

# Accelerating Cancer Research with TruLens and Gemini

biologyAI team

# Accelerating Cancer Research with AI

Problem: Lots of scientific research

PubMed : Millions of scientific articles

How to find relevant information more quickly?

Entrez Edirect : Download hundreds to thousands of abstracts

Gemini : Ask scientific questions about the abstracts

TruLens : Identify relevant articles for Groundedness and Relevance

Example : Extract all proteins and potential cancer antigens

# Target Audience

Bioinformaticians

New kind of scientist that uses large databases to study biology

Many kinds of data, lots of DNA and protein sequence data

People that use Python every day

We want to give them new powers and capabilities

Use Gemini and TruLens to help speed their research

Convenient Chatbot interface to science

# Bispecific T-cell Engagers - BiTEs

Tebentafusp a novel first in class cancer treatment

Could be the platform to cure a wide variety of cancers

Uses a novel mechanism, a bispecific fusion protein

It is an engineered synthetic protein

One side binds to the HLA receptor with gp100 bound

The other binds to T-cells CD3 complex

It brings the T-cells in contact with the cancer cells

And the T-cells kill the cancer cell

# Method

- 1) Find abstracts with Entrez
- 2) For each abstract
  - a) Extract all proteins and antigens with Gemini
  - b) Evaluate against all abstracts with TruLens with gpt-3.5 embeddings

ResearchBot:

Hello! I'm a helpful science robot for bioinformaticians that knows about a new kind of cancer treatment that uses a bispecific fusion protein.

What kind of cancer are you trying to cure?

ResearchBot:

Hello! I'm a helpful science robot for bioinformaticians that knows about a new kind of cancer treatment that uses a bispecific fusion protein.

What kind of cancer are you trying to cure?

Uveal Melanoma

What kind of cancer are you trying to cure?

You:

Uveal Melanoma

ResearchBot:

Uveal melanoma is a rare type of cancer that develops in the uvea, which is the middle layer of the eye. It is the most common type of eye cancer in adults. Uveal melanoma typically begins in the cells that produce melanin, the pigment that gives color to the eye.

The first thing we should do is to look for antigens that might be present in Uveal Melanoma. Let's do a search with Entrez for that. I'll run the following command:

```
for PMID in $(esearch -db pubmed -query "uveal melanoma AND 1980:2000[Date - Pub  
efetch -db pubmed -id $PMID -format abstract > "$PMID.txt"  
done
```



begins in the cells that produce melanin, the pigment that gives color to the eye.

The first thing we should do is to look for antigens that might be present in Uveal Melanoma. Let's do a search with Entrez for that. I'll run the following command:

```
for PMID in $(esearch -db pubmed -query "uveal melanoma AND 1980:2000[Date - Pub  
efetch -db pubmed -id $PMID -format abstract > "$PMID.txt"  
done
```

ResearchBot:

I've downloaded the abstracts. Now I will go through each of them and convert them into a language more tailored for bioinformaticians.

Running...



```
for PMID in $(esearch -db pubmed -query "uveal melanoma AND 1980:2000[Date - Pub  
efetch -db pubmed -id $PMID -format abstract > "$PMID.txt"  
done
```

ResearchBot:

I've downloaded the abstracts. Now I will go through each of them and convert them into a language more tailored for bioinformaticians using Gemini.

ResearchBot:

Now I will use a different LLM, ChatGPT-3.5 to check each of the summaries to see if they have any potential relevance for having a cancer antigen.

Running...

I've downloaded the abstracts. Now I will go through each of them and convert them into a language more tailored for bioinformaticians using Gemini.

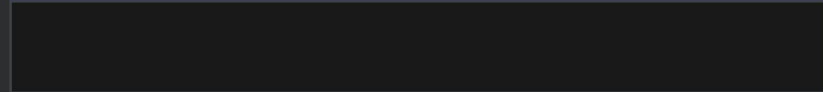
ResearchBot:

Now I will use a different LLM, ChatGPT-3.5 to check each of the summaries to see if they have any potential relevance for having a cancer antigen.

ResearchBot:

I've prepared a TruLens dashboard with my results. [TruLens Dashboard](#)

It appears that there are a number of relevant papers, including one titled "Adoptive T cell immunotherapy of human uveal melanoma targeting gp100."



## 11120562.txt

Records	Average Latency (Seconds)	Total Cost (USD)	Total Tokens	Context Relevance	Groundedness	Answer Relevance
1	2	\$0	1.06k	1.0 ✔ high	0.25 ● low	1.0 ✔ high

Select App

## 11120866.txt

Records	Average Latency (Seconds)	Total Cost (USD)	Total Tokens	Context Relevance	Groundedness	Answer Relevance
3	3	\$0	3.02k	1.0 ✔ high	1.0 ✔ high	1.0 ✔ high

Select App

## 11151214.txt

Records	Average Latency (Seconds)	Total Cost (USD)	Total Tokens	Context Relevance	Groundedness	Answer Relevance
1	2	\$0	1.08k	0.9 ✔ high	0.25 ● low	1.0 ✔ high

Select App

11107402.txt



Filter Applications

11120866.txt ×



# 11120866.txt

Records Feedback Functions

App ID	Response	Groundedness	Answer Relevance	Cor
11120866.txt	"The protein or antigen that ...	1	1	
11120866.txt	"The protein or antigen that ...	1	1	
11120866.txt	"The protein or antigen that ...	1	1	

1 to 3 of 3 | Page 1 of 1

Hint: select a row to display details of a record

1. J Immunol. 2000 Dec 15;165(12):7308-15. doi: 10.4049/jimmunol.165.12.7308.

Adoptive T cell immunotherapy of human uveal melanoma targeting gp100.

Sutmuller RP(1), Schurmans LR, van Duivenvoorde LM, Tine JA, van Der Voort EI, Toes RE, Melief CJ, Jager MJ, Offringa R.

Author information:

(1)Department of Immunohematology and Bloodbank, Leiden University Medical Center, The Netherlands.

HLA-A\*0201-restricted CTL against human gp100 were isolated from HLA-A\*0201/K(b) (A2/K(b))-transgenic mice immunized with recombinant canarypox virus (ALVAC-gp100). These CTL strongly responded to the gp100(154-162) epitope, in the context of both the chimeric A2/K(b) and the wild-type HLA-A\*0201- molecule, and efficiently lysed human HLA-A\*0201(+), gp100(+) melanoma cells in vitro. The capacity of the CTL to eradicate these tumors in vivo was analyzed in A2/K(b)-transgenic transgenic mice that had received a tumorigenic dose of human uveal melanoma cells in the anterior chamber of the eye. This immune-privileged site offered the unique opportunity to graft xenogeneic tumors into immunocompetent A2/K(b)-transgenic mice, a host in which they otherwise would not grow. Importantly, systemic (i.v.) administration of the A2/K(b)-transgenic

-UU-:--- F1 11120866.txt Top L1 Git-main (Org Wrap) 11:10AM 1.41 -----

# Strengths

**High Relevance and Impact:** The system addresses a critical global health issue, cancer, making it highly relevant and impactful.

**Advanced Knowledge Integration:** By integrating thousands of research papers, it ensures a comprehensive understanding of current cancer research.

**Data-Driven Insights:** The system could provide valuable insights by analysing large volumes of data, potentially leading to new discoveries.

**Time Efficiency:** It can significantly speed up the research process by providing quick access to a vast amount of information.

**Interdisciplinary Collaboration:** It encourages collaboration among experts from different fields, fostering a multidisciplinary approach to cancer research.

# Weaknesses

**Complexity in Development:** Developing such a system requires sophisticated algorithms and understanding of both computer science and oncology.

**Data Quality and Bias:** The system's effectiveness depends on the quality and diversity of the input papers. Bias in research could lead to skewed results.

**High Resource Requirement:** It requires substantial computational power and storage for processing and maintaining a large database.

**Maintenance and Updating:** The system needs continuous updates as new research is published, which can be resource-intensive.

**User Learning Curve:** There might be a learning curve for researchers to effectively use the system.



# Opportunities

**Technological Advancements:** Leveraging AI, machine learning, and natural language processing can enhance the system's capabilities.

**Funding and Support:** There is potential for funding from healthcare institutions, governments, and NGOs focused on cancer research.

**Collaborations with Universities and Research Centers:** Partnering with academic institutions can provide access to the latest research and expert feedback.

**Global Reach and Impact:** The system can be used worldwide, contributing to global cancer research efforts.

**Publication and Recognition:** Successful implementation could lead to publications and recognition in both the tech and medical research communities.

# Threats

**Competitive Solutions:** There might be existing or upcoming tools with similar objectives, posing competition.

**Technological Challenges:** Rapid changes in technology can make parts of the system obsolete quickly.

**Data Security and Privacy:** Handling sensitive research data requires robust security measures to prevent breaches.

**Dependence on Research Quality:** The system's effectiveness is directly tied to the quality and integrity of the research papers fed into it.

**Regulatory and Ethical Considerations:** The system must adhere to ethical guidelines and regulations in handling medical research data.

# Market Analysis: AI in Oncology

## TAM: Global Oncology Drugs Market

- 2022 Value: USD 184.95 Billion
- Projected 2030 Value: USD 484.32 Billion
- Includes all cancer R&D activities globally.

## SAM: AI in Oncology

- 2023 Value: USD 1.20 Billion
- Expected Growth: 33.5% CAGR (2024-2030)
- Focused on AI for cancer research, diagnostics, and treatment.

# Generative LLM model for cancer research

- Comprehensive Literature Review**: Utilize LLM to analyze and summarize existing research papers, extracting key information on cancer antigens.
- Identifying Connections**: LLM models can detect hidden correlations between different studies, revealing potential antigen targets.
- Update Alerts**: Set up LLM to provide updates on new research findings in specific cancer types and antigen discovery.
- Data Synthesis**: LLM integrates diverse data sets (genomic, clinical trials, etc.) for comprehensive antigen profiles.
- Hypothesis Generation**: LLM suggests new hypotheses based on existing literature for exploring undiscovered antigens.