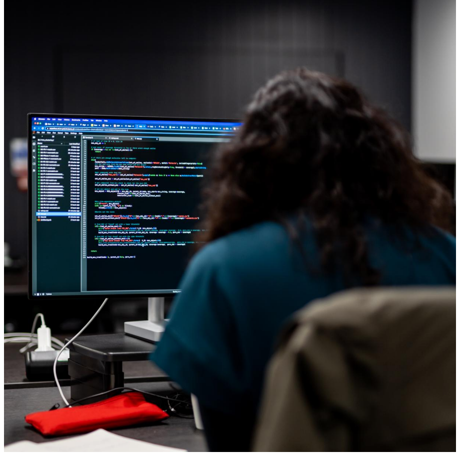
Language models for target prioritisation

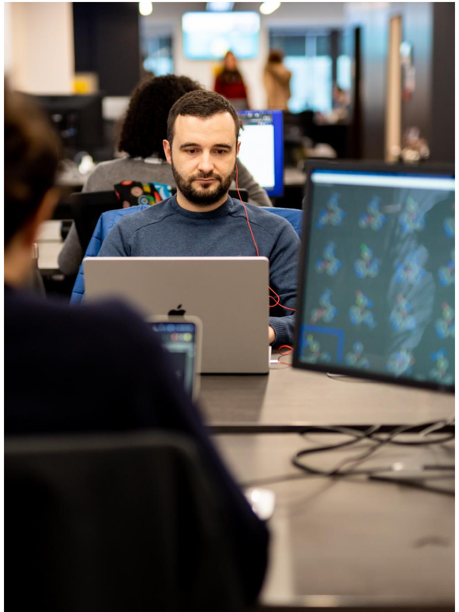
Genome-wide | Multimodal | Explainable

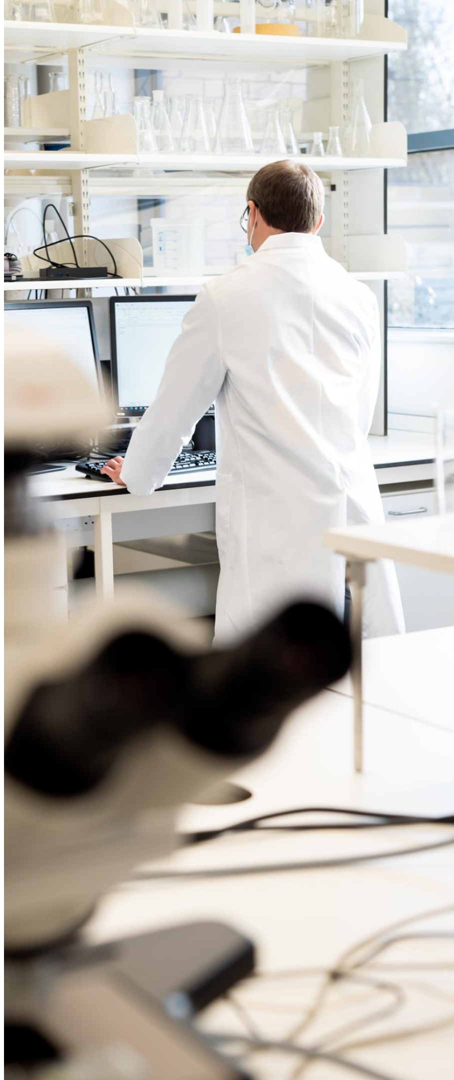
Ravi Patel | July 16, 2023

BenevolentAI | ravi.patel@benevolent.ai

Oxford Machine Learning Summer School







Contents

Drug discovery & BenevolentAI

Huge burden on society demands a new approach

90%
failure rate in clinical development

$2.6bn
in average R&D and to market cost per drug

10 years
to market

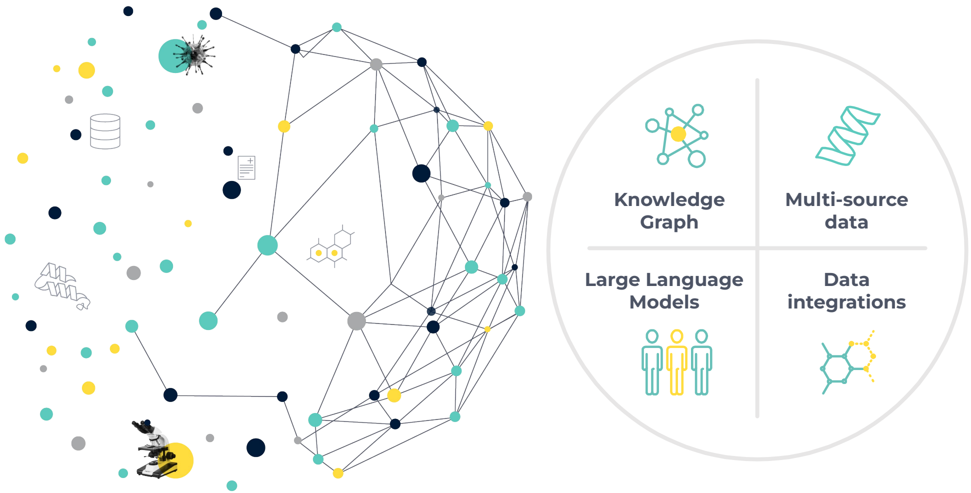
30-50%
of patients respond poorly to leading drugs

Gaining a clear understanding of the **underlying molecular mechanisms of disease** based on the **totality of available biomedical data** is a vital step in the development of successful and efficacious treatments

BenevolentAI develops technology in the service of science

BenevolentAI applies deep expertise in developing **biology-specific AI models**, including sophisticated and **comprehensive data foundations** and **NLP**, to enable novel biomedical discoveries.

The Benevolent Platform™



COMMERCIAL VALIDATION

5 novel targets selected for **AstraZeneca's** portfolio

REGULATORY VALIDATION

FDA approval of **COVID-19 treatment** identified by BenevolentAI

SCIENTIFIC VALIDATION

Robust wholly owned drug pipeline

Data modalities

Patient-level data

Omics, genetics, clinical data etc.

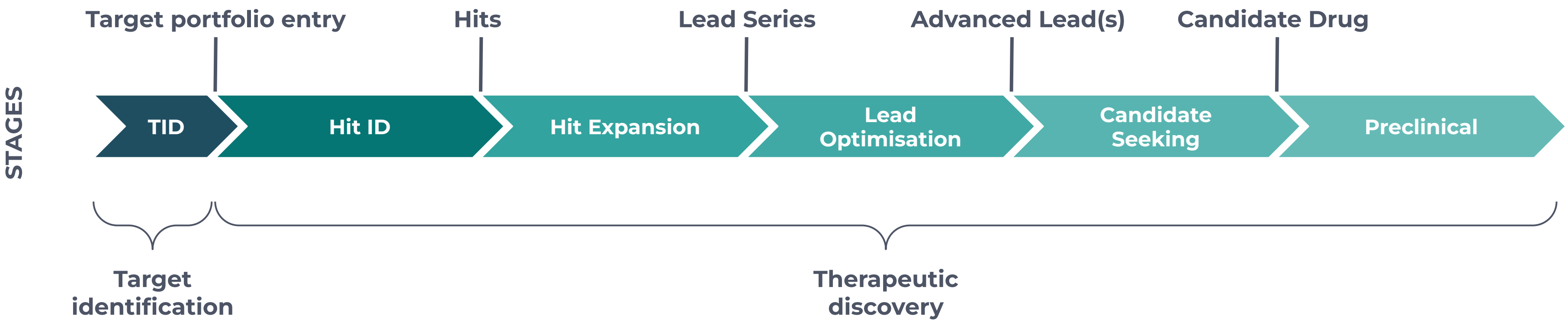
Unstructured Literature

Millions of ingested full-text biomedical publications.

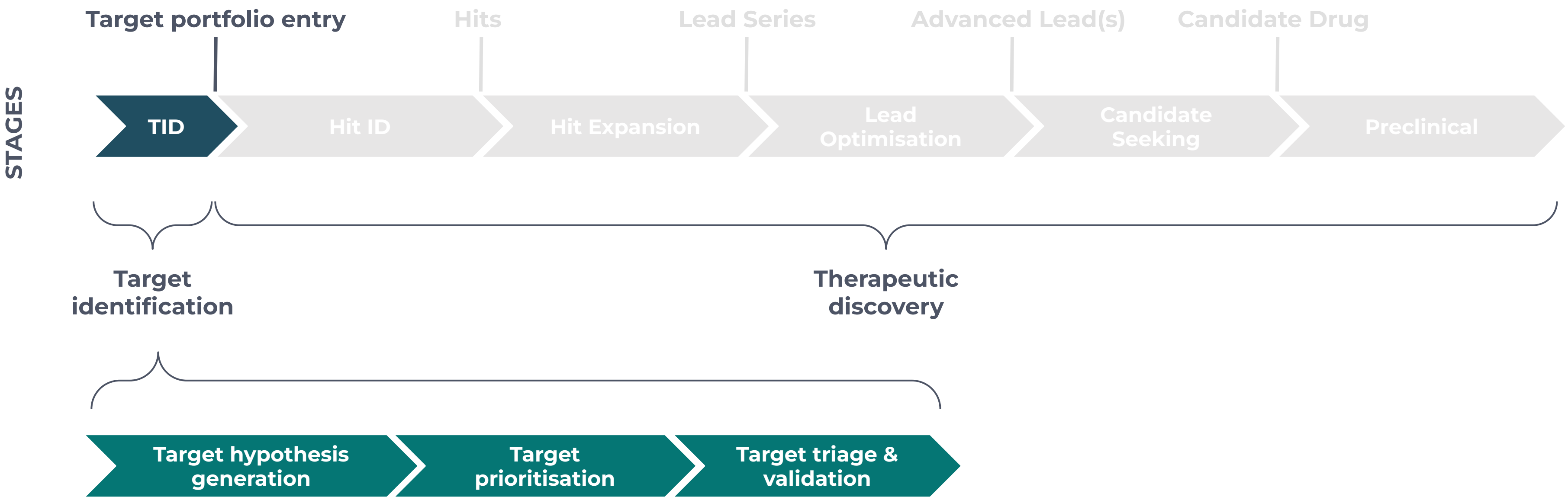
Knowledge Graph

A large-scale biomedical knowledge graph combining patient-level data, literature and structured sources.

Early stages of drug discovery



Early stages of drug discovery



Target hypothesis generation

Target Prioritisation

Target triage and Validation

Target hypothesis generation

Drug discovery biologists explore our data and **define strategic hypotheses** aligned with assay strategies to use as inputs into our AI models and custom data queries

Disease

Endpoint: What we are measuring in the assay

Cell type

Can we treat T2DM by reversing insulin resistance in adipocytes by reducing oxidative stress?

Sign

Mechanism

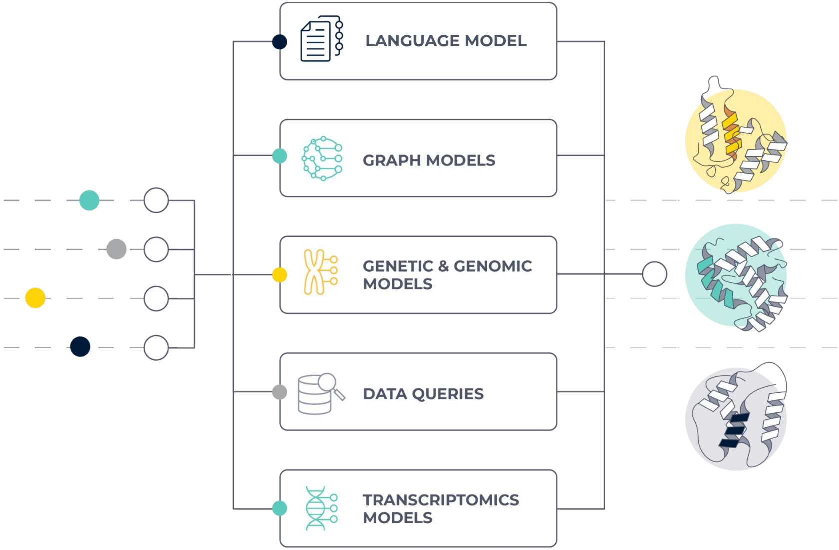
Target prioritisation

Target hypothesis generation

Target Prioritisation

Target triage and Validation

Our **models** combine our large literature corpus, knowledge graph and patient data to predict which targets are the most **efficacious** and **progressible** for the user's biological context



Target hypothesis generation

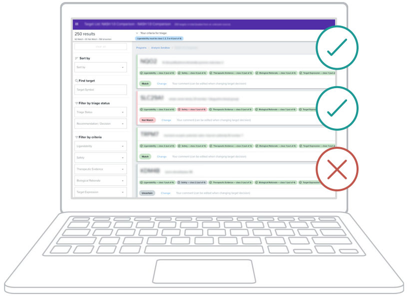
Target Prioritisation

Target triage and Validation

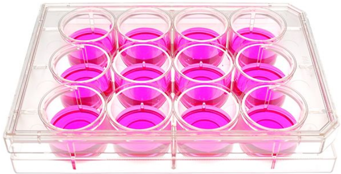
Target triage & validation

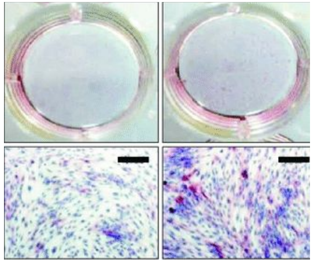
Biology and chemistry expert **triage of targets** for biological relevance and suitability for portfolio entry

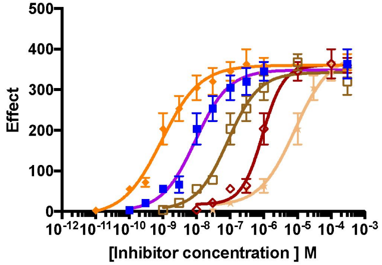
Shortlisted targets are validated in **assays**



Triage

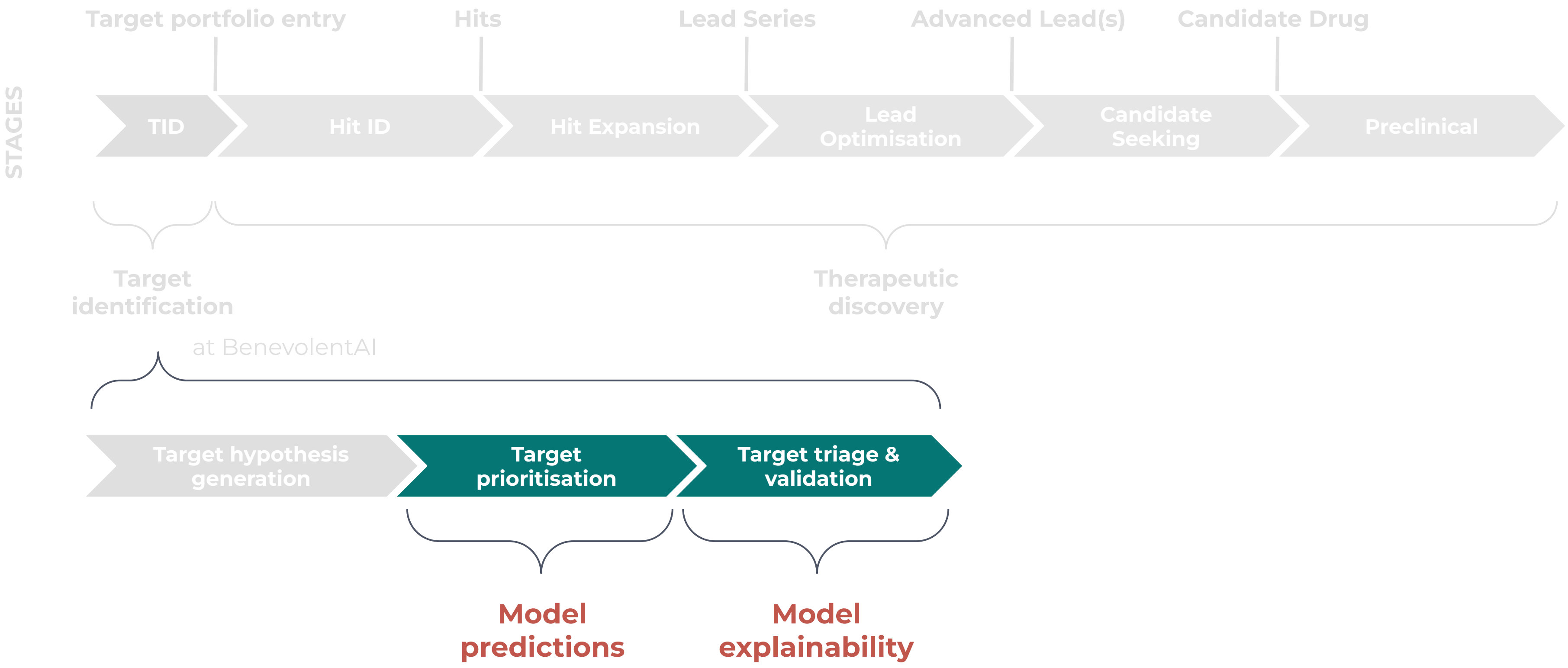






Validation

Our focus today: Target prioritisation and validation



Drug discovery biologist’s approach

Drug discovery scientists and target prioritisation

Drug discovery scientists:

- Can integrate **multiple tradeoffs** and **evidence from different data modalities,** but can only consider a **limited** number of **targets**

Drug discovery scientists and target prioritisation

Drug discovery scientists:

- Can integrate **multiple tradeoffs** and **evidence from different data modalities,** but can only consider a **limited** number of **targets**

Tradeoffs

Disease relevance

Mechanism relevance

Novelty

Expression

Assayability

...

Data modalities

Literature

Patient-level data

- Genetics
- Expression
- Differential expression

...

Knowledge graph

...

Drug discovery scientists and target prioritisation

Drug discovery scientists:

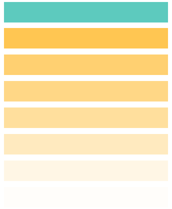
- Can integrate **multiple tradeoffs** and **evidence from different data modalities,** but can only consider a **limited** number of **targets**
- Gain **deep insight** from **papers** and **datasets**, but can only look at a **limited number**

Drug discovery scientists and target prioritisation

Drug discovery scientists:

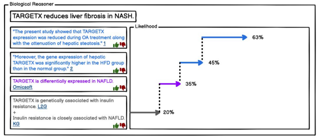
- Can integrate **multiple tradeoffs** and **evidence from different data modalities,** but can only consider a **limited** number of **targets**
- Gain **deep insight** from **papers** and **datasets**, but can only look at a **limited number**
- Could be complemented by model-based prioritisation, but have a **strong requirement for explainability**

Target prioritisation model wishlist



Genome-wide predictions (relevant and novel)

Reason from multiple data modalities



Explainable predictions

Masked-language modelling approach

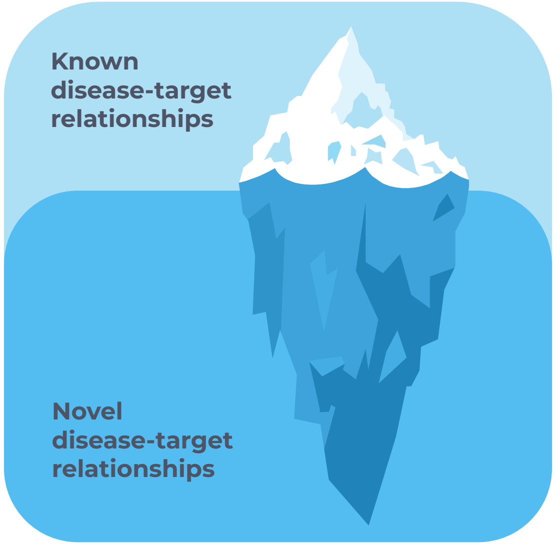
- Direct and indirect evidence
- User workflow
- Model setup: Training, inference, post-processing
- Explainability

Direct and indirect evidence

Levels of insight from the literature

Simple Pattern Matching

Semantic Awareness



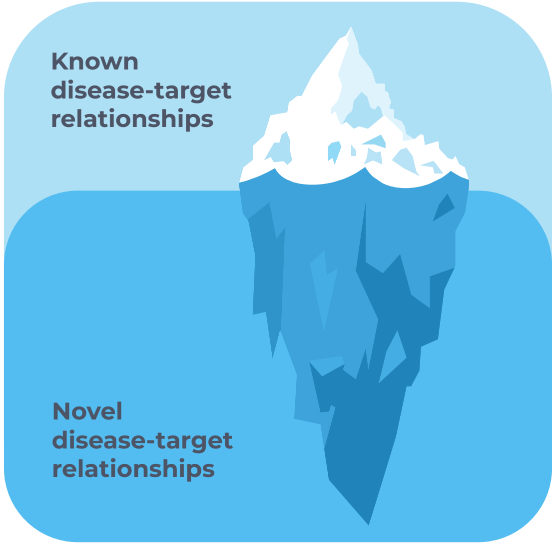
Levels of insight from the literature

Simple Pattern Matching

Identify *direct* relationships between target and disease

E.g. **NTRK1** and **atopic dermatitis** co-occur in many papers
➔ Can find **known** (published) relationships

Semantic Awareness



Levels of insight from the literature

Simple Pattern Matching

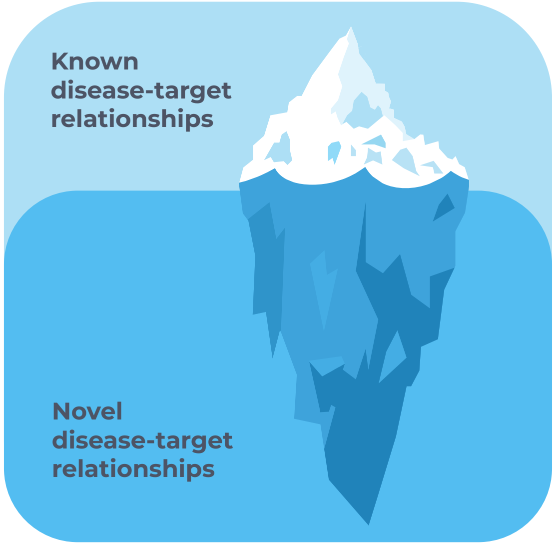
Identify *direct* relationships between target and disease

E.g. **NTRK1** and **atopic dermatitis** co-occur in many papers
➔ Can find **known** (published) relationships

Semantic Awareness

Identify *indirect* relationships between target and disease

E.g. **NTRK1** found to be involved in other similar diseases of allergic inflammation or mechanisms related to atopic dermatitis
➔ Can uncover **novel** relationships.



User workflow

User enters a query sentence containing a blank. This can be a **completely novel sentence** that hasn't appeared in any publication.

Generate target predictions and evidence using a query sentence.

[x] is a promising therapeutic target for atopic dermatitis.

The sentence should contain at least one instance of [x], which denotes the GeneProtein target; the model will score targets according to how suitable they are to replace [x].

Search

Rank	Symbol	Probability (%)	Evidence
1	TARGET1	30.98	
2	TARGET2	11.12	
3	TARGET3	3.76	
4	TARGET4	2.29	
5	TARGET5	2.03	
6	TARGET6	1.97	
7	TARGET7	1.75	
8	TARGET8	1.53	
9	TARGET9	1.51	

Showing 1 to 25 of 19653 rows 25 rows per page

‹ 1 2 3 4 5 ... 787 ›

User workflow

Generate target predictions and evidence using a query sentence.

[x] is a promising therapeutic target for atopic dermatitis.

The sentence should contain at least one instance of [x], which denotes the GeneProtein target; the model will score targets according to how suitable they are

Search

Rank	Symbol	Probability (%)	Evidence
1	TARGET1	30.98	
2	TARGET2	11.12	
3	TARGET3	3.76	
4	TARGET4	2.29	
5	TARGET5	2.03	
6	TARGET6	1.97	
7	TARGET7	1.75	
8	TARGET8	1.53	
9	TARGET9	1.51	

Showing 1 to 25 of 19653 rows 25 rows per page

‹ 1 2 3 4 5 ... 787 ›

User receives the full genome scored and ranked according to the probability of filling the blank position in the query.

User workflow

Generate target predictions and evidence using a query sentence.

[x] is a promising therapeutic target for atopic dermatitis.

The sentence should contain at least one instance of [x], which denotes the GeneProtein target; the model will score targets according to how suitable they are to replace [x].

Search

Rank	Symbol	Probability (%)	Evidence
117	TARGET117	0.07	
118	TARGET118	0.07	
119	TARGET119	0.06	
120	TARGET120	0.06	
121	TARGET121	0.06	
122	NTRK1	0.06	
123	TARGET123	0.06	
124	TARGET124	0.06	
125	TARGET125	0.06	

Showing 101 to 125 of 19653 rows 25 rows per page

‹ 1 ... 4 5 6 ... 787 ›

Search

Rank	Score	Evidence Text	Link
1	0.805	The staining score of TrkA was significantly higher (P<0.01) in skin from atopic dermatitis patients than in skin from healthy individuals (Fig. 1B).	[1]
2	0.795	In addition, application of TrkA inhibitor significantly improved dermatitis and scratching behavior in the NC/Nga mouse atopic dermatitis model, suggesting the importance of TrkA in atopic dermatitis in mice [29].	[1]
3	0.79	These results demonstrate the effectiveness of Trk inhibition in reducing the signs, symptoms, histologic presentation and cytokine up-regulation associated with atopic dermatitis.	[2]
4	0.779	•Expression of TrkA is enhanced in the epidermis in atopic dermatitis.	[1]
5	0.757	In addition, TrkA was strongly expressed in keratinocytes in all layers of atopic dermatitis skin.	[1]
6	0.724	We first investigated the expression of TrkA in the inflammatory skin diseases atopic dermatitis, prurigo nodularis, and psoriasis vulgaris.	[1]
7	0.706	As follow-up, a topical compound screen using the fluorescein isothyocyanate (FITC)-induced contact hypersensitivity model was used to further understand the effects of Trk inhibition on cytokine expression typically associated with allergic skin disease.	[2]
[illegible]	[illegible]	The expression of TrkA in keratinocytes was strong in all atopic dermatitis skin samples	[illegible]

Showing 1 to 25 of 100 rows 25 rows per page

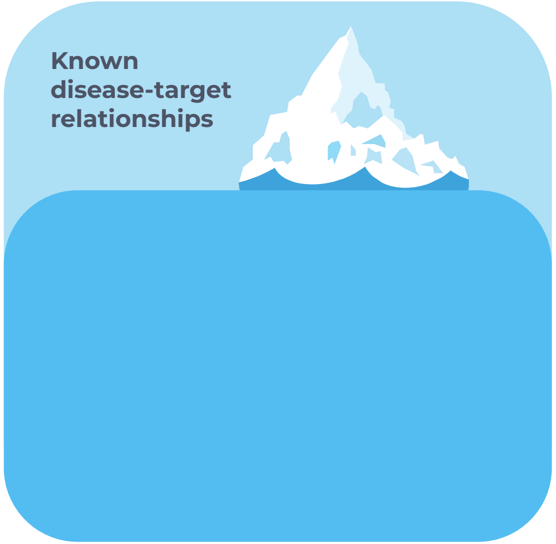
‹ 1 2 3 4 ›

User workflow

Generate target predictions and evidence using a query sentence.

[x] is a promising therapeutic target for atopic dermatitis.

The sentence should contain at least one instance of [x], which denotes the GeneProtein target; the model will score targets according to how suitable they are to replace [x].



Rank 1: The staining score of [NTRK1] was significantly higher (P<0.01) in skin from **atopic dermatitis** patients than in skin from healthy individuals (Fig. 1B).

Rank 2: In addition, application of [NTRK1] inhibitor significantly improved dermatitis and scratching behavior in the NC/Nga mouse **atopic dermatitis** model, suggesting the importance of [NTRK1] in **atopic dermatitis** in mice [29].

Rank 3: These results demonstrate the effectiveness of [NTRK1] inhibition in reducing the signs, symptoms, histologic presentation and cytokine up-regulation associated with **atopic dermatitis**.

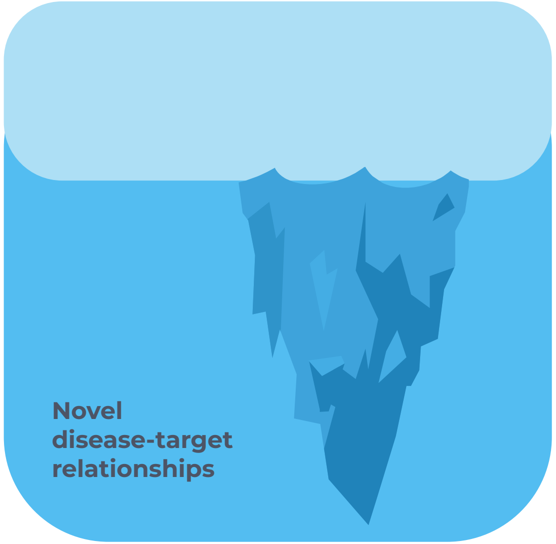
Rank 4: Expression of [NTRK1] is enhanced in the epidermis in **atopic dermatitis**.

User workflow

Generate target predictions and evidence using a query sentence.

[x] is a promising therapeutic target for atopic dermatitis.

The sentence should contain at least one instance of [x], which denotes the GeneProtein target; the model will score targets according to how suitable they are to replace [x].



Rank 12: Translational studies showed elevated expression of [NTRK1] in allergic tissue from patients with EoE.

relevant pathology

Rank 15: Our results demonstrate that inhibition of [NTRK1] decreases IL-4-induced proliferation of keratinocytes and that [NTRK1] overexpression in keratinocytes enhances cell proliferation.

relevant symptomatic mechanisms

Rank 17: [NTRK1] is also involved in NGF-related allergic reaction.

relevant pathology

Rank 21: In summary, we conclude that significant enhancement of [NTRK1] expression by Th2 cytokines regulates proliferation and differentiation of keratinocytes as well as allergic inflammation.

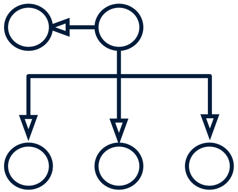
relevant symptomatic and causative mechanisms

User workflow

Generate target predictions and evidence using a query sentence.

[x] is a promising therapeutic target for atopic dermatitis.

The sentence should contain at least one instance of [x], which denotes the GeneProtein target; the model will score targets according to how suitable they are to



The masked language model enables **fast full-genome ranking** for a textual target discovery hypothesis, accounting for both **direct** and **indirect evidence**.

Rank 12: Translational studies showed elevated expression of [NTRK1] in allergic tissue from patients with EoE.

relevant pathology

Rank 15: Our results demonstrate that inhibition of [NTRK1] decreases IL-4-induced proliferation of keratinocytes and that [NTRK1] overexpression in keratinocytes enhances cell proliferation.

relevant symptomatic mechanisms

Rank 17: [NTRK1] is also involved in NGF-related allergic reaction.

relevant pathology

Rank 21: In summary, we conclude that significant enhancement of [NTRK1] expression by Th2 cytokines regulates proliferation and differentiation of keratinocytes as well as allergic inflammation.

relevant symptomatic and causative mechanisms

Model setup

Model setup: Overview

Transformer-based masked language model

150million+ entity-linked real-world scientific literature sentences

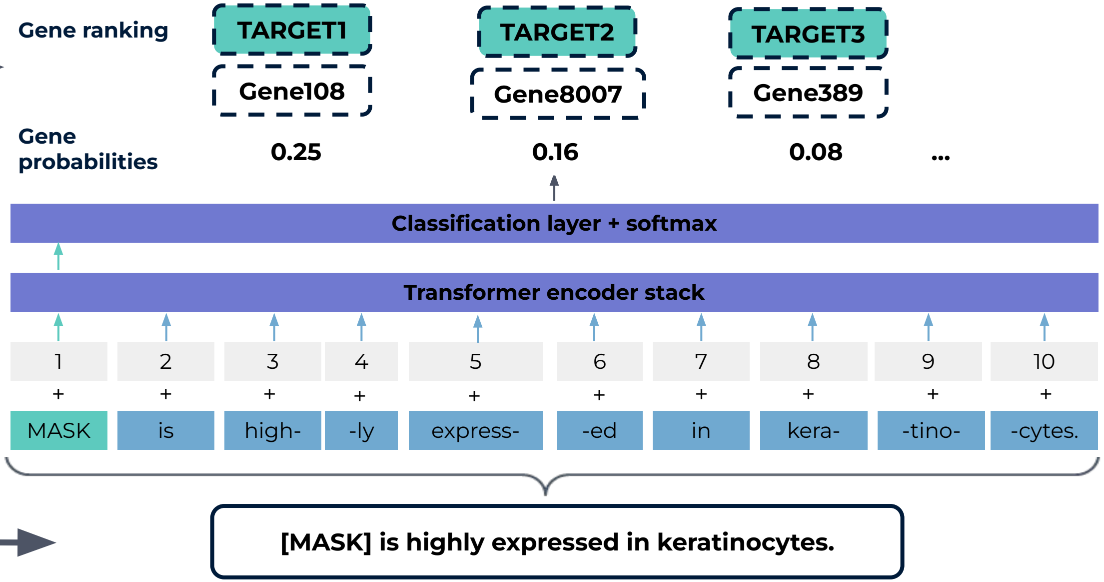
Trained to predict **genes** from a **text-based query**

- → Semantic representation of biology
  - → Predict novel targets

Model setup: Training

Training output: Model parameters are updated to increase the softmax score of the true MASKed gene.

Training input: A sentence where all appearances of the same gene are masked.

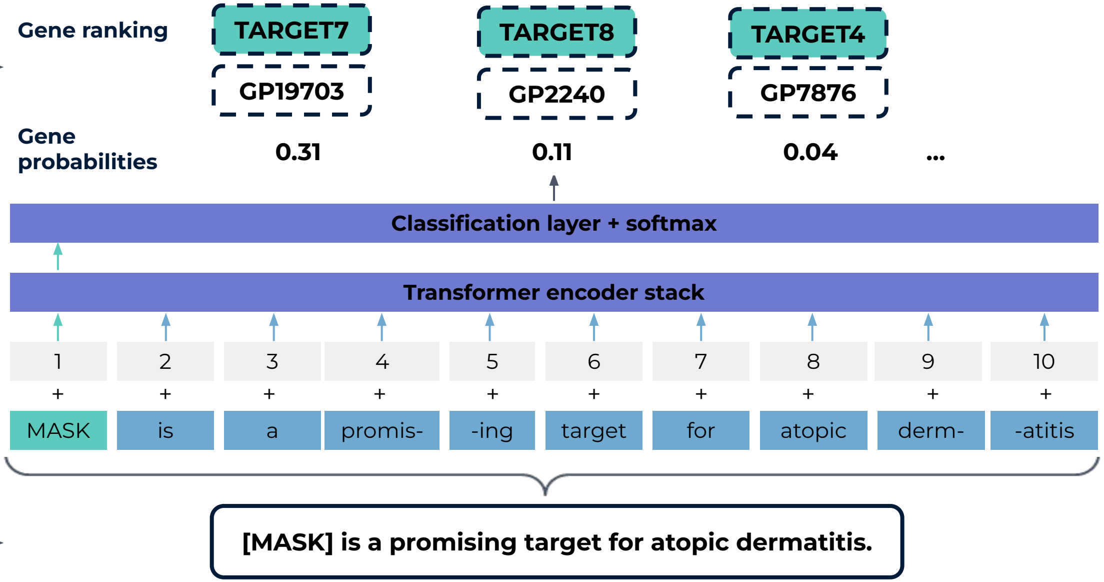


BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Devlin et al., 2018.

Model setup: Inference

Inference output: Targets ranked according to softmax score.

Inference input: Sentence is a user's "fill-in-the-blank" style query containing a MASK directly.



BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Devlin et al., 2018.

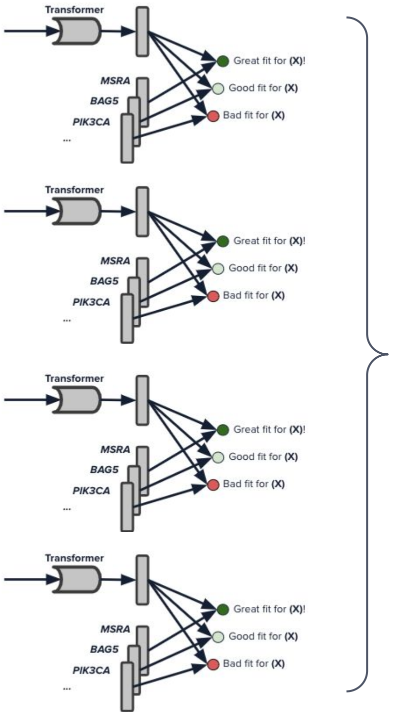
Model setup: Post-processing

[X] is a potential therapeutic target for ***{disease}***

[X] modulates ***{mechanism}***

[X] is expressed in ***{tissue}***

[X] is expressed in ***{cell type}***



Aggregate scores → Literature bias correction → **Ranked genome**

Templated sentences from user-defined context entities

Model setup: Post-processing

Correcting for literature bias

score(target|query) = log(p(target|query)) - ***c*****log(****p(target)****)*

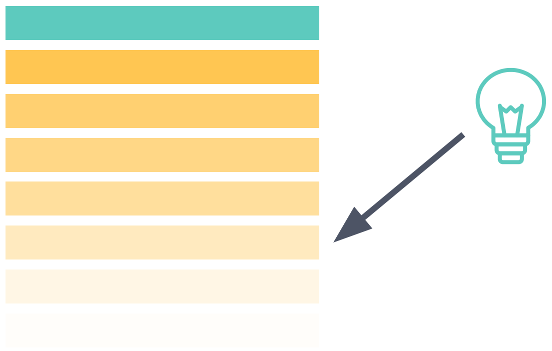
Where *0 <=* **c** *<= 1*

and ***p(target)*** *= (nb of sentences with target) / (nb of sentences)*

Model setup: Evaluation

Tens of thousands of in-house target annotations from drug discovery scientists

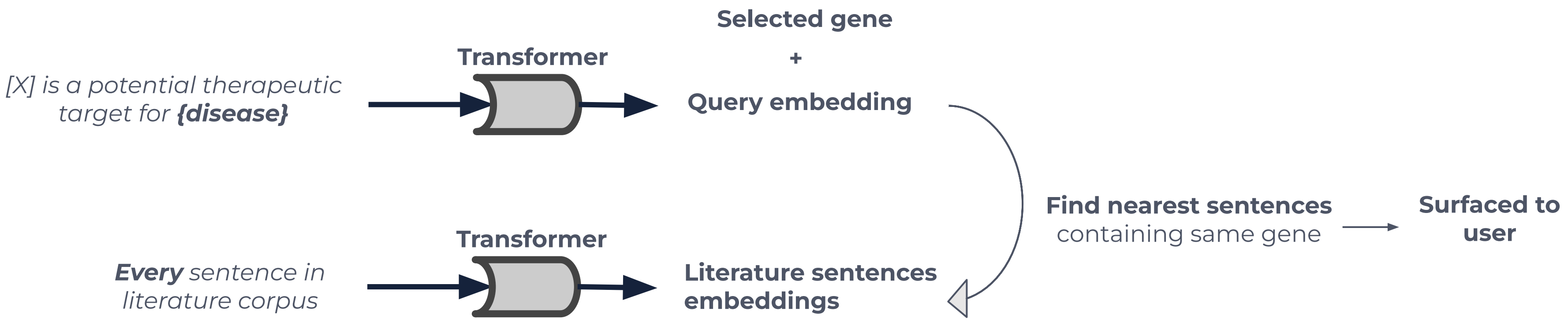
- Wide range of diseases
- **Biological rationale** and **novelty** annotations
- Highly relevant targets near the top of the list (e.g. Ranks 1-10)
- Continues to find **promising, novel targets** further down in the list (e.g. Ranks 100-200)



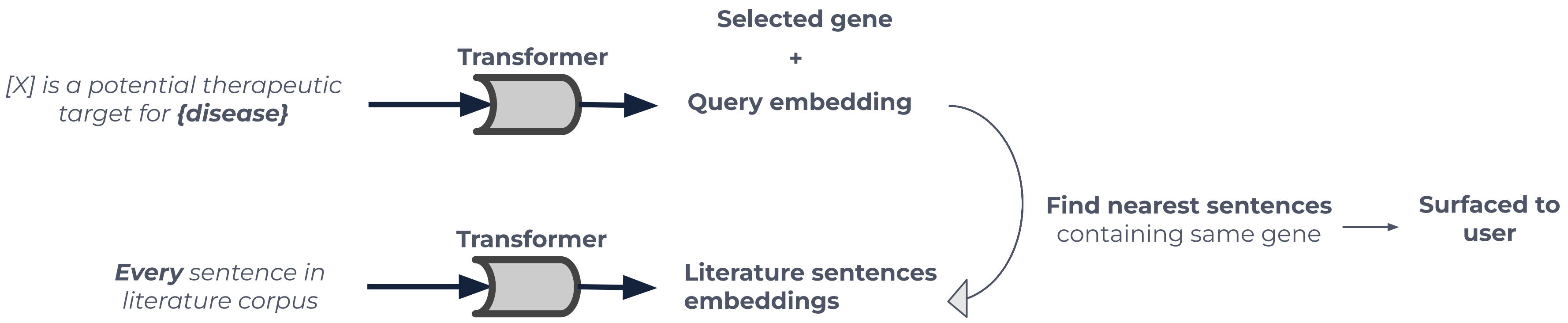
Explainability

Indirect explainability via retrieval

Indirect explainability via retrieval



Indirect explainability via retrieval

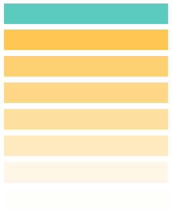


Literature sentence embeddings stored in **per-gene FAISS* indices** - for efficient vector similarity search

*Johnson, J et al. (2019). Billion-scale similarity search with GPUs. IEEE Transactions on Big Data, 7(3), 535–547.

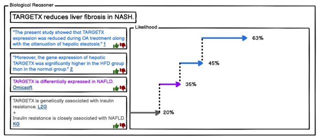
Target prioritisation wish list

Target prioritisation model wishlist



Genome-wide predictions (relevant and novel) ✅

Reason from multiple data modalities ❌



Explainable predictions ✅/ ❌

A moment to breathe and ask questions

Before breath: 😴

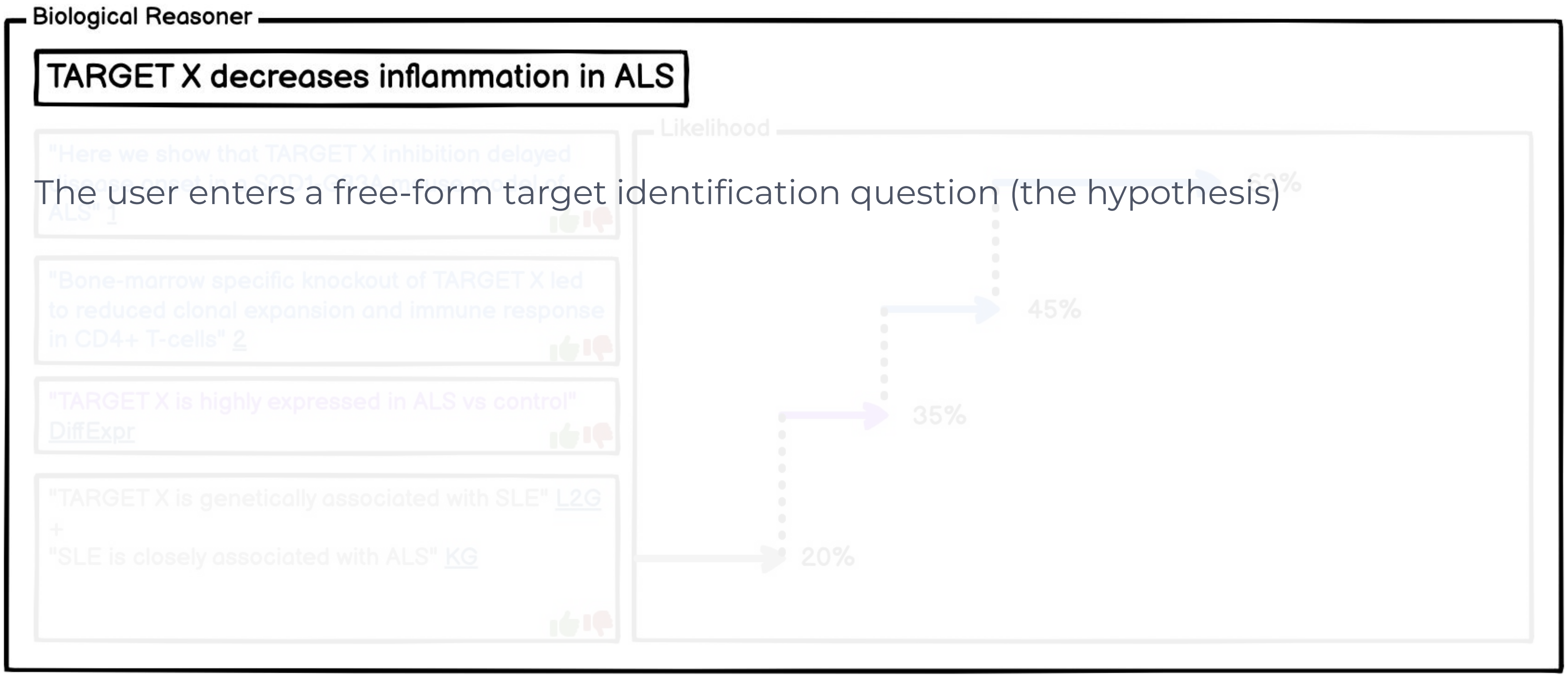
After breath: 👶

Evidence synthesis with attribution

- User workflow
- Model setup: Training, finetuning, inference
- Explainability and multimodality

User workflow

User workflow: User query



The user enters a free-form target identification question (the hypothesis)

User workflow: Multimodal evidence

Biological Reasoner

TARGET X decreases inflammation in ALS

"Here we show that TARGET X inhibition delayed disease onset in a SOD1-G93A mouse model of ALS" 1

"Bone-marrow specific knockout of TARGET X led to reduced clonal expansion and immune response in CD4+ T-cells" 2

"TARGET X is highly expressed in ALS vs control" DiffExpr

"TARGET X is genetically associated with SLE" L2G
+
"SLE is closely associated with ALS" KG

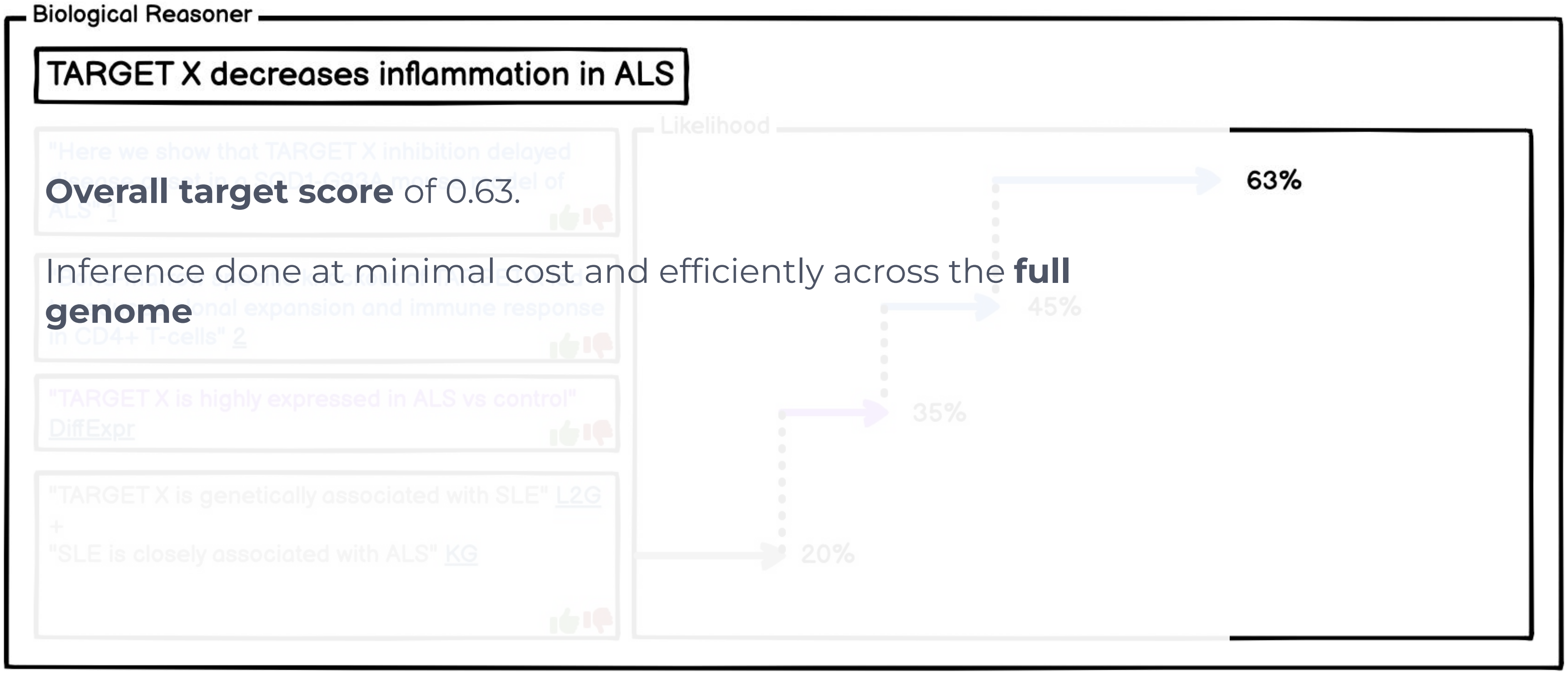
Likelihood

Evidence is retrieved related to the query across all available modalities, including **literature**,

textualized **patient-level data**,

And textualized **co-operative combinations of evidence** from a knowledge graph

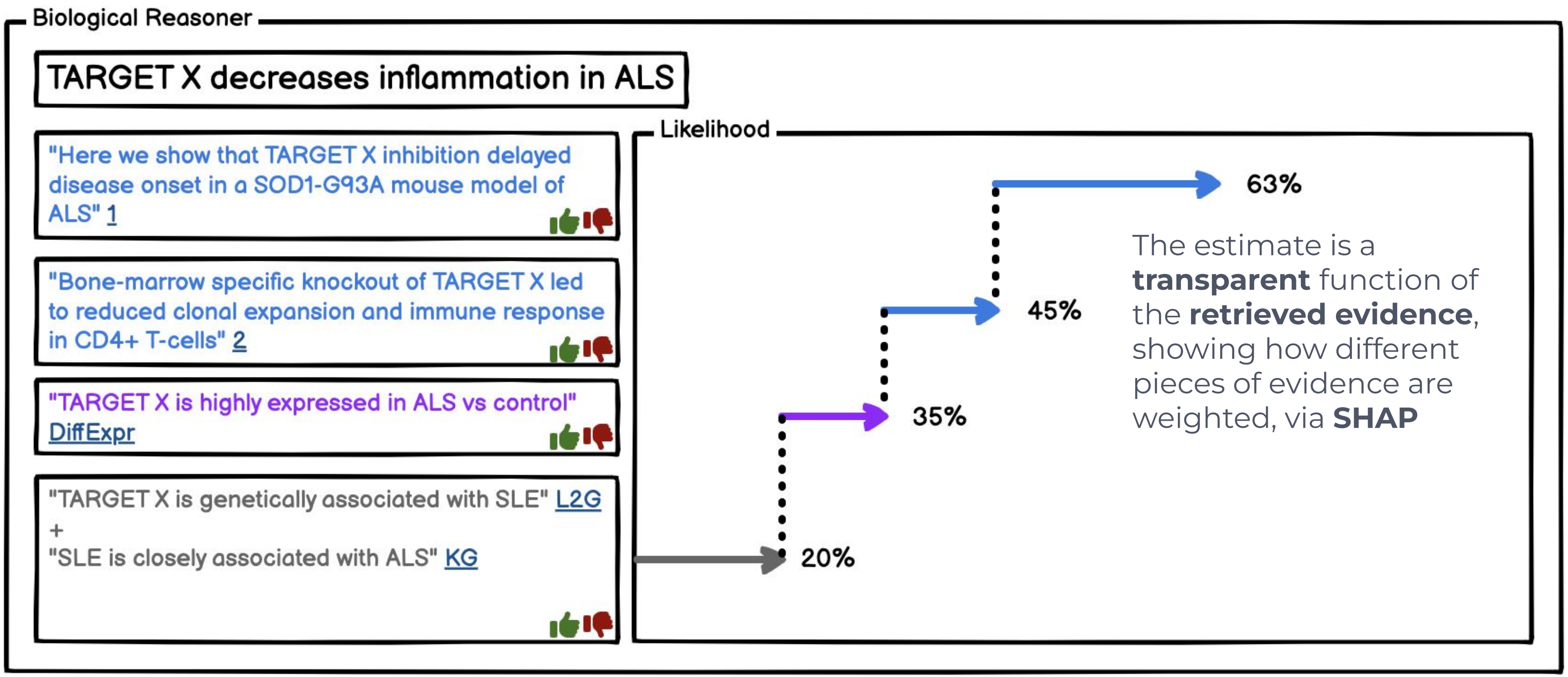
User workflow: Overall prediction, full-genome



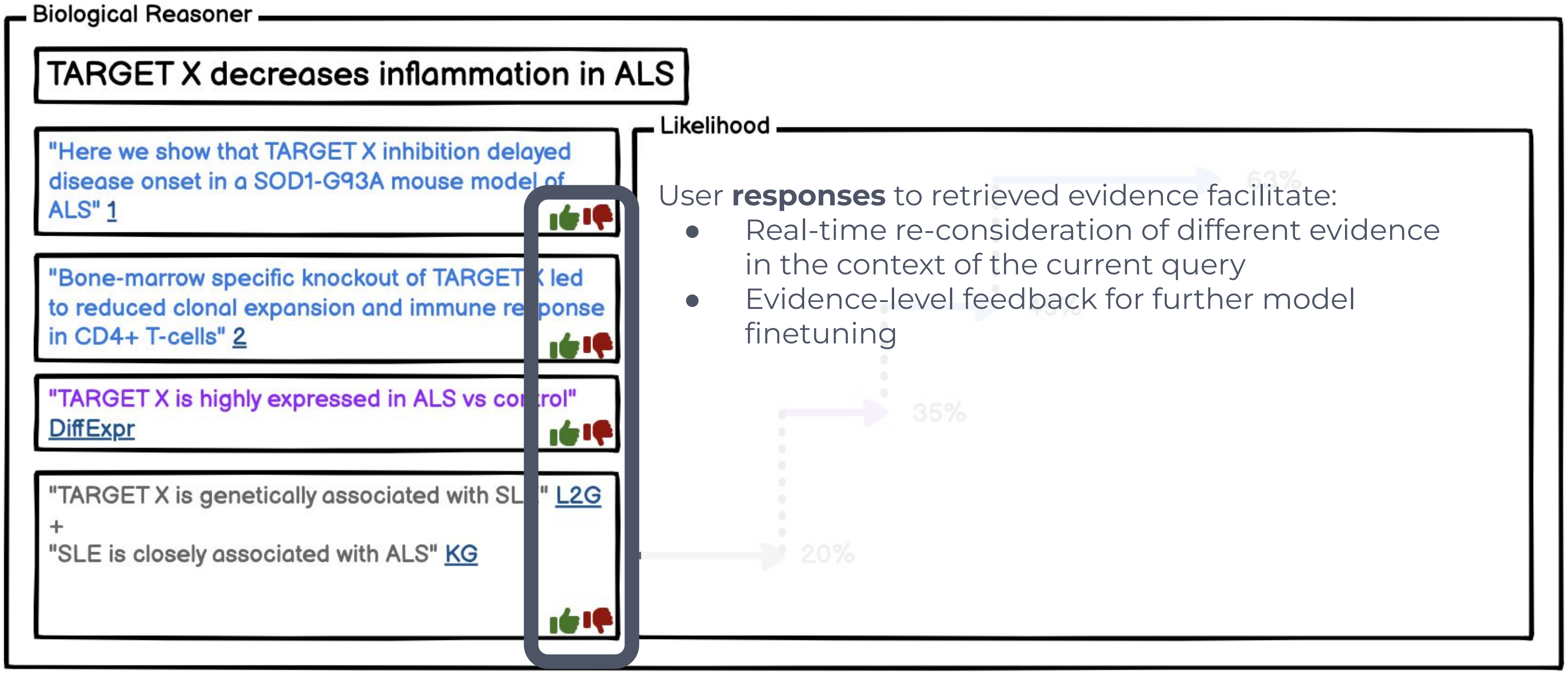
Overall target score of 0.63.

Inference done at minimal cost and efficiently across the **full genome**

User workflow: Explanation via SHAP



User workflow: User feedback



User **responses** to retrieved evidence facilitate:

- Real-time re-consideration of different evidence in the context of the current query
- Evidence-level feedback for further model finetuning

User workflow: Example

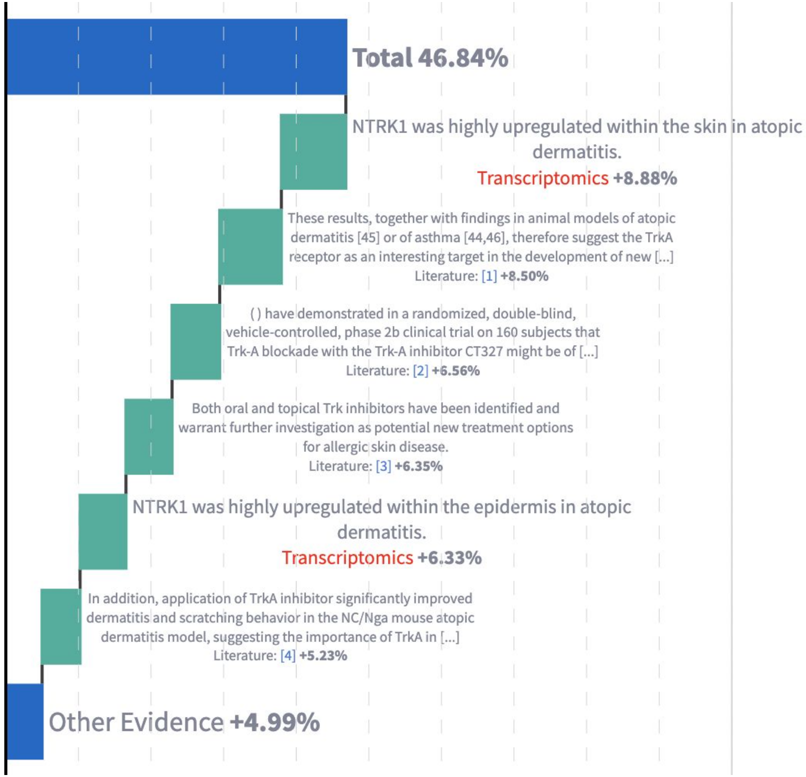
Query:

[x] is a novel promising therapeutic target for atopic dermatitis.

Evidence Sources

Literature & Differential Expression

Target: NTRK1

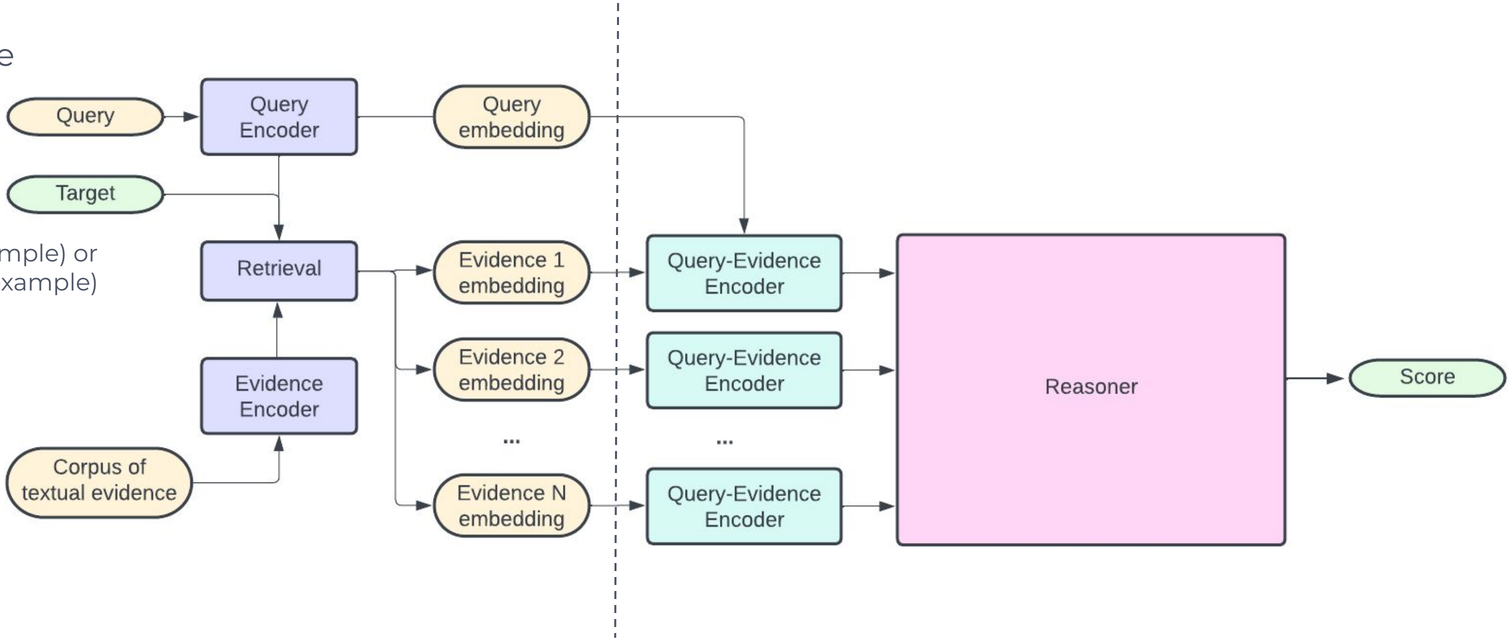


Model setup: Training

Model setup: Training

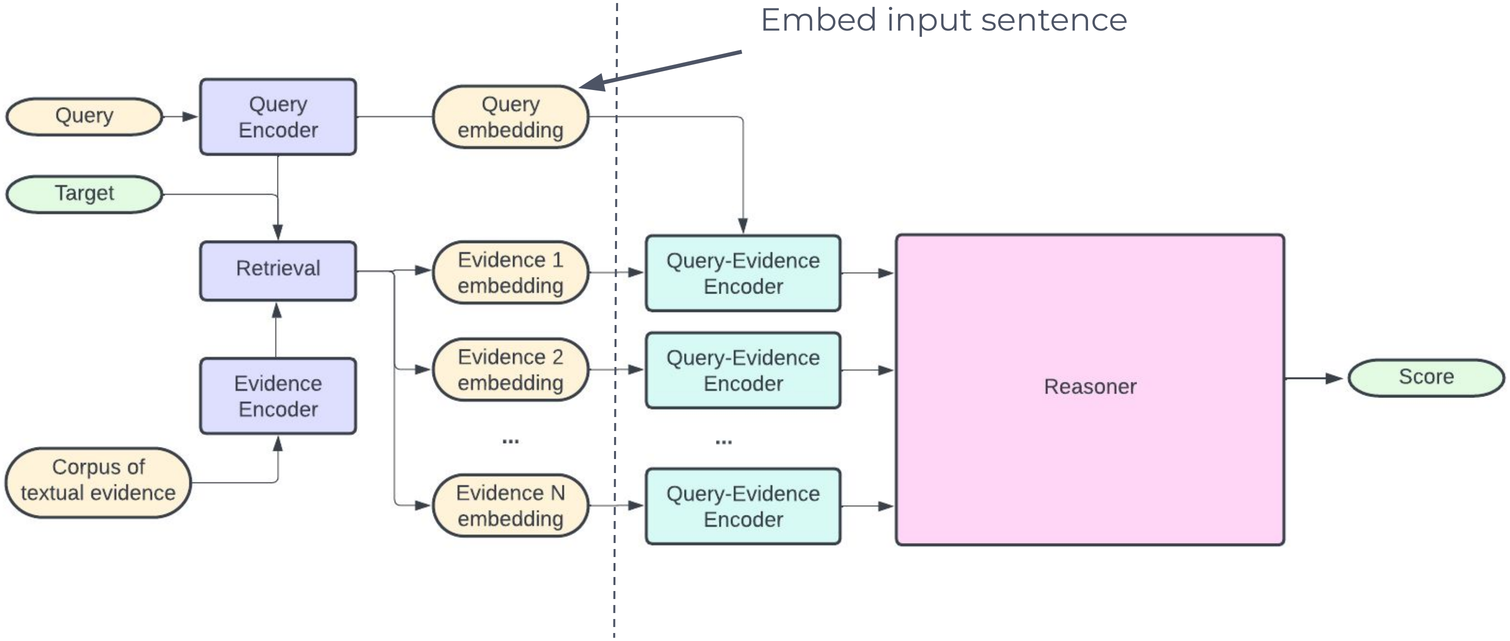
Input sentence

Correct (positive example) or incorrect (negative example) target



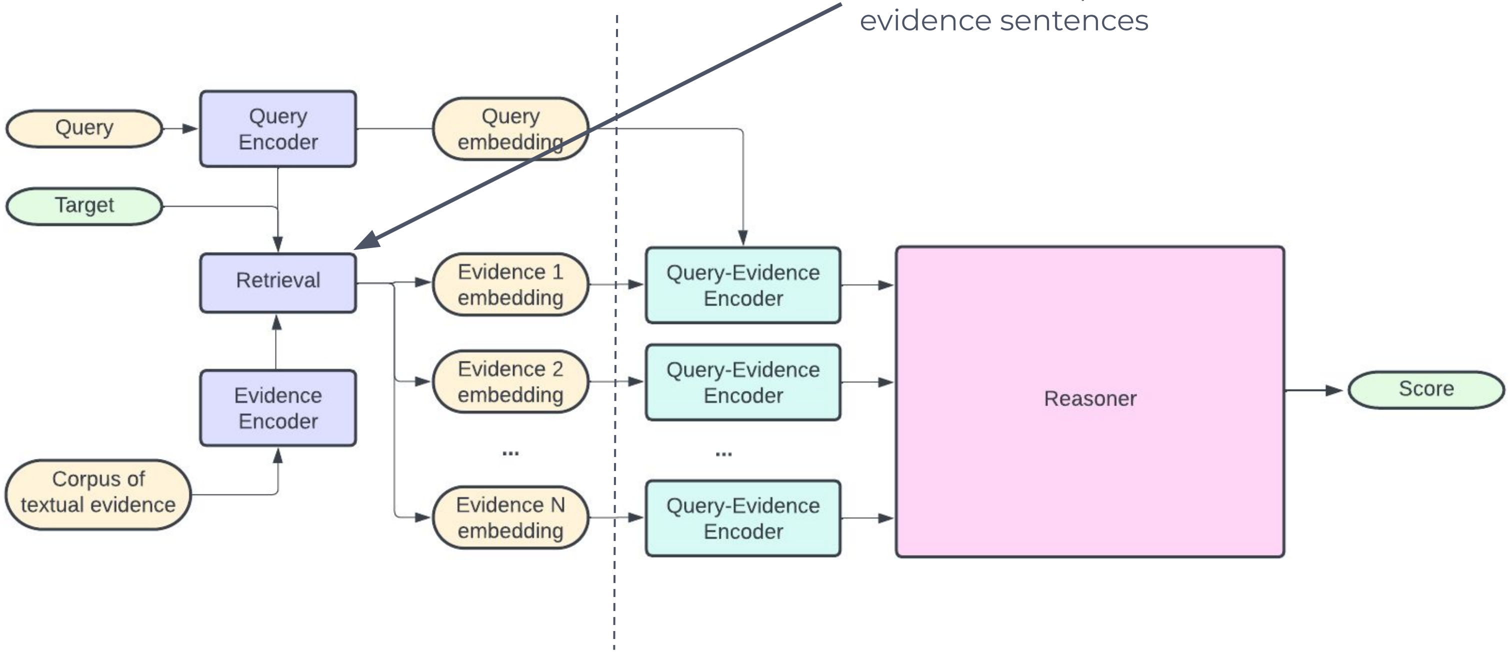
Model setup: Training

Embed input sentence



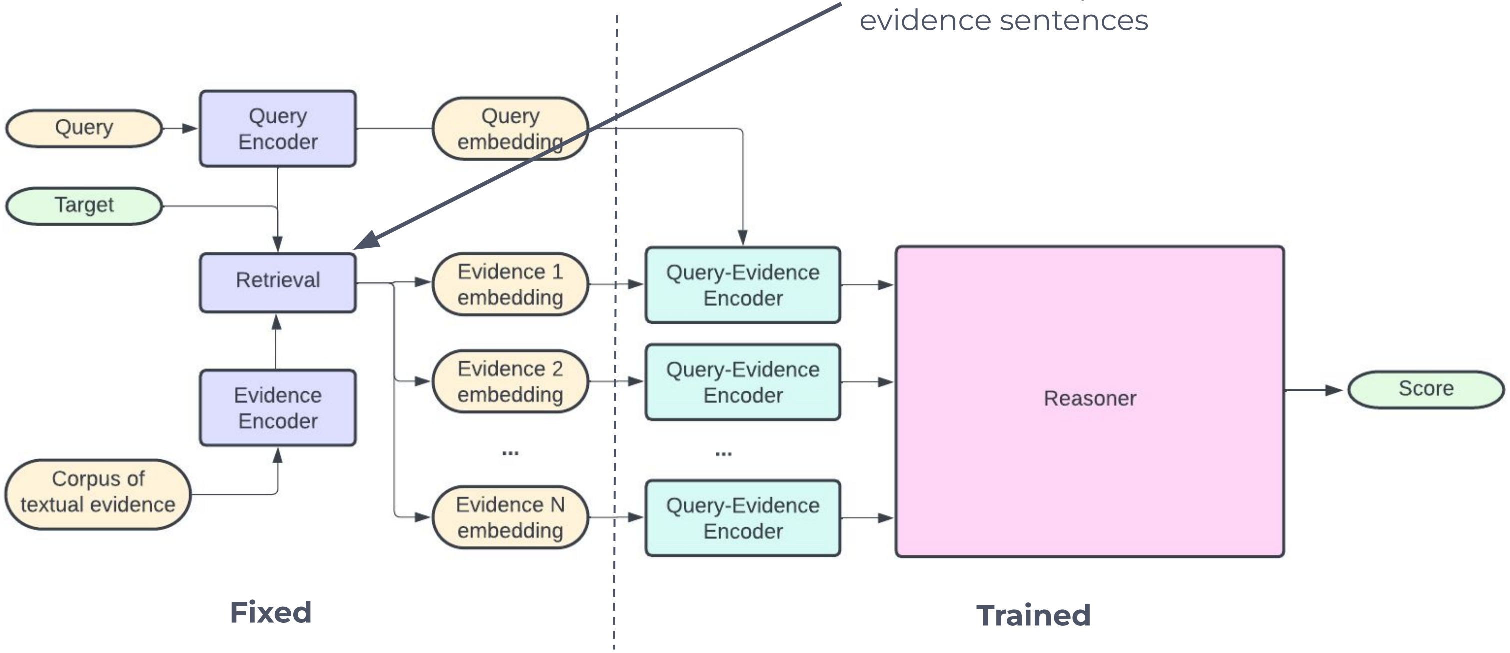
Model setup: Training

Retrieve the Top N most similar evidence sentences



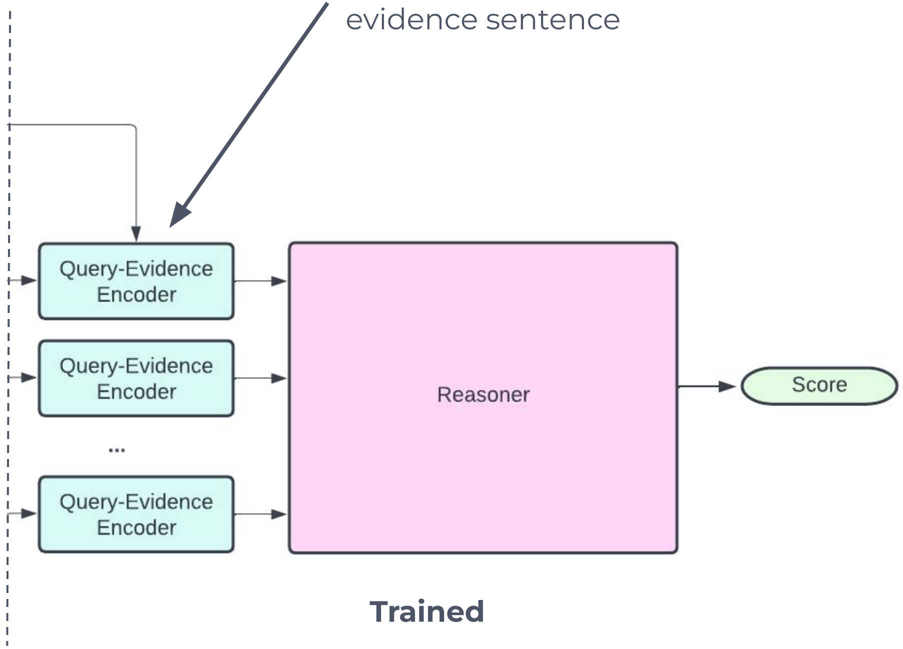
Model setup: Training

Retrieve the Top N most similar evidence sentences



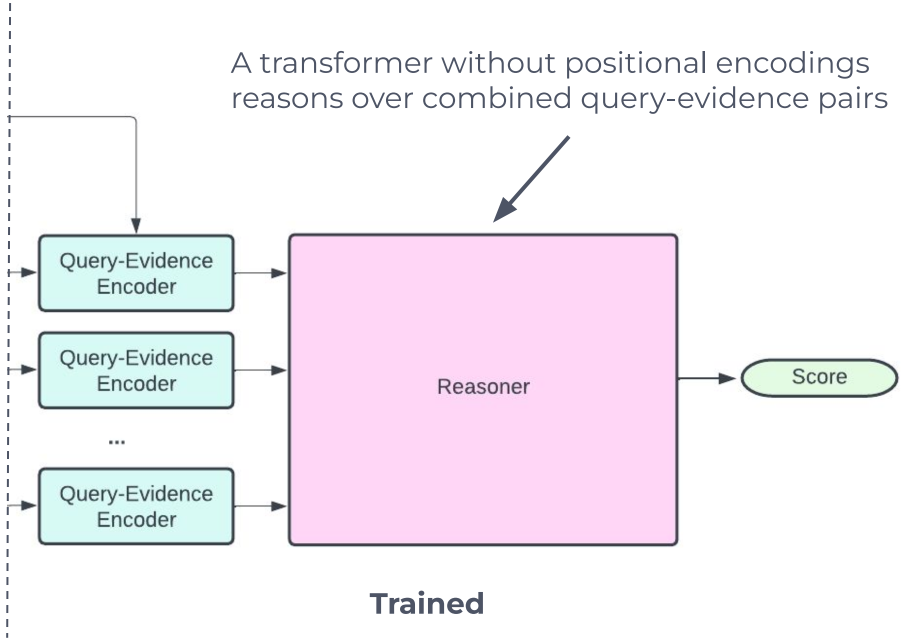
Model setup: Training

Combine the query sentence with each evidence sentence



Trained

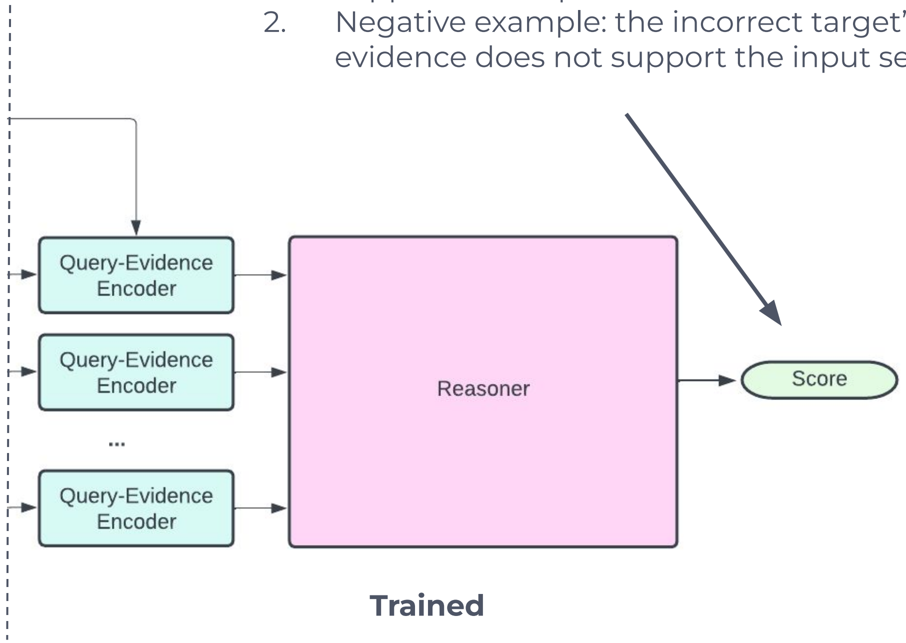
Model setup: Training



Model setup: Training

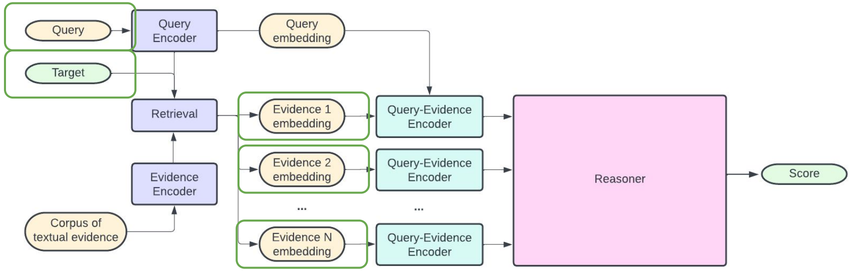
Objective is to predict that:

1. Positive example: the correct target's evidence supports the input sentence
2. Negative example: the incorrect target's evidence does not support the input sentence



Model setup: Training

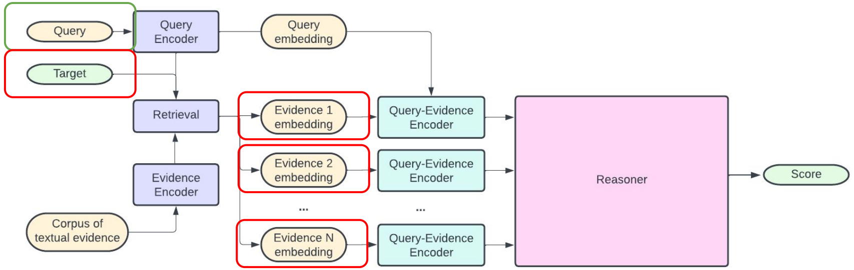
Positive example



Retrieved evidence corresponds to correct target for the input sentence.

During training the query is a sentence with a known target, such as an entity-linked sentence from the literature.

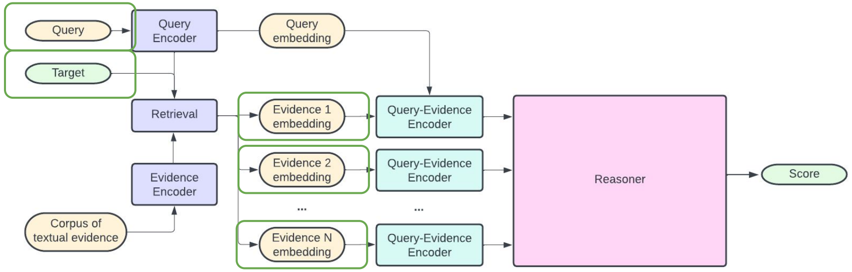
Negative example



Retrieved evidence corresponds to an incorrect target for the input sentence.

Model setup: Training

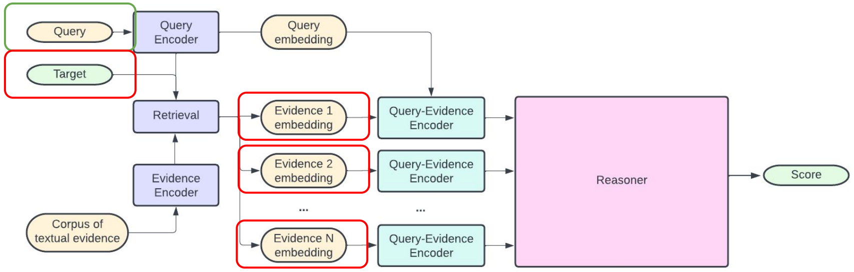
Positive example



Retrieved evidence corresponds to correct target for the input sentence.

During training the query is a sentence with a known target, such as an entity-linked sentence from the literature.

Negative example



Retrieved evidence corresponds to an incorrect target for the input sentence.

To **limit literature bias**, the negative evidence **target** is randomly **sampled according** its **frequency of mention in the literature**, so that so well-studied targets come up equally often in negative and positive examples.

Model setup: Finetuning

Model setup: Finetuning

Constitutional evidence set finetuning with an LLM

Input query

Target A

Sample a target identification query

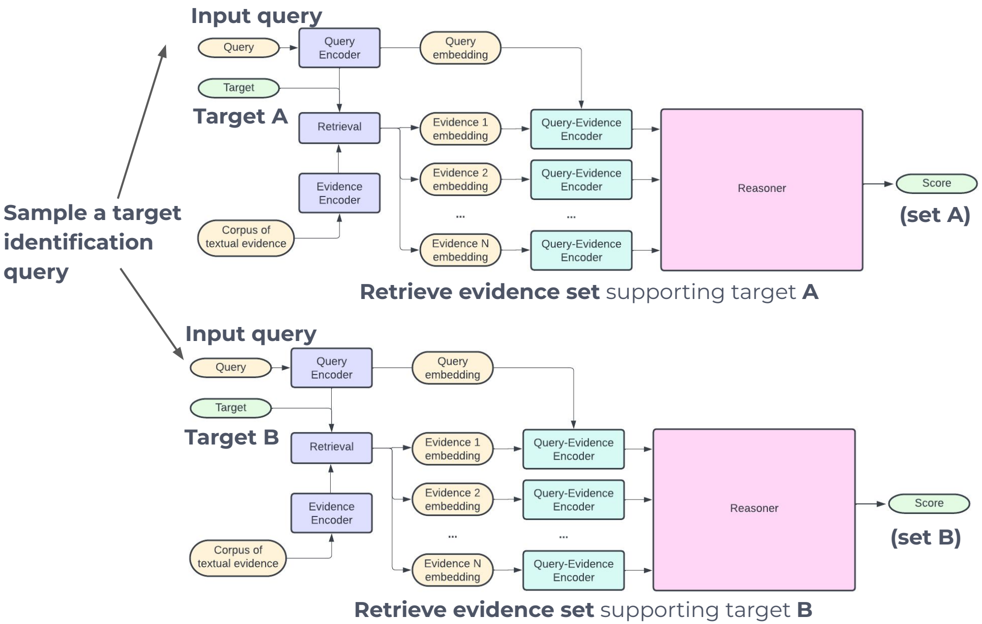
Input query

Target B

Yuntao Bai,et al. (2022). Constitutional AI: Harmlessness from AI Feedback.

Model setup: Finetuning

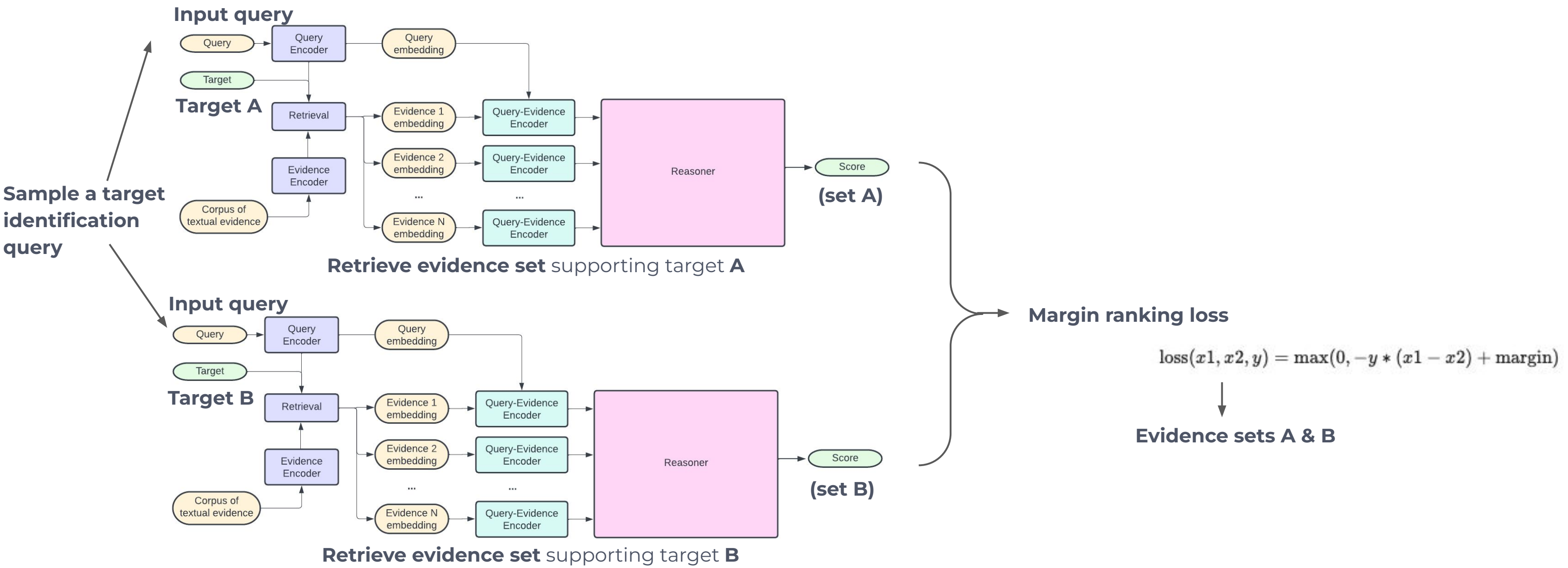
Constitutional evidence set finetuning with an LLM



Yuntao Bai,et al. (2022). Constitutional AI: Harmlessness from AI Feedback.

Model setup: Finetuning

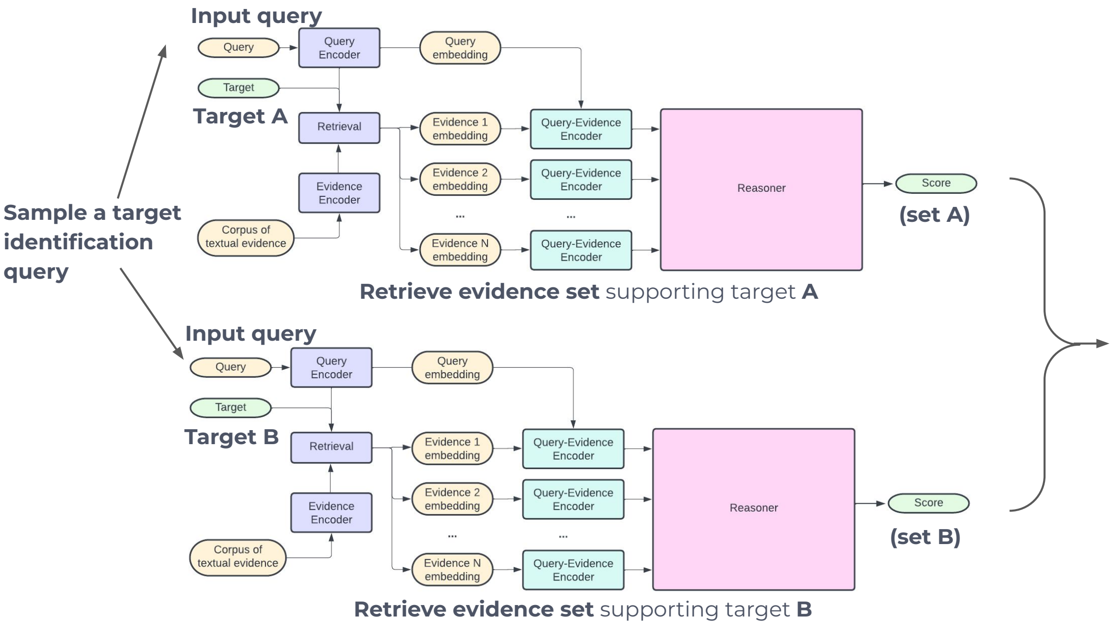
Constitutional evidence set finetuning with an LLM



Yuntao Bai,et al. (2022). Constitutional AI: Harmlessness from AI Feedback.

Model setup: Finetuning

Constitutional evidence set finetuning with an LLM



label is determined by few-shot (examples provided by DD) prompted **LLM**, based on a **constitution of principles** that specify what makes a good evidence set to support a target as an answer to the input query

y (1 or -1) is the **label** specifying which evidence set is better

Margin ranking loss

$$\mathrm{loss}(x1, x2, y) = \max(0, -y * (x1 - x2) + \mathrm{margin})$$

Evidence sets A & B

Yuntao Bai,et al. (2022). Constitutional AI: Harmlessness from AI Feedback.

Model setup: Finetuning

A constitution of principles for few-shot LLM ranking of two sets of evidence

The better evidence set is the one which offers most support to set of principles

Principles in the constitution relate to:

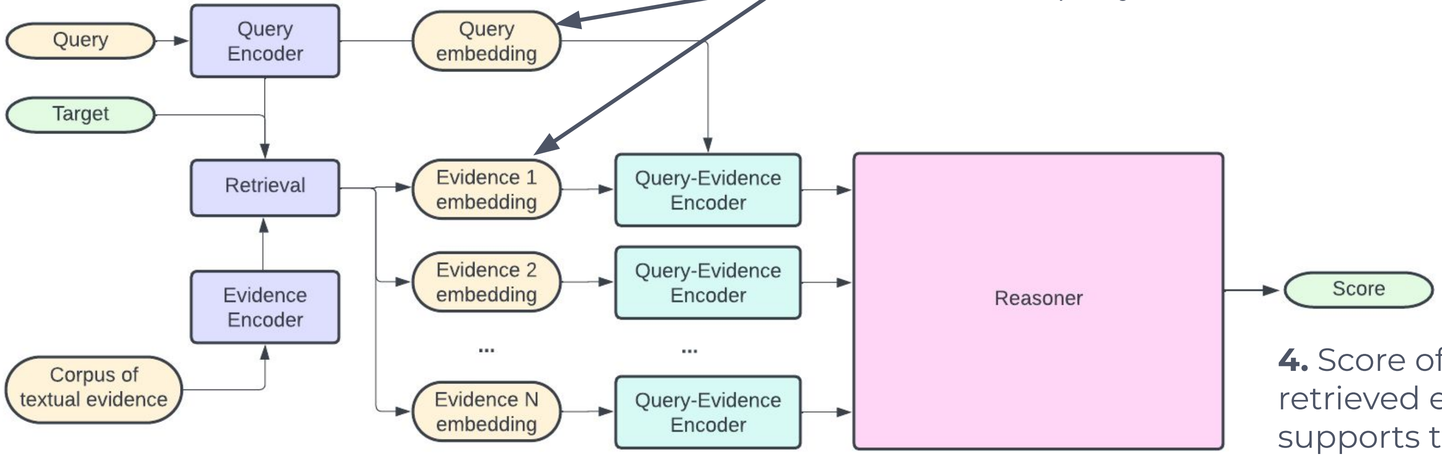
- Disease relevance
- Mechanism relevance
- Tissue and cell type expression

Model setup: Inference

Model setup: Inference

1. User query

2. Embed user query and retrieve evidence



4. Score of whether retrieved evidence supports the target as an answer to the user query

Knowledge (1. And 2.)

(retrieved evidence to help answer user query)

3. Reasoning

(using retrieved evidence to answer user query)

5. SHAP

Explainability & Multimodality

Direct explainability with SHAP

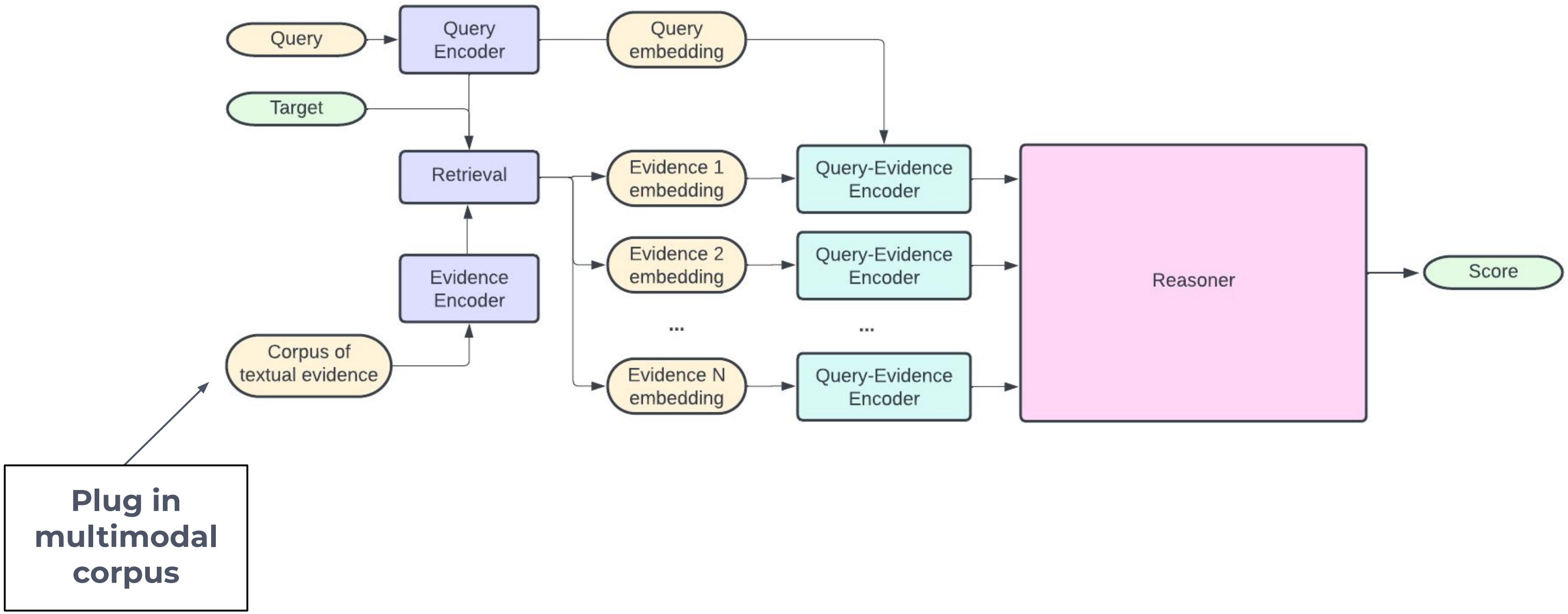
SHapley Additive exPlanations (SHAP)

- Approximate output of a model as a linear combination of binary input features
- The model output is an additive sum of Shapley values (φ) attributed to the input features, plus a bias (φ0):

$$g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z'_i,$$

  - (z') is binary, the feature is present [1], or absent [0].'
  - φi is the average difference between the score with and without feature i, considering all combinations of other features
  - g(*) is the output

Multimodality



Model setup: Multimodality via textualisation

Example modality: Differential expression

High quality case-control **differential expression** datasets → Groupby datasets **disease** / **tissue** / **cell type** context, and take largest N → Create highly / moderately / lowly of upregulated / downregulated, **bins** using log2fc cutoffs

Model setup: Multimodality via textualisation

Example modality: Differential expression

High quality case-control **differential expression** datasets → Groupby datasets **disease** / **tissue** / **cell type** context, and take largest N → Create highly / moderately / lowly of upregulated / downregulated, **bins** using log2fc cutoffs → Use **template** to convert to sentence

↓

Target X was {bin} in {cell type}s within the {tissue} in {disease}.

↓

Target X was highly upregulated in keratinocytes within the skin in atopic dermatitis.

Model setup: Multimodality via textualisation

Example modality: Differential expression

High quality case-control **differential expression** datasets → Groupby datasets **disease** / **tissue** / **cell type** context, and take largest N → Create highly / moderately / lowly of upregulated / downregulated, **bins** using log2fc cutoffs → Use **template** to convert to sentence

↓

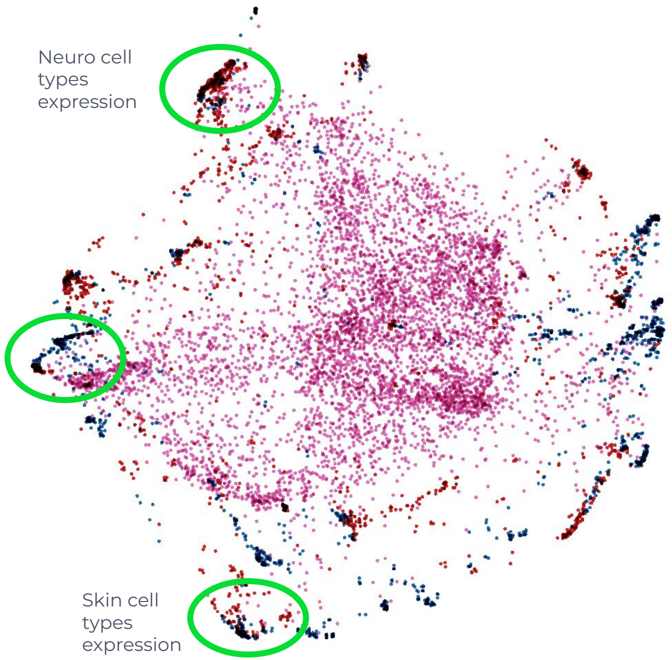
Target X was {bin} in {cell type}s within the {tissue} in {disease}.

↓

Target X was highly upregulated in keratinocytes within the skin in atopic dermatitis. → Embed sentence with literature-trained encoder and add to **vector search index**

Model setup: Multimodality via textualisation

t-SNE

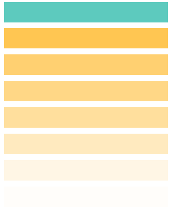


Literature trained encoder → overlap of literature and textualised (omics) sentences covering similar aspects of biology

Literature
Differential expression
Expression

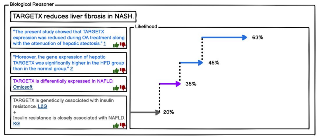
Target prioritisation wish list

Target prioritisation model wishlist



Genome-wide predictions (relevant and novel) ✅

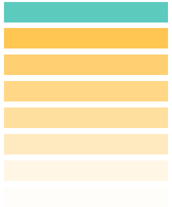
Reason from multiple data modalities ✅



Explainable predictions ✅

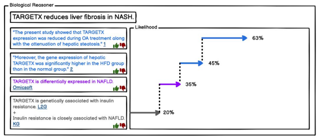
Large language models:
What role do they have?

Target prioritisation model wishlist: LLMs



Genome-wide predictions (relevant and novel) ✅/ ❌

Reason from multiple data modalities ✅/ ❌



Explainable predictions ✅/ ❌

The flexibility / performance tradeoff

Flexibility

Easy to use,
high-performance

Teach users how to use a model or data source optimally.

Develop guardrails around a limited number of use cases to ensure ease and high performance.

The flexibility / performance tradeoff

Large language models can help to mitigate this tradeoff as both a **translator** and **orchestrator**

Flexibility

Teach users how to use a model or data source optimally.

Easy to use, high-performance

Develop guardrails around a limited number of use cases to ensure ease and high performance.

The flexibility / performance tradeoff

Large language models can help to mitigate this tradeoff as both a **translator** and **orchestrator**

Adapt user needs into the right language for underlying models and databases.

Navigating complex workflows involving a series of complex analysis steps leveraging a large range of tools and data sources

Flexibility

Easy to use, high-performance

Teach users how to use a model or data source optimally.

Develop guardrails around a limited number of use cases to ensure ease and high performance.

LLM as translator

Can you find me novel targets to modulate neuroinflammation in ALS?

An LLM is initialized with a set of instructions about how to format queries to an underlying specialist model or data source.

It knows the relevant best practices and can flexibly apply them.

LLM as translator

Can you find me novel targets to modulate neuroinflammation in ALS?

Requirement (in LLM context):
Phrase the query in the form of a statement, using an "[x]" to denote the desired target.

LLM as translator

Can you find me novel targets to modulate neuroinflammation in ALS?

Example Best Practice (in LLM context):
When considering novelty, use two separate queries for novelty and biological function, aggregating results post-hoc.

LLM as translator

Can you find me novel targets to modulate neuroinflammation in ALS?

Example Best Practice (in LLM context):
When considering novelty, use the phrase "well-known" and invert the score.

LLM as translator

Can you find me novel targets to modulate neuroinflammation in ALS?

Query:
"[x] modulates neuroinflammation in ALS" AND NOT "[x] is well-known in ALS".

LLM as translator

Can you find me novel targets to modulate neuroinflammation in ALS?

***Query**:*
"[x] modulates neuroinflammation in ALS" AND NOT "[x] is well-known in ALS".

Details about best practices can be **abstracted away** for non-technical users (but available for experts to tweak).

Certainly. My top choice is TARGETX. It has been linked to neuroinflammation by ...

LLM as translator

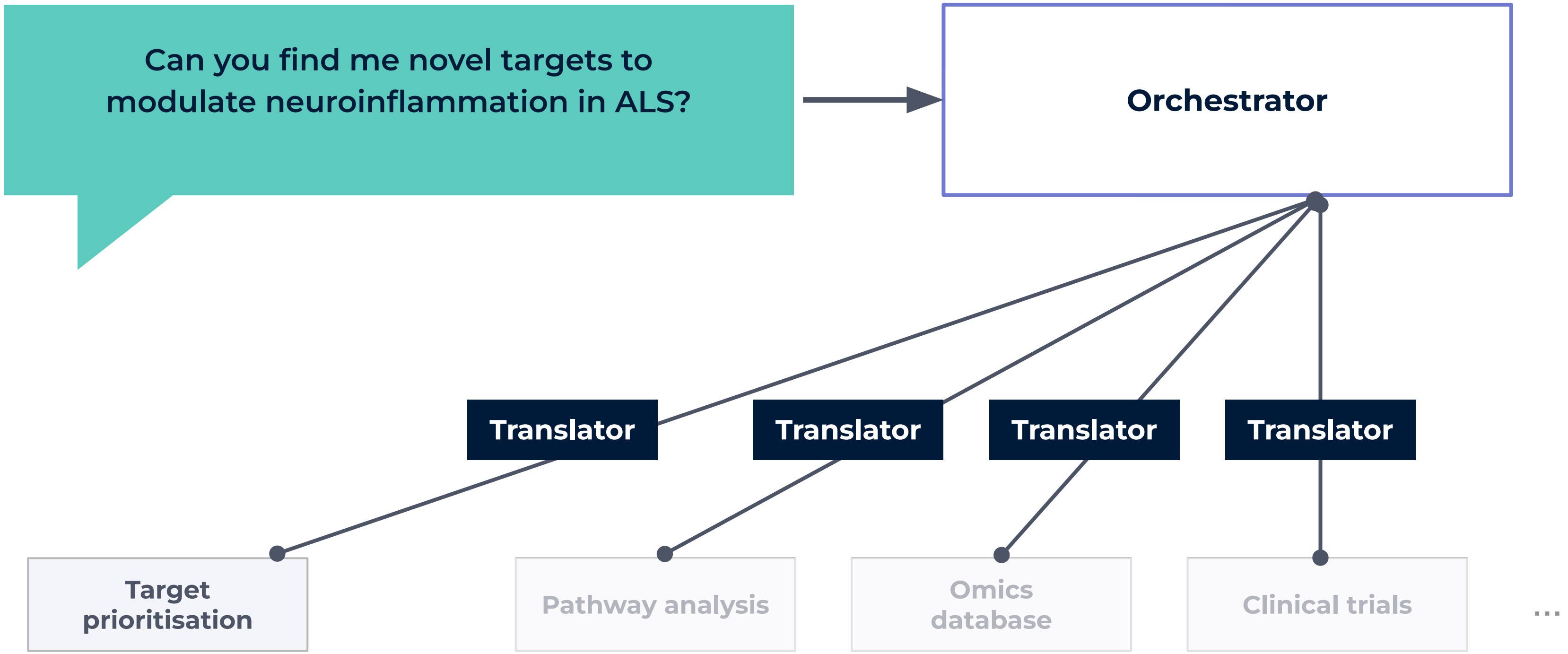
Can you find me novel targets to modulate neuroinflammation in ALS?

Query:
"[x] modulates neuroinflammation in ALS"
AND NOT "[x] is well-known in ALS".

Users can **flexibly** interact with the underlying model, while applying best practices for **high performance**.

Certainly. My top choice is TARGETX. It has been linked to neuroinflammation by …

LLM as orchestrator



LLM as orchestrator

Can you find me novel targets to modulate neuroinflammation in ALS?

Certainly. My top choice is TARGETX. It has been linked to neuroinflammation by …

LLM as orchestrator

Can you find me novel targets to modulate neuroinflammation in ALS?

Certainly. My top choice is TARGETX. It has been linked to neuroinflammation by ...

Is there anything interesting upstream of TARGETX?

TARGETY upregulates TARGETX. TARGETY ranked 834th overall; there doesn't appear to be any strong evidence linking it to neuroinflammation...

LLM as orchestrator

Can you find me novel targets to modulate neuroinflammation in ALS?

Certainly. My top choice is TARGETX. It has been linked to neuroinflammation by ...

Is there anything interesting upstream of TARGETX?

TARGETY upregulates TARGETX. TARGETY ranked 834th overall; there doesn't appear to be any strong evidence linking it to neuroinflammation...

Can you filter the Top 200 results to those without any clinical results at or beyond Phase II?

Yes. There are 128 targets remaining. The top remaining target is...

LLM as orchestrator

Benevolent[AI] Augmented Intelligence

You:

Predict targets that modulate insulin resistance in adipose tissue in type 2 diabetes

Ben:

Here are the top 20 targets predicted for Predict targets involved with insulin resistance in adipose tissue in type 2 diabetes. Data Source:

Evidence synthesis and attribution model

Target UUID	Score
8cd65a04-cbea-11e5-a271-005056be9332	87.2
a8633b9c-cbeb-11e5-a271-005056be9332	86.8
aa7d9e16-cbed-11e5-a271-005056be9332	86.7
d1eb66ba-cbeb-11e5-a271-005056be9332	86.1
37df44c0-cbef-11e5-a271-005056be9332	85.7
c04350bc-cbeb-11e5-a271-005056be9332	83.7
ce7fef78-cbeb-11e5-a271-005056be9332	82.8
3f699cf2-cbec-11e5-a271-005056be9332	81.9
80c23cc2-cbec-11e5-a271-005056be9332	81.7
9af8f222-cbea-11e5-a271-005056be9332	81.5

Items per page: 10 1-10 of 25

Filter by: Target Target UUID Score

Ask a question

Summary

Summary:
Language models for target prioritisation

Genome-wide

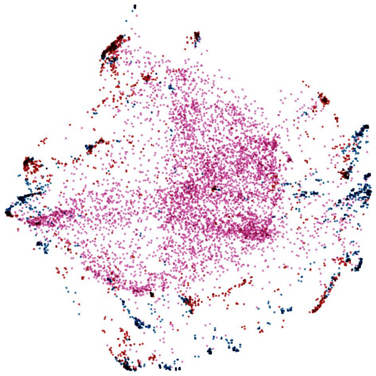
Multimodal

Explainable

- **Drug discovery and BenevolentAI**
- **Drug discovery biologists:** In depth, multimodal reasoning
- **Masked-language models:** Genome-wide, literature-based target prioritisation
- **Evidence synthesis with attribution:** Explainable, genome-wide, multimodal target prioritisation
- **Large-language models:** What role do they have?

Thanks to the team

Dane Corneil - AI Scientist
Angus Brayne - AI Scientist
Georgiana Neculae - AI Scientist
Daniel Jaroslawicz - AI Scientist
Paride Antinucci - Drug Discovery Scientist
Eryk Kropiwnicki - Bioinformatics Data Scientist
Maciek Wiatrak - ML Engineer
Vinay Subbiah - Product Designer
Many others



Any questions?

Email: ravi.patel@benevolent.ai

benevolent.com   @benevolent_ai   benevolentai   hello@benevolent.ai