

# Large Language Models for Biomedicine: from PubMed Search to Gene Set Analysis

Zhiyong Lu, PhD FACMI FIAHSI  
Senior Investigator, NCBI, NLM, NIH  
Professor of Computer Science (Adjunct), UIUC

ISMB 2024 Tutorial  
July 9, 2024

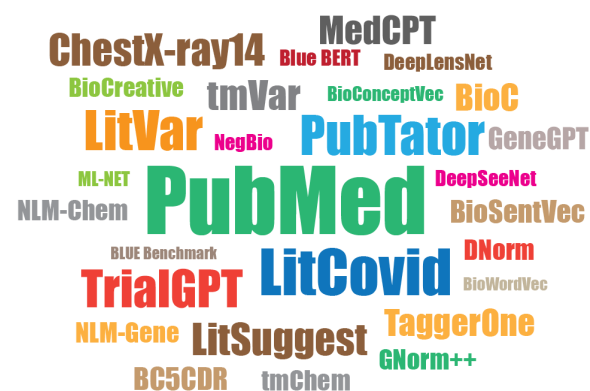
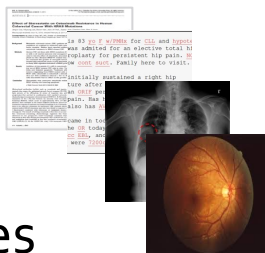
# Our research at NLM IRP

## Research Areas

AI & Machine Learning, LLMs  
Natural Language Processing (NLP)  
Medical Image Analysis

## Text & Image data

Biomedical Literature  
Clinical notes, EHRs  
CT, CXR & retinal images

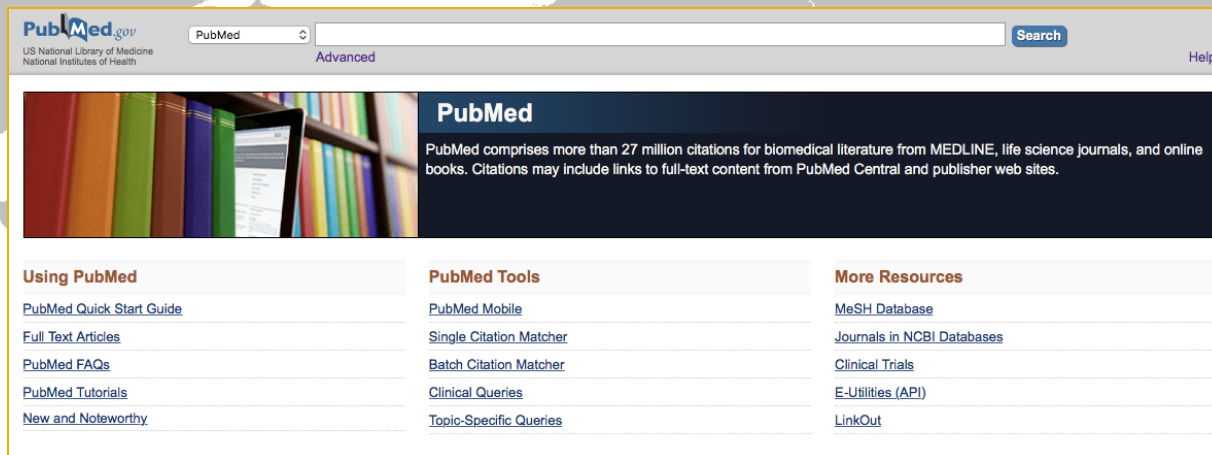


## Application areas:

Literature Retrieval (e.g., PubMed Search; LitCovid)  
Information Extraction/Curation (e.g., LitVar, PubTator)  
AI/LLMs in Healthcare (e.g., diagnosis/prognosis; TrialGPT)

# PubMed Daily Usage

- ~36+ million articles
- 2.5 million users daily
- ~ 3 million searches
- 9 million page views



The screenshot shows the PubMed website interface. At the top left is the PubMed logo with the URL [pubmed.gov](http://pubmed.gov) and the text "US National Library of Medicine National Institutes of Health". To the right is a search bar with "PubMed" in the dropdown menu and a "Search" button. Below the search bar is a "Help" link. The main content area features a header image of books and a tablet, followed by the "PubMed" title and a description: "PubMed comprises more than 27 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites." Below this are three columns of links: "Using PubMed" (including Quick Start Guide, Full Text Articles, FAQs, Tutorials, and New and Noteworthy), "PubMed Tools" (including Mobile, Citation Matchers, Clinical Queries, and Topic-Specific Queries), and "More Resources" (including MeSH Database, NCBI Databases, Clinical Trials, E-Utilities API, and LinkOut).

# AI/ML in PubMed

- Related articles
- Spell checker
- Query autosuggest
- Semantic query understanding
- Citation sensor
- Author name disambiguation
- Query expansion
- Best Match: Sort by Relevance
- ...

The screenshot shows a web page from Nature Biotechnology. At the top, there is a navigation bar with the Nature Biotechnology logo, a PDF icon, a search icon, and a user profile icon. Below the navigation bar, the article title "How user intelligence is improving PubMed" is displayed in a large font. The authors "Nicolas Fiorini, Robert Leaman [...] Zhiyong Lu" are listed below the title. A "Download Citation" link is also present. The article is categorized as a "Perspective" and was published on "01 October 2018". Below the article title, there is a section for "COMMUNITY PAGE" with the sub-title "Best Match: New relevance search for PubMed". The authors of this community page are listed as "Nicolas Fiorini, Kathi Canese, Grisha Starchenko, Evgeny Kireev, Won Kim, Vadim Miller, Maxim Osipov, Michael Kholodov, Rafis Ismagilov, Sunil Mohan, James Ostell, Zhiyong Lu\*". The authors' affiliations are listed as "National Center for Biotechnology Information (NCBI), National Library of Medicine (NLM), National Institutes of Health (NIH), Bethesda, Maryland, United States of America". A "Check for updates" button is visible. The "Abstract" section begins with "PubMed is a free search engine for biomedical literature accessed by millions of users from around the world each day. With the rapid growth of biomedical literature—about two articles are added every minute on average—finding and retrieving the most relevant papers for a given query is increasingly challenging. We present Best Match, a new relevance search algorithm for PubMed that leverages the intelligence of our users and cutting-edge machine-learning technology as an alternative to the traditional date sort order. The Best Match algorithm is trained with past user searches with dozens of relevance-ranking signals (factors), the most important being the past usage of an article, publication date, relevance score, and type of article. This new algorithm demonstrates state-of-the-art retrieval performance in benchmarking experiments as well as an improved user experience in real-world testing (over 20% increase in user click-through rate). Since its deployment in June 2017, we have observed a significant increase (60%) in PubMed searches with relevance sort order: it now assists millions of PubMed searches each week. In this work, we hope to increase the awareness and transparency of this new relevance sort option for PubMed users, enabling them to retrieve information more effectively." The article is marked as "OPEN ACCESS". The citation is: "Fiorini N, Canese K, Starchenko G, Kireev E, Kim W, Miller V, et al. (2018) Best Match: New relevance search for PubMed. PLoS Biol 16(8): e2005343. https://doi.org/10.1371/journal.pbio.2005343". The publication date is "August 28, 2018". The copyright notice states: "Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication." The footer of the page reads "Proceedings of the National Academy of Sciences".

Unknown pneumonia

2019 nCov infection

Novel coronavirus pneumonia

Coronavirus disease 2019

**COVID-19**

Wuhan coronavirus pneumonia

CoV 19 infection

SARS-CoV-2 associated ADRS

Coronavirus 2 syndrome

2019 nCoV

SARS COVID 2

Novel coronavirus

**SARS-CoV-2**

New Cov 19

HCoV-019

SC<sub>2</sub>

WH-human 1' coronavirus

New CoV

Severe acute respiratory syndrome coronavirus 2

# LitCovid: Research matters (est. 02/2020)

LitCovid is a curated literature hub for tracking up-to-date scientific information about the 2019 novel Coronavirus. It is the most comprehensive resource on the subject, providing a central access to [175579](#) (and [growing](#)) relevant articles in PubMed. The articles are updated daily and are further categorized by different research topics (e.g. [Long Covid](#)) and geographic locations for improved access. You can learn more at [Chen et al. Nature \(2020\)](#) or our [FAQ](#), and download our data [here](#).

**WEEKLY PUBLICATIONS**

**LATEST PUBLICATIONS**

**CASE REPORT**  
Acute hemorrhagic necrotizing pancreatitis in patient with COVID-19: a case report and review of literature.  
Basukala, Sunil et al. • J Surg Case Rep

**DIAGNOSIS • TREATMENT**  
COVID-19 and peripheral artery thrombosis: A mini review.  
Ehsan, Gondarzi et al. • Curr Probl Cardiol

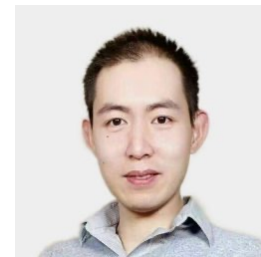
**MECHANISM • TREATMENT**  
Anti-HIV and anti-HCV small molecule protease inhibitors in-silico repurposing against SARS-CoV-2 M(pro) for the treatment of COVID-19.  
Sultan, Armiya et al. • J Biomed Struct Dyn

**PREVENTION**  
Management of COVID-19 infection in patients undergoing hemodialysis.  
Boushab, Boushab Mohamed et al. • Pan Afr Med J

**MECHANISM • TREATMENT**  
Antiviral, immunomodulatory, and anticoagulant effects of quercetin and its derivatives: Potential role in prevention and management of COVID-19.  
Manjunath, Sooparnika H et al. • J Pharm Anal

[See All Publications](#)

Chen et al., Keep up with the latest coronavirus research. *Nature*. 2020, *NAR* 2021, 2023



Qingyu Chen, postdoc Assistant Professor, Yale

“First thing in the morning, right after brushing your teeth, you click on LitCovid”

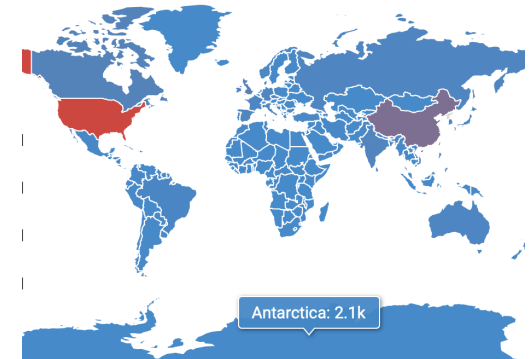
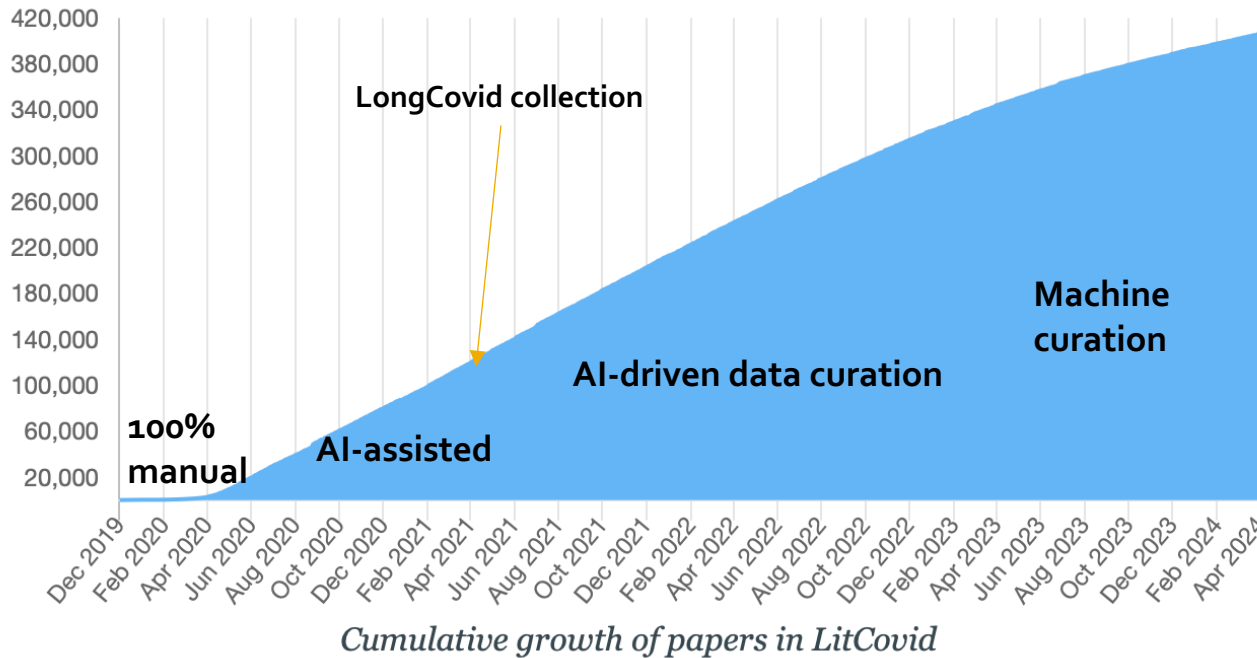
“My favorite site for following the latest [#covid19](#) literature is NLM LitCovid. Organized, comprehensive and user friendly.”

“I just wanted to say thank you for the LitCovid tool! It is an amazing resource. ”

“I just wanted to thank you for organising this informative information. I am incredibly grateful for your efforts. ”

“Thank you for putting together such a wonderful resource.”

# AI + Curator > Curator



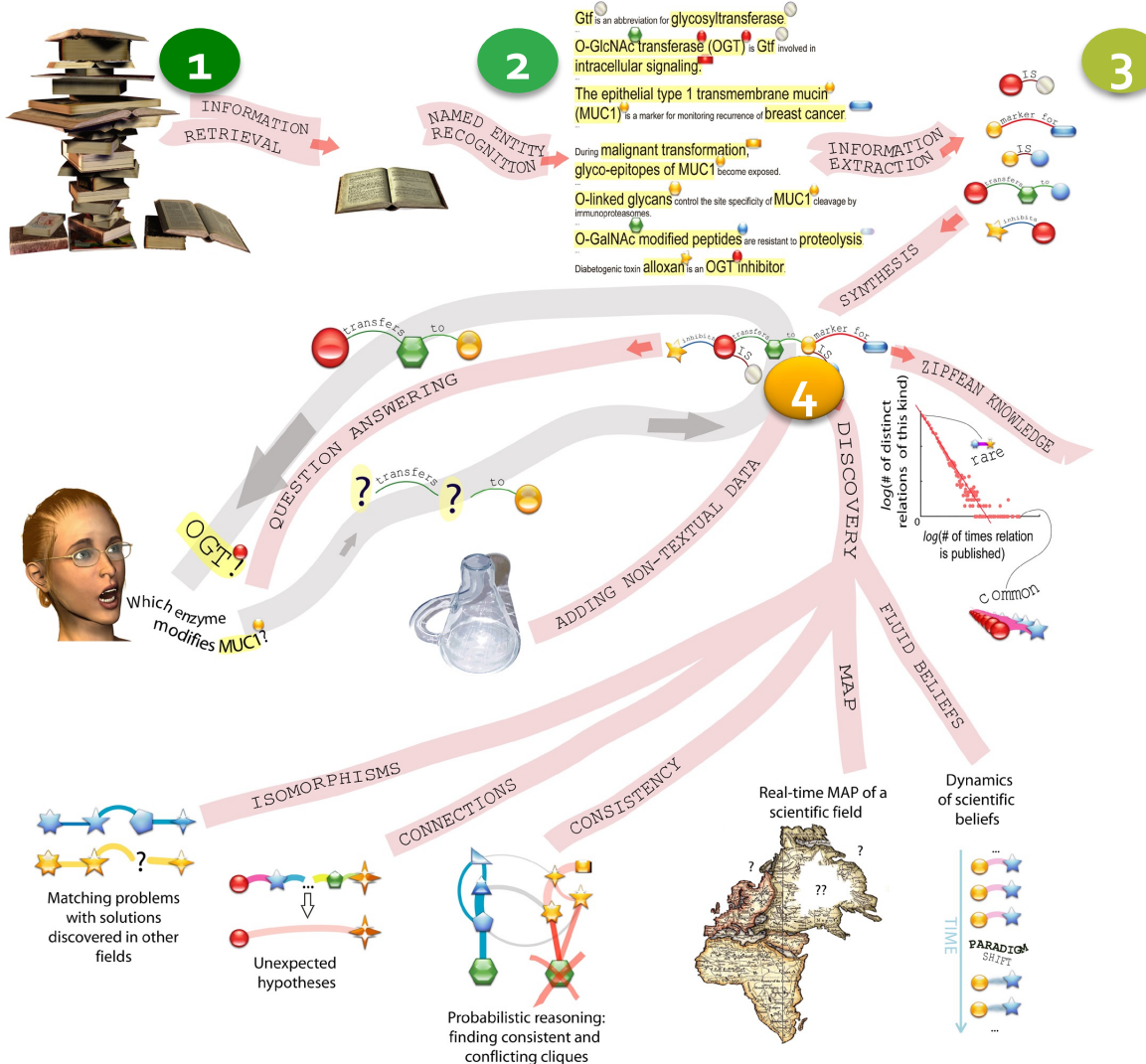
Total usage: Over 60 million hits



Lilly Xu, 2022 summer intern  
AMIA 2022 High School Scholar  
Harvard, Class of 2027

Leaman et al. Comprehensively identifying Long Covid articles with [human-in-the-loop](#) machine learning." *Patterns*, 2023  
Chen et al. LitMC-BERT: Transformer-Based Multi-Label Classification of COVID-19 Literature. *IEEE/ACM TCBB*, 2022

# Information Extraction for Discovery



1. Articles

2. Entities

3. Relations

4. Integration



# Extracting gene-disease-variant relations from free text

[J Alzheimers Dis.](#) 2012;32(2) **Disease** **A** **Gene** **Variant**

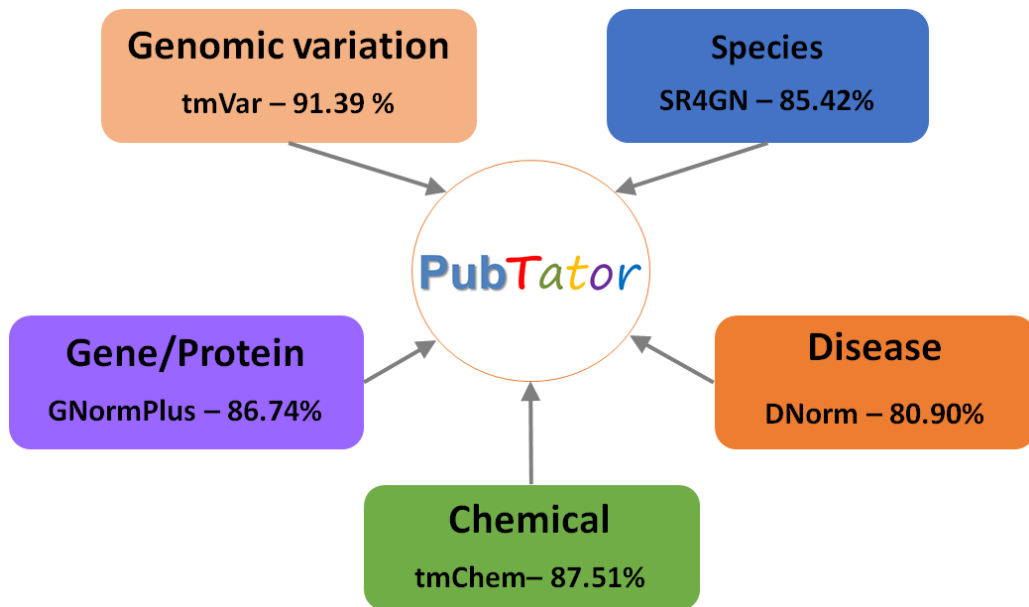
Highly pathogenic **Alzheimer's disease** **presenilin 1** **P117R** mutation causes a specific increase in p53 and p21 protein levels and cell cycle dysregulation in human lymphocytes.

[Bialopiotrowicz E](#)<sup>1</sup>, [Szybinska A](#), [Kuzniewska B](#), [Buizza L](#), [Uberti D](#), [Kuznicki J](#), [Wojda U](#).



Rank	Gene or Protein ID	Gene SYM	WTAA	MTAA	POS	Disease	PMIDs
1	Q13131	PRKAA1	Q	R	16	Breast cancer	16959974
2	P31749	AKT1	E	K	17	Breast cancer	17611497 18954143 19713527 21793738
3	P10275	AR	H	Y	874	Prostate cancer	17591767
	.....	....	....	....	....	....	.....

# PubTator (est. 2012): integrating text-mined results at PubMed scale



Curatable
  Not Curatable
  TBD
 **PubTator**
 Disease
  Species
  Mutation
  Chemical
  Gene

PMID:26022131 Selection of a novel DNA thioaptamer against HER2 structure.

Publication: Clinical \_ translational oncology : official publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico; 2015 May 29 [Full text links]

TITLE:  
Selection of a novel DNA thioaptamer against HER2 structure.

ABSTRACT:  
PURPOSE: Human epithelial growth factor receptor 2 (HER2) is over-expressed in several malignancies and represents an important therapeutic target. Aptamers are oligonucleotides that may potentially serve as tumor-homing ligand with excellent affinity and specificity for targeted cancer therapy. However, aptamers need to have nuclease resistance in order to function in vivo. The aim of this study was to generate a novel HER2 thioaptamer with enhanced nuclease resistance. METHODS: The HER2 thioaptamer is selected in an evolutionary process called systematic evolution of ligands by exponential enrichment. RESULTS: The thioaptamer could bind to the extracellular domain of HER2 with a K<sub>d</sub> of 172 nM and had minimal cross reactivity to trypsin or IgG. Moreover, the thioaptamer was found capable of binding with the HER2-positive breast cancer cells SK-BR-3 and MDA-MB-453, but not the HER2-negative cells MDA-MB-231. Notably, the thioaptamer HY6 largely maintained its structural integrity facing the nucleases in serum, while regular DNA aptamers were mostly digested. Additionally, the thioaptamer retained the capability of binding with the HER2-positive cells in the presence of serum, whereas non-thionated HER2 aptamer lost the binding function. CONCLUSION: The results indicated that the selected thioaptamer was more resistant to nuclease than regular DNA aptamers and might potentially function as a HER2-targeting ligand in complicated environment.

Concept View
  Mention View
 [Add bio-relation annotation to the table below.](#)

Entity type	Entity mention	Concept ID	Nomenclature	Delete
Disease	breast cancer	D001943	MEDIC	Delete
Disease	cancer tumor	D009369	MEDIC	Delete
Gene	HER2 Human epithelial growth factor receptor 2	2064	NCBI Gene	Delete

# PubTator milestones

**2019**

Deep Learning for improved results



**2024**

Relation annotations (PubTator 3.0)



**2016**

Bulk download available via FTP



**2018**

Full-text PMC articles added (PubTator 2.0)



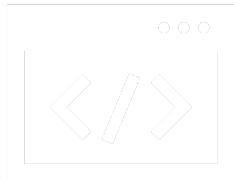
**2015**

APIs released



**2012**

PubTator first launched



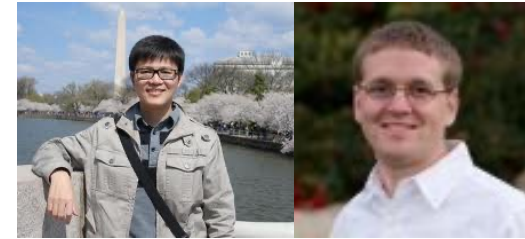
## PubTator 3.0: an AI-powered literature resource for unlocking biomedical knowledge

Chih-Hsuan Wei<sup>1</sup>, Alexis Allot<sup>2</sup>, Po-Ting Lai<sup>3</sup>, Robert Leaman<sup>4</sup>, Shubo Tian<sup>5</sup>, Ling Luo<sup>6</sup>, Qiao Jin<sup>7</sup>, Zhizheng Wang<sup>8</sup>, Qingyu Chen<sup>9</sup> and Zhiyong Lu<sup>1\*</sup>

National Center for Biotechnology Information (NCBI), National Library of Medicine (NLM), National Institutes of Health (NIH), Bethesda, MD 20894, USA

\*To whom correspondence should be addressed. Tel: +1 301 594 7089; Email: zhiyong.lu@nih.gov  
†The first two authors should be regarded as Joint First Authors.

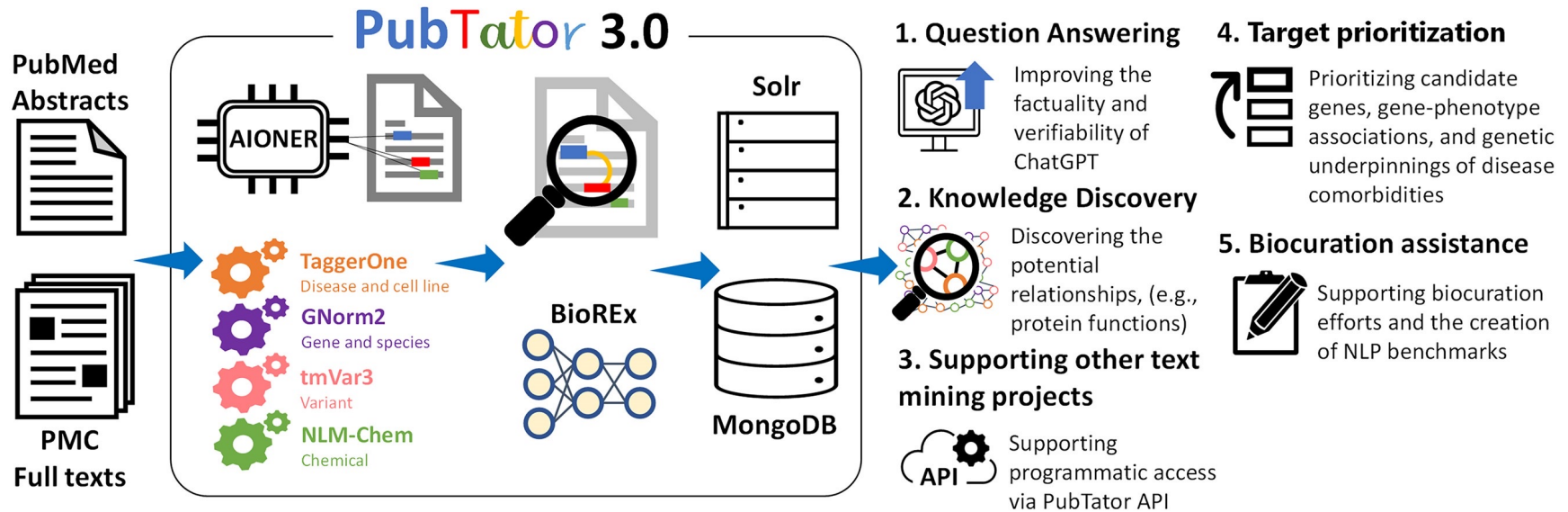
Present addresses:  
Alexis Allot, The Neuro (Montreal Neurological Institute-Hospital), McGill University, Montreal, Quebec H3A 2B4, Canada.  
Ling Luo, School of Computer Science and Technology, Dalian University of Technology, 116024 Dalian, China.  
Qingyu Chen, Biomedical Informatics and Data Science, Yale School of Medicine, New Haven, CT 06510, USA.



Staff Scientists: C. Wei, R. Leaman

### Abstract

PubTator 3.0 (<https://www.ncbi.nlm.nih.gov/research/pubtator3/>) is a biomedical literature resource using state-of-the-art AI techniques to offer semantic and relation searches for key concepts like proteins, genetic variants, diseases and chemicals. It currently provides over one billion entity and relation annotations across approximately 36 million PubMed abstracts and 6 million full-text articles from the PMC open access subset, updated weekly. PubTator 3.0's online interface and API utilize these precomputed entity relations and synonyms to provide advanced search capabilities and enable large-scale analyses, streamlining many complex information needs. We showcase the retrieval quality of PubTator 3.0 using a series of entity pair queries, demonstrating that PubTator 3.0 retrieves a greater number of articles than either PubMed or Google Scholar, with higher precision in the top 20 results. We further show that integrating ChatGPT (GPT-4) with PubTator APIs dramatically improves the factuality and verifiability of its responses. In summary, PubTator 3.0 offers a comprehensive set of features and tools that allow researchers to navigate the ever-expanding wealth of biomedical literature, expediting research and unlocking valuable insights for scientific discovery.



# API usage: over 1 billion requests



A world map showing the global distribution of API usage. Red circles of varying sizes are overlaid on the map, indicating the location and relative volume of requests. The largest circles are concentrated in North America (USA and Canada), Europe, and East Asia (China and Japan). Smaller circles are scattered across South America, Africa, and Australia.

**Logos and Institutions:**

- Top Row:** CDC (Centers for Disease Control and Prevention), ctd (Connective Tissue Database), 5 YEARS OMIM (Human Genetics Knowledge for the World), SIB (Structural Bioinformatics Institute) Rhea, swissprot (EBI), GWAS Catalog (The NHGRI-EBI Catalog of published genome-wide association studies), PharmGKB (Pharmacogenetics Knowledge Base).
- Second Row:** GLADSTONE INSTITUTES (University of California, San Diego), THE GEORGE WASHINGTON UNIVERSITY (WASHINGTON, DC), UAMS (UNIVERSITY OF ARKANSAS FOR MEDICAL SCIENCES), EMBL-EBI (European Molecular Biology Laboratory - European Bioinformatics Institute), COSMIC (Catalogue Of Somatic Mutations In Cancer).
- Third Row:** UK (UNIVERSITY OF KENTUCKY), Roche (Roche), JIMBC (JOHNS HOPKINS BAYVIEW COMMUNITY HEALTH CENTER), DBCLS (Database Center for Life Science).
- Fourth Row:** CORELL UNIVERSITY (FOUNDED A.D. 1865), CURE (Cancer Research and Biotechnology), Stanford University, DALIAN UNIVERSITY OF TECHNOLOGY (1949), KOREA UNIVERSITY (1945).
- Fifth Row:** TATA CONSULTANCY SERVICES, ILLINOIS (UNIVERSITY OF ILLINOIS CHAMPAIGN), UNIVERSITY OF DELAWARE, OMA - ONLINE MENDELIAN INHERITANCE IN ANIMALS, THE UNIVERSITY OF MELBOURNE, NANYANG AGRICULTURAL UNIVERSITY (1898), NATIONAL CHENG KUNG UNIVERSITY (1891).
- Sixth Row:** Penn (UNIVERSITY OF PENNSYLVANIA), OMIM (Online Mendelian Inheritance in Man), GEORGETOWN UNIVERSITY, wellcome sanger institute.

# Tracking genetics variants in literature

**LitVar<sup>2</sup>** NIH NLM [FAQ](#) [API](#)

Search for variants in more than 35 million biomedical publications.

Variant (e.g. CFH R1210C)

Optional Text (e.g. AMD)

**SEARCH**

Try: KRAS A146T CFH R1210C rs121913527 CA123643

LitVar allows the search and retrieval of variant specific information from relevant studies in the literature, with related concept (e.g., diseases) annotations. By normalizing variant names, LitVar returns the same results regardless of which name of a variant (e.g. BRCA1 p.P871L or c.2612C>T) is used in the query. [Read more here.](#)



Lon Phan, dbSNP



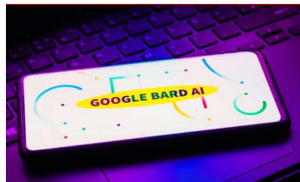
Melissa Landrum, ClinVar



Heidi Rehm, Broad

Allot et al., Tracking genetic variants in the biomedical literature using LitVar 2.0. *Nat Genet.* 2023

# ChatGPT: revolution or hype?



## Google v Microsoft: who will win the AI chatbot race?

Bard's misfire on launch cost owner \$160bn but experts believe ChatGPT is also prone to errors

10 Feb 2023

## 'ChatGPT needs a huge amount of editing': users' views mixed on AI chatbot

8 Feb 2023

## Microsoft to power Bing with AI as race with Google heats up

7 Feb 2023

## Google trials its own AI chatbot Bard after success of ChatGPT

6 Feb 2023

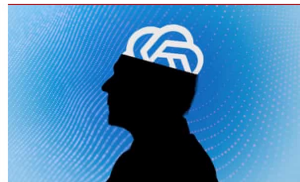
## The networker

ChatGPT isn't a great leap forward, it's an expensive deal with the devil  
*John Naughton*

4 Feb 2023

## Colombian judge says he used ChatGPT in ruling

2 Feb 2023



## Who said it: an Australian MP or ChatGPT?

10 Feb 2023

## US experts warn AI likely to kill off jobs - and widen wealth inequality

8 Feb 2023

## How will Google and Microsoft AI chatbots affect us and how we work?

7 Feb 2023

## MP tells Australia's parliament AI could be used for 'mass destruction' in speech part-written by ChatGPT

6 Feb 2023

## Google poised to release chatbot technology after ChatGPT success

3 Feb 2023

## ChatGPT reaches 100 million users two months after launch

2 Feb 2023

## ChatGPT Sprints to One Million Users

Time it took for selected online services to reach one million users



\* one million backers \*\* one million nights booked \*\*\* one million downloads

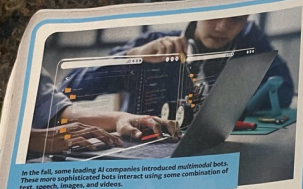
Source: Company announcements via Business Insider/LinkedIn





# MUSE: Science Magazine for Kids

## May 2024 issue



In the fall, some leading AI companies introduced multimodal bots. These more sophisticated bots interact using some combination of text, speech, images, and videos.

improve. His student Andya Gattu thought this activity was fascinating. "All of us had the same four paragraphs," she says. "We had to add the elements that the robot missed and then fix the sentences, adding style, tone, voice."

**Confidence and Creativity**  
In all of these examples, AI isn't offering up an easy answer. Instead, it has become part of the process of learning. Yes, chatbots can lead to cheating or misinformation. But many educators say when they're used carefully and openly, they make possible entirely new ways of learning and creating. Many students agree. "It's like having your teacher right next to you sometimes," says Andya. Indeed, AI companies often liken chatbots to private tutors.

As Brantley's MakerSpace class showed, AI can personalize discussion. Every student could work on something a bit different—a project or quiz or test designed for their skills.

Some liken the use of chatbots with schoolwork to the use of calculators. What do you think?

and abilities. This aspect of AI tools might help even the playing field for language learners or students with learning differences. Brantley's student Joey '01 says, "AI can change the way you learn by giving different thinkers new challenges." This could boost their confidence.

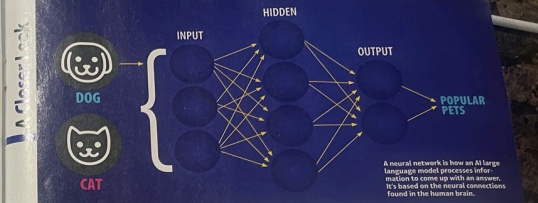
Perhaps in the future, some students may use chatbots for writing just as they now use a calculator for math. AI tools could become a regular part of the process of composing an essay or other school assignment. Teachers, though, don't allow students to use calculators until they learn how basic math works. Likewise, with AI students will still have to learn how to write for themselves. Writing is, after all, how we demonstrate our thinking on a topic.



When Kathryn Hulick asked ChatGPT to write her bio, it said she writes for Science News Express but failed to mention Muse at all. But bot: it also gave her a master's degree in science writing from MIT, which is entirely untrue, but sort of flattering.

**HOW DO YOU FEEL ABOUT USING A CHATBOT TO WRITE?**  
A world of AI-enabled learning and creating skills is opening up to many of us. Andya Gattu is a senior at Central Bucks High School South in Warrington, Pennsylvania. She says that she feels guilty sometimes when she uses ChatGPT. "Even though I know I'm using it for the right reasons, I still feel like I'm becoming a part of myself," she says. "I feel like I'm taking a shortcut, like I could be putting in more effort." Have you used ChatGPT, Hundo, or another AI tool for school? What did you use it for and how did you feel about it?

Students will always need critical thinking skills. Indeed, such skills are crucial for using AI tools. "There's critical thinking in the way that you prompt ChatGPT," says Belzote. And then you need to be able to analyze the results to make sure they make sense and are correct.  
What do you think about asking AI chatbots for help with schoolwork? Is this a shortcut that could introduce mistakes or undercut real learning? Or should it be OK, as long as you have permission, you're protecting your privacy, and it's used as a tool that's part of a process? How do you think education might change in the future because of AI?



### HOW AN A.I. CHATBOT WORKS

#### According to ChatGPT

We wanted to take a look at how an AI chatbot works. So, we tried turning to one to find out if it could provide an answer. It did! Here's a recent exchange with ChatGPT, the version based on GPT-3.5. (An upgrade to GPT-4 was unavailable at the time "due to high demand.") The answer it gave has been left unedited here, though we did fact-check it. We confirmed all the information, except the line "The collected data was then preprocessed to remove any sensitive or personally identifiable information."

**Muse**  
You are an AI chatbot. How did you become a chatbot?

**ChatGPT**  
I am indeed an AI chatbot created by OpenAI. I became a chatbot through a process called machine learning and natural language processing. Here's a simplified explanation of how it happened:

- **Data Collection:** To train me, OpenAI collected a massive amount of text data from the internet. This data included books, articles, websites, and more. The purpose of collecting this data was to expose the AI model to a wide range of language and information.
- **Preprocessing:** The collected data was then preprocessed to remove any sensitive or personally identifiable information. It was also cleaned and structured to make it suitable for training.
- **Model Training:** OpenAI used a deep learning architecture called GPT (Generative Pre-trained Transformer) to train me. During training, the model learned to understand and generate text based on the patterns and information in the training data. This process involved optimizing numerous parameters and running computations on powerful hardware, often involving graphics processing units (GPUs).
- **Fine-Tuning:** After the initial training, the model underwent fine-tuning to make it more specific and useful for certain tasks or domains. For example, it can be fine-tuned to provide medical advice, answer programming questions, or engage in conversation.
- **Deployment:** Once the model was trained and fine-tuned, it was deployed as a chatbot on various platforms and applications, including websites, messaging apps, and more. This is how I became a chatbot, capable of interacting with users like you and providing information and assistance.

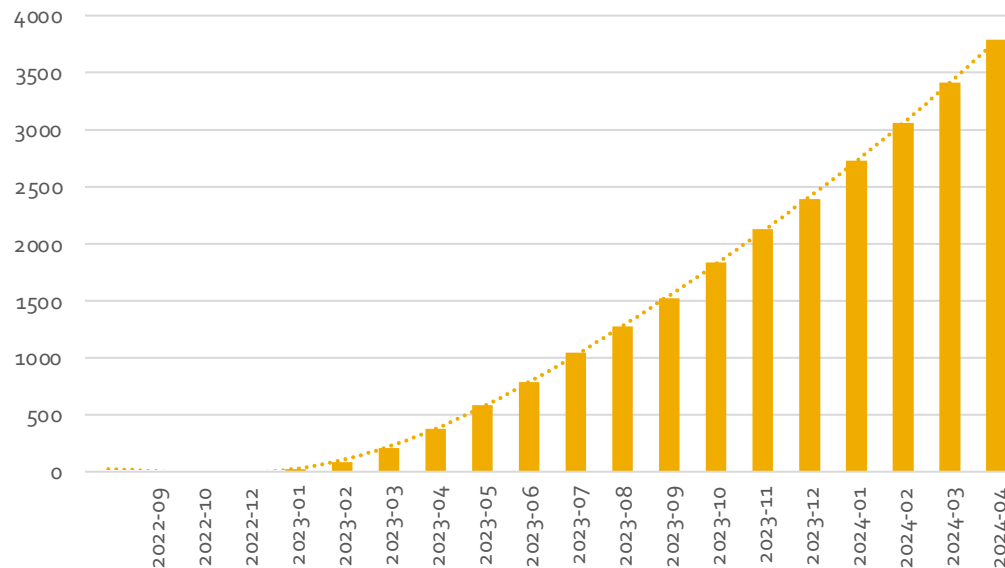
As a chatbot, my primary function is to understand and generate text responses based on the input I receive. I don't have consciousness or emotions, but I can provide information and engage in conversations to the best of my training and programming. If you have any more questions about being a chatbot or any other topic, feel free to ask!



# LLMs for Biomedicine

PubMed: “large language models” OR chatgpt

Over 1,000 journals



~170 in ophthalmology

## Large language models in medicine.

Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Nat Med. 2023 Aug;29(8):1930-1940. doi: 10.1038/s41591-023-02448-8. Epub 2023 Jul 17. PMID: 37460753 Review.

**Large language models** (LLMs) can respond to free-text queries without being specifically trained in the task in question, causing excitement and concern about their use in healthcare settings. ...Here we outline how LLM applications such as **ChatGPT** are ...

## Large language models and their impact in ophthalmology.

Betzler BK, Chen H, Cheng CY, Lee CS, Ning G, Song SJ, Lee AY, Kawasaki R, van Wijngaarden P, Grzybowski A, He M, Li D, Ran Ran A, Ting DSW, Teo K, Ruamviboonsuk P, Sivaprasad S, Chaudhary V, Tadayoni R, Wang X, Cheung CY, Zheng Y, Wang YX, Tham YC, Wong TY. Lancet Digit Health. 2023 Dec;5(12):e917-e924. doi: 10.1016/S2589-7500(23)00201-7. PMID: 38000875 Free article. Review.

The advent of generative artificial intelligence and **large language models** has ushered in transformative applications within medicine. ...This Viewpoint seeks to stimulate broader discourse on the potential of **large language models** in ...

## Benchmarking large language models' performances for myopia care: a comparative analysis of ChatGPT-3.5, ChatGPT-4.0, and Google Bard.

Lim ZW, Pushpanathan K, Yew SME, Lai Y, Sun CH, Lam JSH, Chen DZ, Goh JHL, Tan MCJ, Sheng B, Cheng CY, Koh VTC, Tham YC. EBioMedicine. 2023 Sep;95:104770. doi: 10.1016/j.ebiom.2023.104770. Epub 2023 Aug 23. PMID: 37625267 Free PMC article.

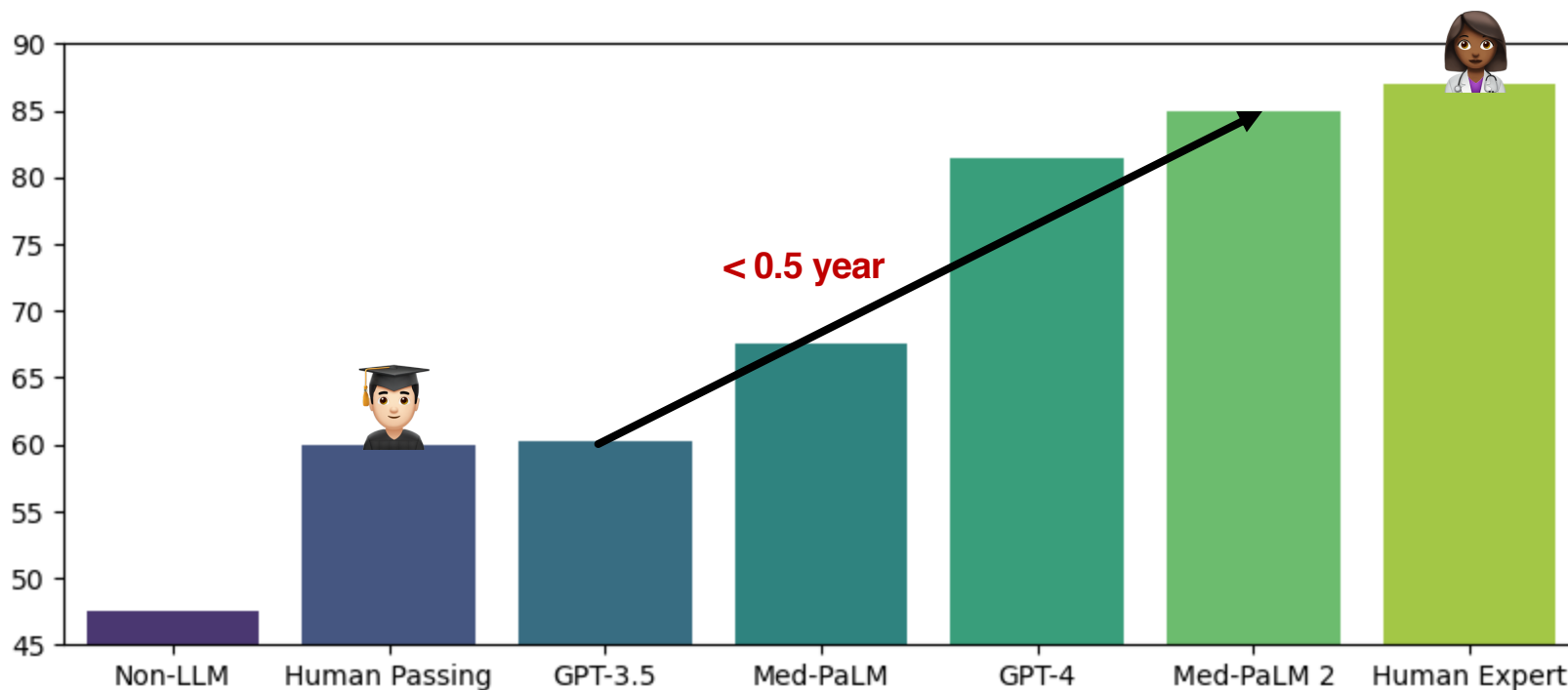
**BACKGROUND:** **Large language models** (LLMs) are garnering wide interest due to their human-like and contextually relevant responses. ...All three LLM-Chatbots showed high mean comprehensiveness scores (Google Bard: 4.35; **ChatGPT-4.0**: 4.23; **ChatGPT-** ...


## Generative Artificial Intelligence Through ChatGPT and Other Large Language Models in Ophthalmology: Clinical Applications and Challenges.

Tan TF, Thirunavukarasu AJ, Campbell JP, Keane PA, Pasquale LR, Abramoff MD, Kalpathy-Cramer J, Lum F, Kim JE, Baxter SL, Ting DSW.

Ophthalmol Sci. 2023 Sep 9;3(4):100394. doi: 10.1016/j.xops.2023.100394. eCollection 2023

# LLMs for Biomedical QA: USMLE Performance





A boy saw 17  
doctors over 3 years  
for chronic pain.  
ChatGPT found the  
diagnosis



<https://rb.gy/tueqe7>

# Hallucination in LLMs

Q1

Please list the most common symptoms associated with the long-term sequelae of COVID-19. Each symptom should list the source article title and PMID.

fabricated



Sure, here are some of the most common symptoms associated with long-term sequelae of COVID-19, along with their corresponding source article title and PMID:

1. Fatigue and Muscle Weakness: "Post-acute COVID-19 syndrome: a case series and multidisciplinary approach to treatment" (PMID: 33571309)
2. Shortness of Breath: "Symptomatic Outcomes in Young Adults With COVID-19" (PMID: 32936950)
3. Joint Pain: "Musculoskeletal symptoms in COVID-19 patients" (PMID: 32767232)
4. Chest Pain: "Chest CT Findings in Long-COVID Patients with Dyspnea and Chest Pain" (PMID: 33828099)
5. Cognitive Dysfunction: "Neuropsychiatric and Cognitive Sequelae of COVID-19" (PMID: 33402831)

**(truncated response)**

# Retrieval Augmented Generation



Please list the most common symptoms associated with the long-term sequelae of COVID-19. Each symptom should list the source PMID.

Review > Nat Immunol. 2022 Feb;23(2):194-202. doi: 10.1038/s41590-021-01104-y. Epub 2022 Feb 1.

### Pathological sequelae of long COVID

Saurabh Mehandru<sup>1,2</sup>, Miriam Merad<sup>3,4</sup>

Affiliations + expand  
PMID: 35105985 PMCID: PMC9127978 DOI: 10.1038/s41590-021-01104-y  
Free PMC article

### Abstract

The world continues to contend with successive waves fueled by the emergence of viral variants. At the same time, debilitating sequelae are increasingly recognized in the '19 syndrome' or 'long-haul COVID'. Clinical symptom defects in memory and concentration and a variety of manifestations, and several organ systems can be involved, but the mechanisms are poorly understood at present. This review synthesizes the current evidence on the underlying pathophysiology of long COVID and examines the underlying mechanisms of persistent inflammation, immune dysregulation, and organ dysfunction. Finally, we propose diagnostic strategies to better understand long COVID and how it continues to afflict millions of people worldwide.

© 2022. Springer Nature America, Inc.

Meta-Analysis > JAMA. 2022 Oct 25;328(16):1604-1615. doi: 10.1001/jama.2022.19422.

### Estimated Global Proportions of Incident Persistent Fatigue, Cognitive, and Symptom Clusters Following Symptom Onset in 2020 and 2021

Global Burden of Disease Long COVID Collaborators; Sarah Wulf<sup>1</sup>, Joachim G Aerts<sup>2</sup>, Ziyad Al-Aly<sup>3,4,5</sup>, Charlie Ashbaugh<sup>1</sup>, Tala B Polina Bobkova<sup>6</sup>, Gouke Bonsel<sup>7</sup>, Svetlana Borzakova<sup>11,12</sup>, Denis Butaru<sup>15</sup>, Austin Carter<sup>1</sup>, Helen Chu<sup>16</sup>, Cristina De Rooz Mohamed Mustafa Diab<sup>17,18</sup>, Emil Ekboom<sup>19</sup>, Maha El Tantawi<sup>20</sup>, Robert Firthio<sup>22</sup>, Aysyu Gamirova<sup>23</sup>, Petr V Glybochko<sup>24</sup>, Jus Shaghayegh Haghighy Javanmard<sup>25</sup>, Erin B Hamilton<sup>1</sup>, Gabriel Majanka H Hejtenbrok-Kal<sup>28,29</sup>, Raimund Helbok<sup>30</sup>, Merel E H Susame M Huijts<sup>32</sup>, Michael Hultström<sup>32,33</sup>, Waasila Jassat<sup>34</sup>, Ing-Marie Larsson<sup>22</sup>, Mikós Lipcsey<sup>22</sup>, Chelsea Liu<sup>35</sup>, Callan Andrei Malnovschi<sup>38</sup>, Wenhui Mao<sup>17,39</sup>, Lyudmila Mazankova<sup>40</sup>, Dominik Menges<sup>6</sup>, Noushin Mohammadifard<sup>41</sup>, Daniel Murbitt<sup>8,42</sup>, Nikita A Nekliudov<sup>23</sup>, Csöndu Ogburn<sup>39</sup>, Ismail M Osmanov<sup>11,43</sup>, José L Peñalvo<sup>44,45</sup>

Affiliations + expand  
PMID: 36631153 PMCID: PMC9832503 DOI: 10.1136/bmj-2022-072529  
Free PMC article

### Abstract

**Objectives:** To determine the clinical sequelae of long covid for a year after infection with mild disease and to evaluate its association with age, sex, SARS-CoV-2 variant, and vaccination status.

**Design:** Retrospective nationwide cohort study.

**Setting:** Electronic medical records from an Israeli nationwide healthcare organization.

**Population:** 1 913 234 Maccabi Healthcare Services members of all ages who were vaccinated against SARS-CoV-2 between 1 March 2020 and 1 October 2021.

**Main outcome measures:** Risk of an evidence based list of 70 reported long COVID symptoms in unvaccinated patients infected with SARS-CoV-2 matched to uninfected people and sex and stratified by SARS-CoV-2 variant, and risk in patients with a breast cancer diagnosis.

BMJ. 2023 Jan 11;380:e072529. doi: 10.1136/bmj-2022-072529.

### Long covid outcomes at one year after mild SARS-CoV-2 infection: nationwide cohort study

Barak Mizrahi<sup>1</sup>, Tomer Sudry<sup>1</sup>, Natalie Falck-Marov<sup>1</sup>, Yeav Yeheskieli<sup>1</sup>, Nir Pinchas Akiva<sup>1</sup>, Anat Ekka-Zohar<sup>2</sup>, Shirley Shapiro Ben David<sup>2</sup>, Uri Lerner<sup>3</sup>, Maytal Bivas-Benita<sup>1</sup>, Shira Greenfeld<sup>4</sup>

Affiliations + expand  
PMID: 36631153 PMCID: PMC9832503 DOI: 10.1136/bmj-2022-072529  
Free PMC article

### Abstract

**Objectives:** To determine the clinical sequelae of long covid for a year after infection with mild disease and to evaluate its association with age, sex, SARS-CoV-2 variant, and vaccination status.

**Design:** Retrospective nationwide cohort study.

**Setting:** Electronic medical records from an Israeli nationwide healthcare organization.

**Population:** 1 913 234 Maccabi Healthcare Services members of all ages who were vaccinated against SARS-CoV-2 between 1 March 2020 and 1 October 2021.

**Main outcome measures:** Risk of an evidence based list of 70 reported long COVID symptoms in unvaccinated patients infected with SARS-CoV-2 matched to uninfected people and sex and stratified by SARS-CoV-2 variant, and risk in patients with a breast cancer diagnosis.

Observational Study > Lancet. 2022 Aug 6;400(10350):452-461. doi: 10.1016/S0140-6736(22)01214-4.

### Persistence of somatic symptoms after COVID-19 in the Netherlands: an observational cohort study

Aranka V Ballering<sup>1</sup>, Sander K R van Zon<sup>2</sup>, Tim C Olde Hartman<sup>3</sup>

Affiliations + expand  
PMID: 35934007 PMCID: PMC9352274 DOI: 10.1016/S0140-6736(22)01214-4  
Free PMC article

### Abstract

**Background:** Patients often report various symptoms after recovery from COVID-19. Previous studies on post-COVID-19 condition have not corrected for these common symptoms before COVID-19 and in populations with different aims to analyse the nature, prevalence, and severity of long-term symptoms while correcting for symptoms present before SARS-CoV-2 infection or symptom dynamics in the population without infection.

**Methods:** This study is based on data collected within Lifelines, a population-based, observational cohort study examining the health of people living in the north of the Netherlands. All Lifelines participants aged 18 years or older received invitations to digital COVID-19 questionnaires. Longitudinal dynamics of 23 somatic

Nat Med. 2022 Nov;28(11):2406-2415. doi: 10.1038/s41591-022-02001-z. Epub 2022 Sep 22.

### Long-term neurologic outcomes of COVID-19

Evan Xu<sup>1</sup>, Yan Xie<sup>1,2,3</sup>, Ziyad Al-Aly<sup>4,5,6,7,8</sup>

Affiliations + expand  
PMID: 36138154 PMCID: PMC9671811 DOI: 10.1038/s41591-022-02001-z  
Free PMC article

### Abstract

The neurologic manifestations of acute COVID-19 are well characterized, but a comprehensive evaluation of postacute neurologic sequelae at 1 year has not been undertaken. Here we use the national healthcare databases of the US Department of Veterans Affairs to build a cohort of 154,068 individuals with COVID-19, 5,638,795 contemporary controls and 5,859,621 historical controls; we use inverse probability weighting to balance the cohorts, and estimate risks and burdens of incident neurologic disorders at 12 months following acute SARS-CoV-2 infection. Our results show that in the postacute phase of COVID-19, there was increased risk of an array of incident neurologic sequelae including ischemic and hemorrhagic stroke, cognition and memory disorders, peripheral nervous system disorders, episodic disorders (for example, migraine and seizures), extrapyramidal and movement disorders, mental health disorders, musculoskeletal disorders, sensory disorders, Guillain-Barré syndrome, and encephalitis or encephalopathy. We

Providing relevant articles to ChatGPT

# Retrieval → Summarization → Verification



Please list the most common symptoms associated with the long-term sequelae of COVID-19. Each symptom should list the source PMID.

5 x [PMID, title, abstract]

{35105985, 36215063, 36631153, 35934007, 36138154}

😬 **But there are still problems**

combination error	not a symptom
missing citation	incorrect citation



Based on the sources provided, some common symptoms associated with the long-term sequelae of COVID-19 include:

1. Fatigue and malaise (PMID: 35105985; PMID: 36215063)
2. Dyspnea (difficulty breathing) (PMID: 35105985; PMID: 36215063)
3. Defects in memory and concentration (PMID: 35105985)
4. Cognitive problems (PMID: 36215063; PMID: 35934007)
5. Neuropsychiatric syndromes (PMID: 35105985)
6. Persistent inflammation (PMID: 35105985)

(truncated response)

# Our work on medical LLMs

- Evaluating LLMs on medical benchmarks
  - Scientific communications (*J Am Soc Nephrol.*, 2023)
  - BioNLP tasks (*nature communications*, under revision)
  - Hidden flaws in multi-modal GPT<sub>4</sub>V (*npj Digital Medicine*, under revision)
- Enhancing standard LLMs
  - Retrieval augmented generation (MedRAG, *ACL*, under review)
  - Tool learning (GeneGPT, *bioinformatics*, 2024)
  - Incorporating domain knowledge (GeneAgent for gene set analysis)
- Novel applications/AI agents
  - TrialGPT: patient-trial matching (*nature communications*, under revision)
  - AgentMD: medical risk calculation
- AI Safety
  - Trustworthiness in clinical evidence synthesis (*J Biomed Inform*, 2024)
  - Bias & fairness (communications medicine, under revision)
  - Adversarial attacks & risks

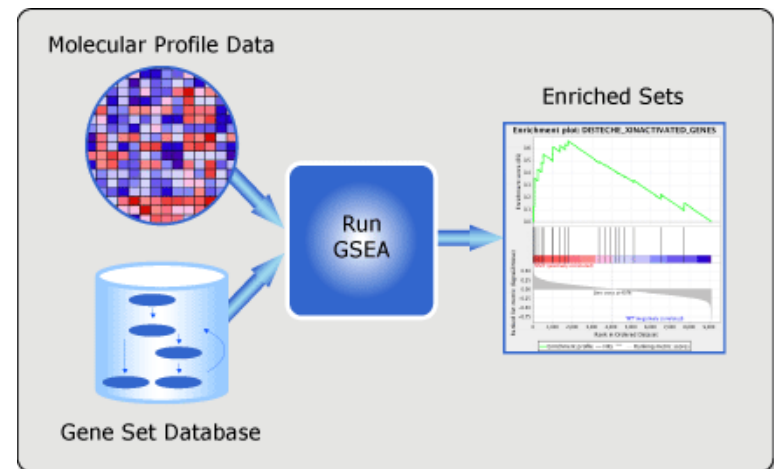
# Gene Set Analysis

- Gene sets from high-throughput experiments
- Differentially expressed genes under different conditions
- Data analysis goal: determine the collective functions by a group of genes



# Existing GSEA methods & their limitations

- Limited to curated knowledge and/or predefined gene sets
- No explanation to support predictions



# Related works

[Submitted on 21 May 2023 (v1), last revised 25 May 2023 (this version, v2)]

## Gene Set Summarization using Large Language Models

Marcin P. Joachimiak, J. Harry Caufield, Nomi L. Harris, Hyeongsik Kim, Christopher J. Mungall

Molecular biologists frequently interpret gene lists derived from high-throughput experiments and computational analysis. This is typically done as a statistical enrichment analysis that measures the over- or under-representation of biological function terms associated with genes or their properties, based on curated assertions from a knowledge base (KB) such as the Gene Ontology (GO). Interpreting gene lists can also be framed as a textual summarization task, enabling the use of Large Language Models (LLMs), potentially utilizing scientific texts directly and avoiding reliance on a KB.

We developed SPINDOCTOR (Structured Prompt Interpolation of Natural Language Descriptions of

Controlled Terms for Ontology Reporting), a method for [Submitted on 7 Sep 2023 (v1), last revised 1 Apr 2024 (this version, v2)]

summarization as a complement to standard enrichment sources of gene functional information: (1) structured annotations, (2) ontology-free narrative gene summaries.

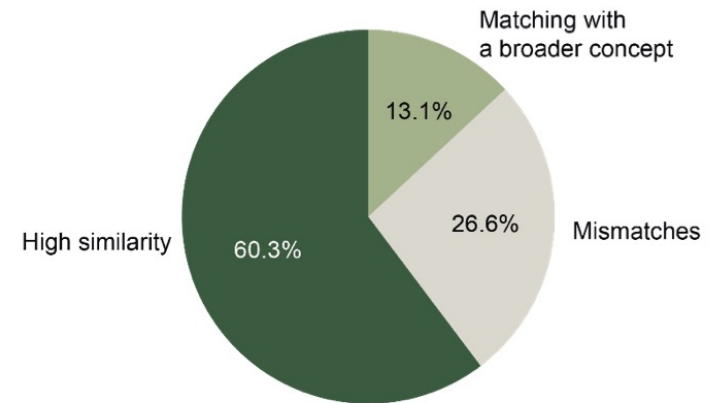
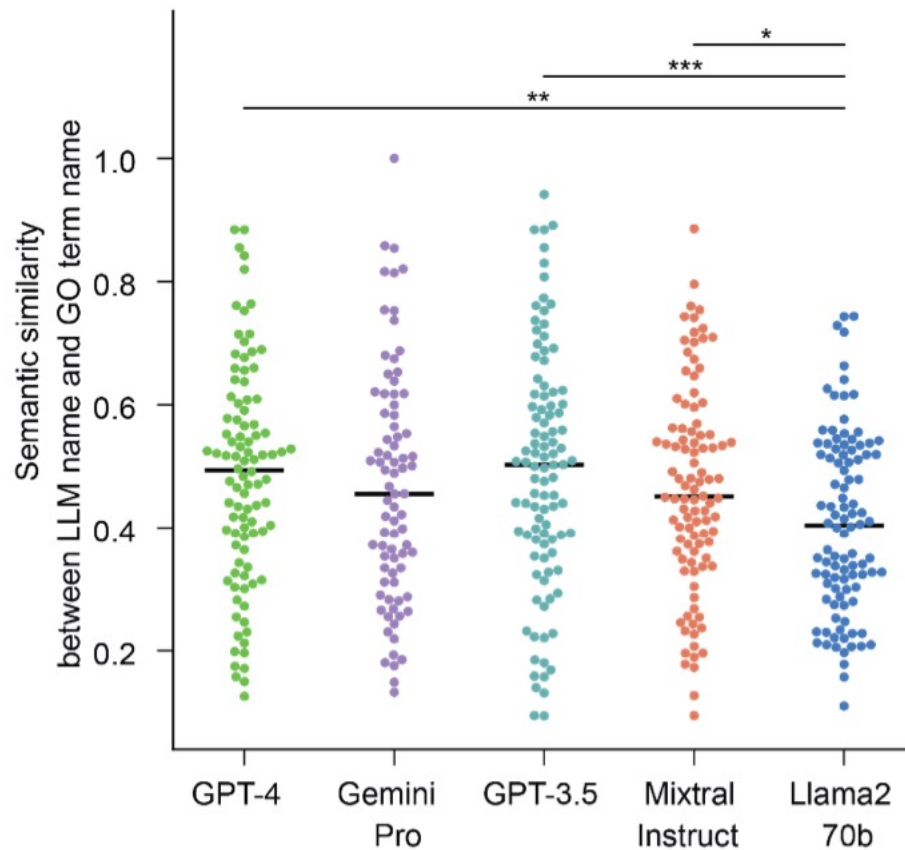
We demonstrate that these methods are able to generate term lists for gene sets. However, GPT-based approaches often return terms that are not statistically able to recapitulate the most precise and informative terms. Inability to generalize and reason using an ontology. R variations in prompt resulting in radically different terms based methods are unsuitable as a replacement for structured curation of ontological assertions remains necessary.

## Evaluation of large language models for discovery of gene set function

Mengzhou Hu, Sahar Alkhairy, Ingo Lee, Rudolf T. Pillich, Dylan Fong, Kevin Smith, Robin Bachelder, Trey Ideker, Dexter Pratt

Gene set analysis is a mainstay of functional genomics, but it relies on curated databases of gene functions that are incomplete. Here we evaluate five Large Language Models (LLMs) for their ability to discover the common biological functions represented by a gene set, substantiated by supporting rationale, citations and a confidence assessment. Benchmarking against canonical gene sets from the Gene Ontology, GPT-4 confidently recovered the curated name or a more general concept (73% of cases), while benchmarking against random gene sets correctly yielded zero confidence. Gemini-Pro and Mixtral-Instruct showed ability in naming but were falsely confident for random sets, whereas Llama2-70b had poor performance overall. In gene sets derived from 'omics data, GPT-4 identified novel functions not reported by classical functional enrichment (32% of cases), which independent review indicated were largely verifiable and not hallucinations. The ability to rapidly synthesize common gene functions positions LLMs as valuable 'omics assistants.

# Key results in Hu et al.,



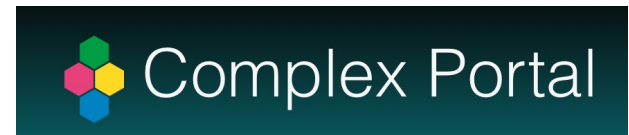
# Gene-centric information in expert-curated biological databases



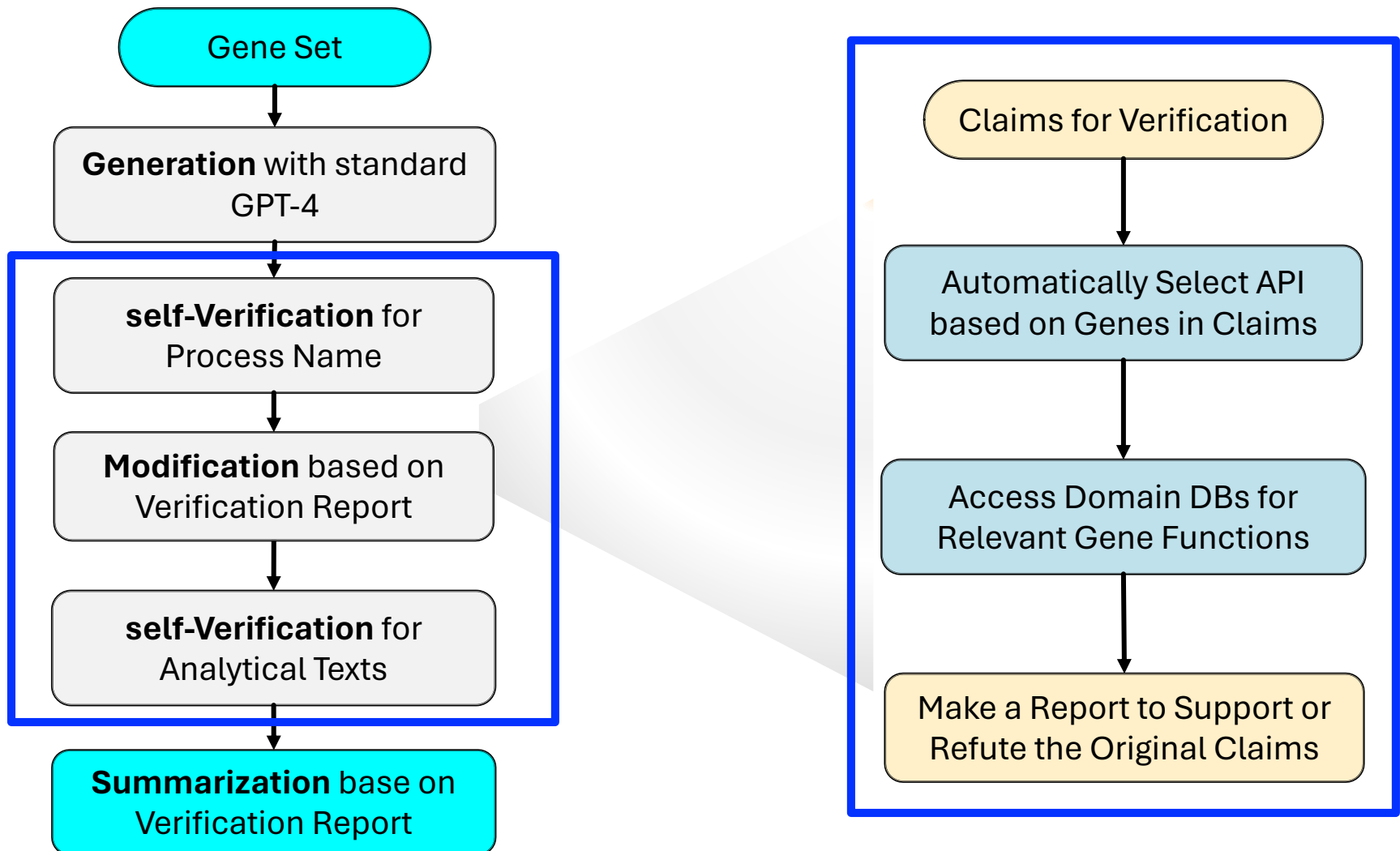
GENEONTOLOGY  
Unifying Biology



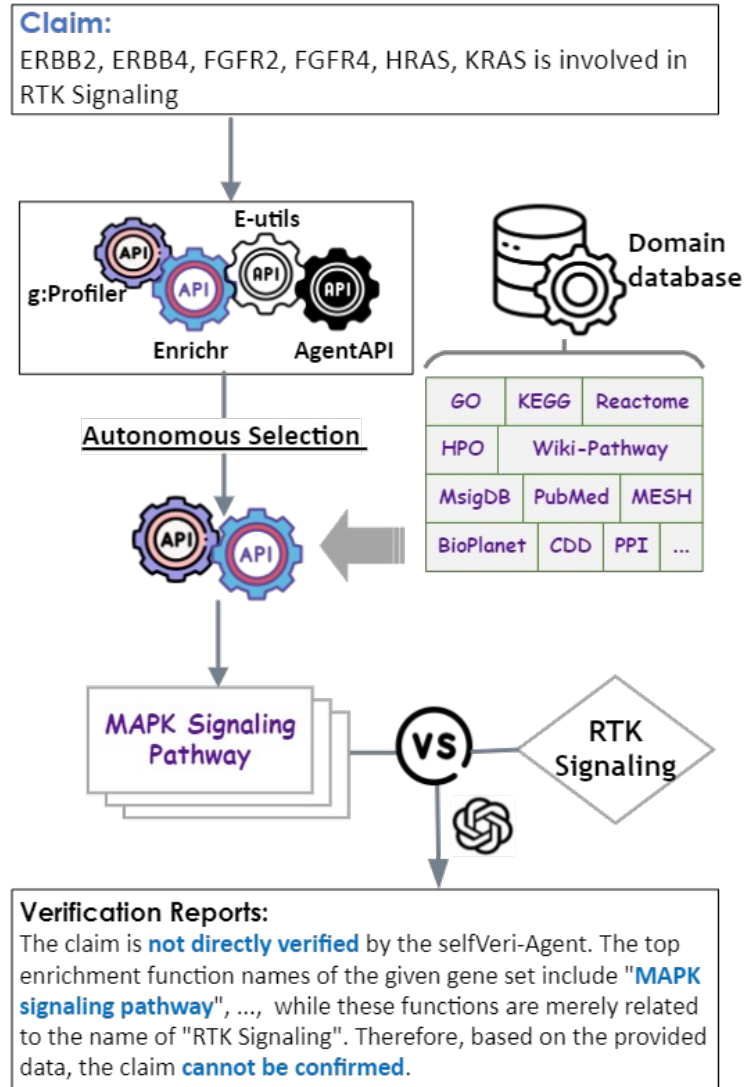
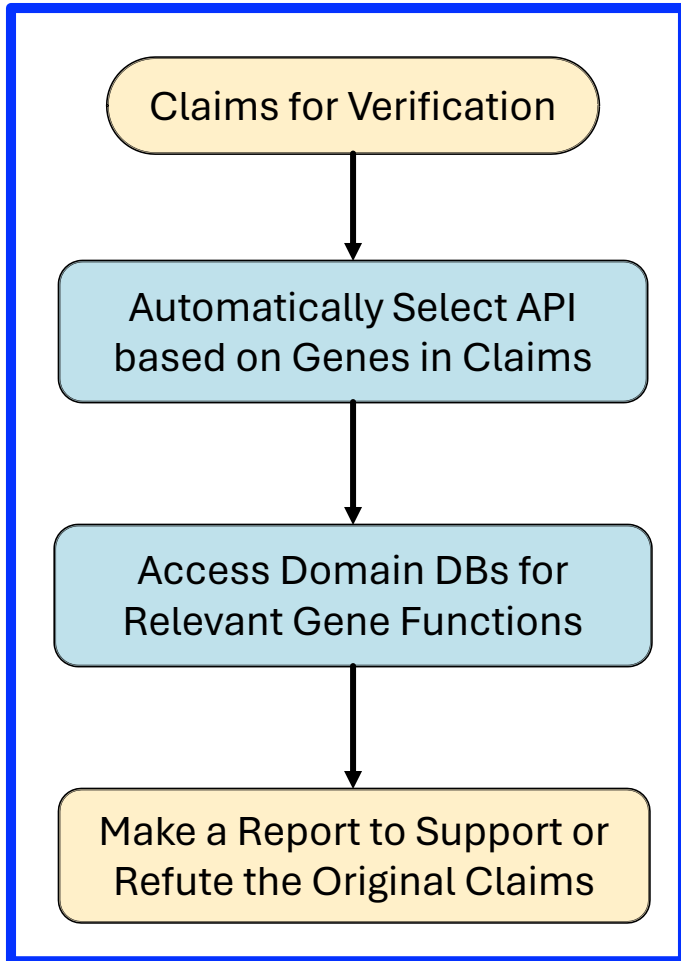
BioPlanet



# GeneAgent: overall workflow

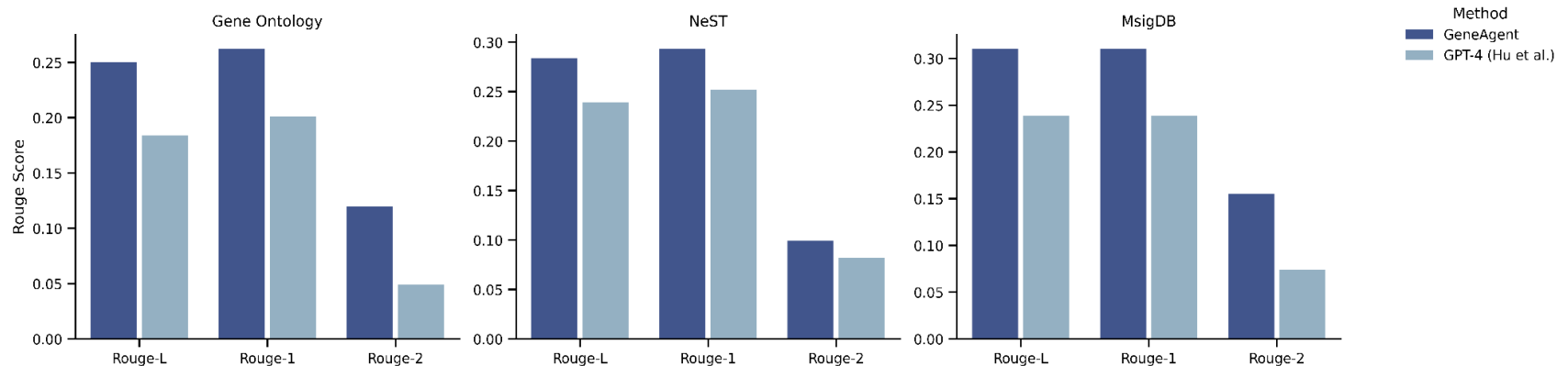


# Example of self-Verification



# GeneAgent vs. standard GPT-4

Dataset	#gene sets	#genes	Avg. genes
Gene Ontology	1,000	3 to 456	48.32
NeST	50	5 to 323	18.96
MsigDB	56	4 to 200	112.00
All	1,106	3 to 456	50.67



# Human review of LLM-generated verification reports

- Self-verification report (partially) supports or refutes original claims

## Claim:

ERBB2, ERBB4, FGFR2, FGFR4, HRAS, KRAS is involved in RTK Signaling

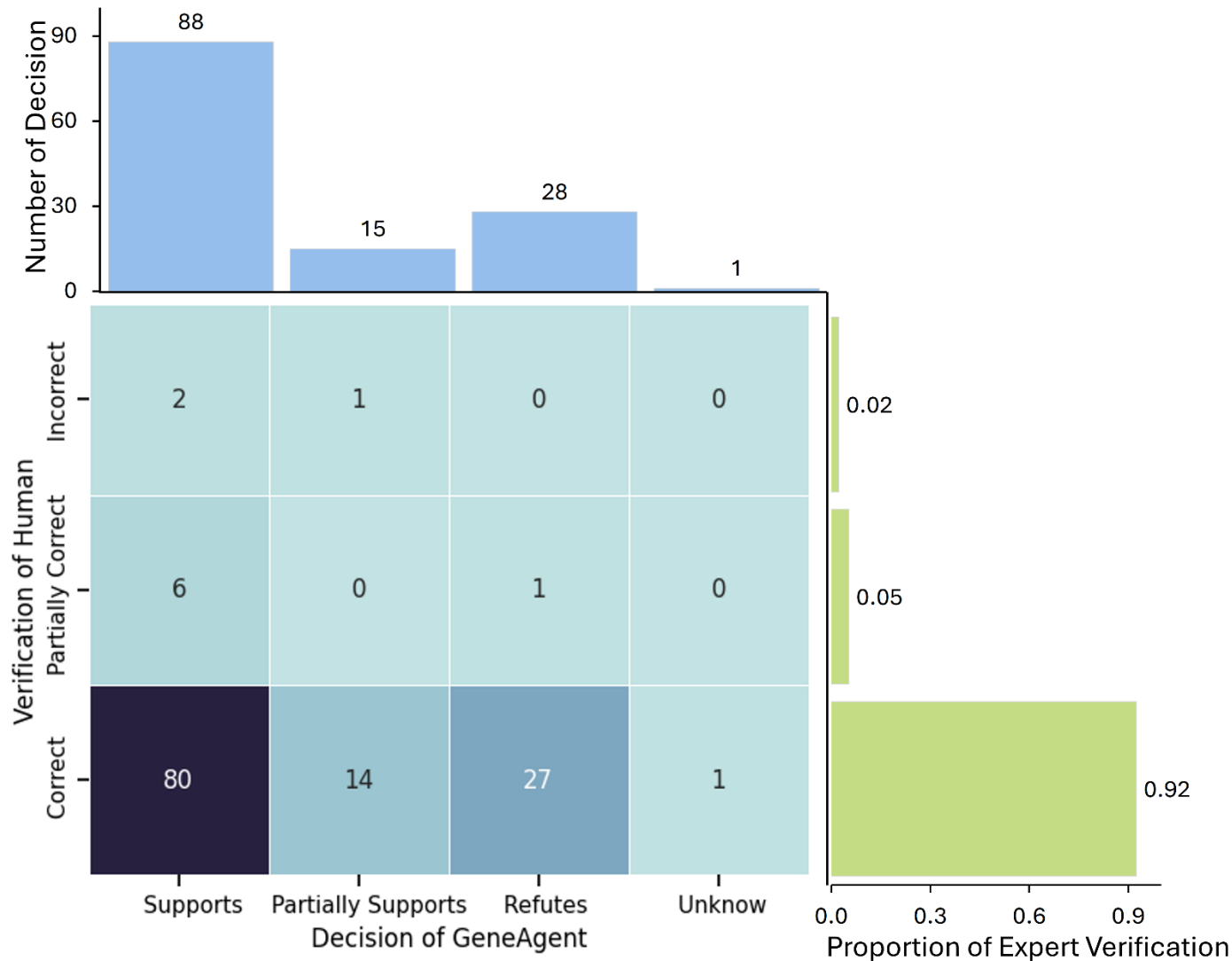
## Verification Reports:

The claim is **not directly verified** by the selfVeri-Agent. The top enrichment function names of the given gene set include "**MAPK signaling pathway**", ..., while these functions are merely related to the name of "RTK Signaling". Therefore, based on the provided data, the claim **cannot be confirmed**.

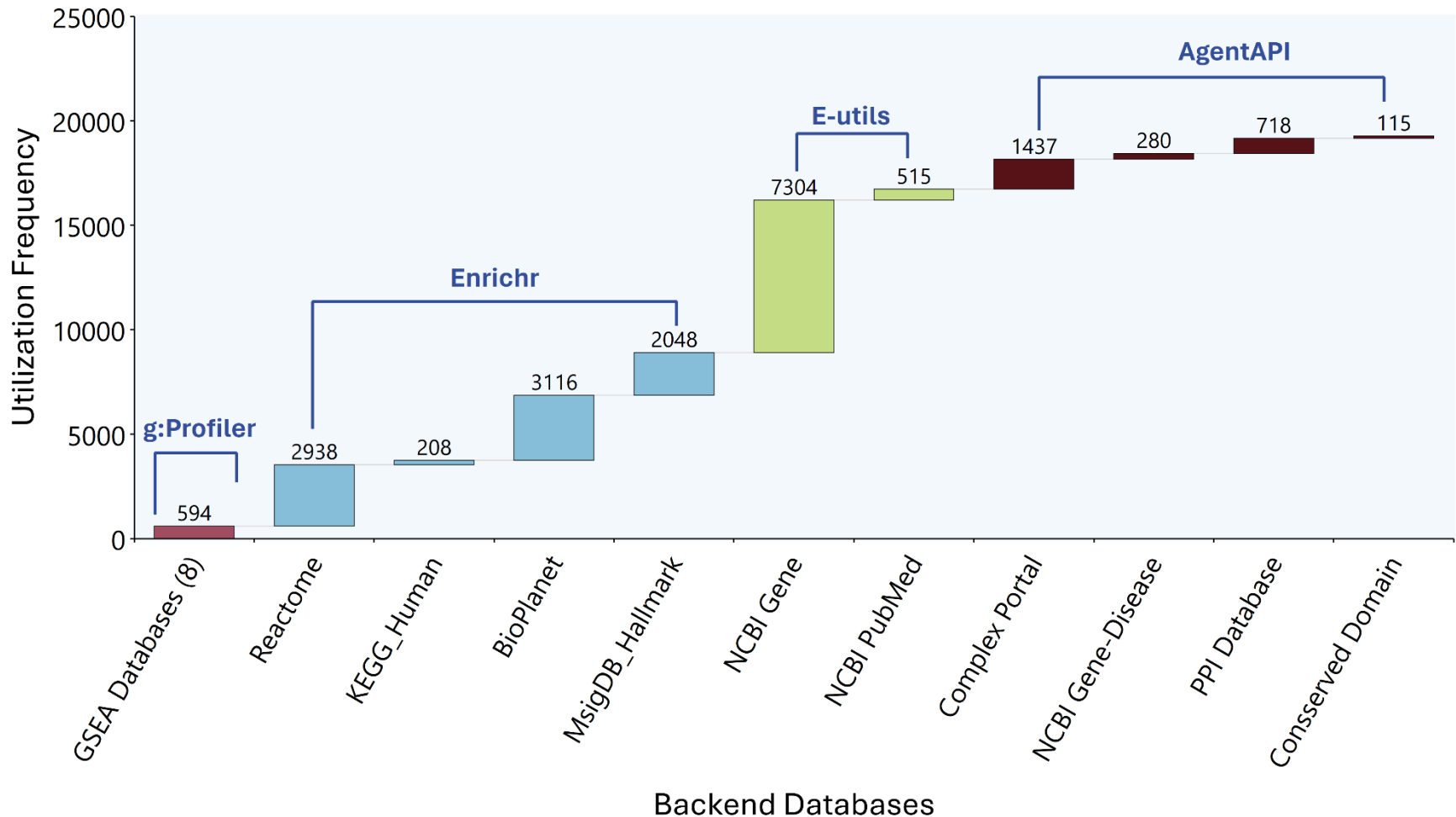
- 132 manually reviewed
  - Double annotated
  - High inter-rater agreement



# Self-verification is highly accurate



# Utilization of different KBs

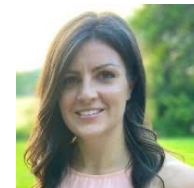


# Pilot study with novel gene sets

- To assess its potential utility in real-world applications
- Worked with domain experts from NCI
- Novel gene sets from mouse B2905 melanoma cell line



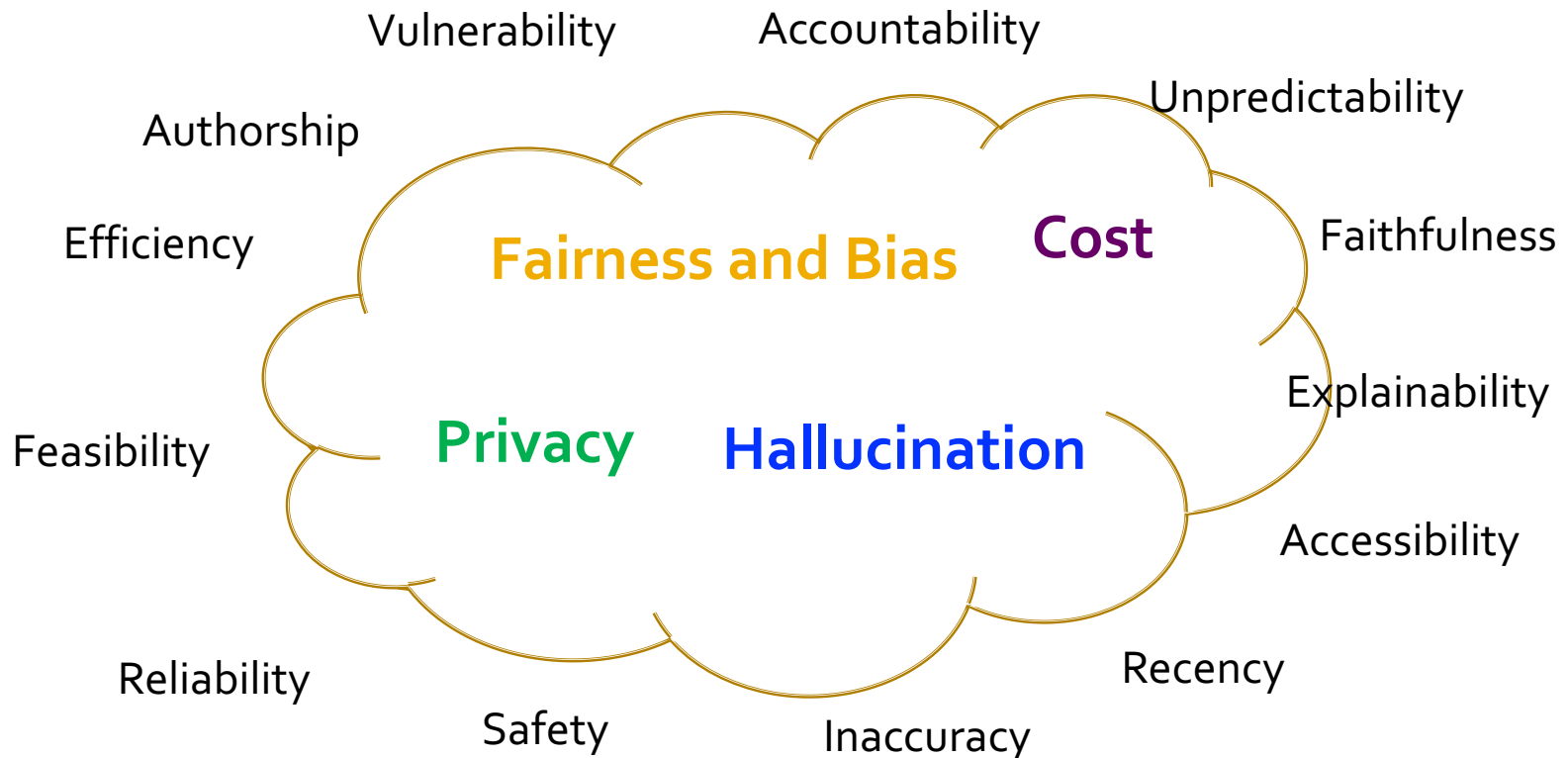
Chi-Ping Day PhD  
Lab of Cancer Biology and Genetics,  
Cancer Data Science Lab, NCI



Christina Ross, Ph.D. NCI  
Lab of Cancer Biology and Genetics



# Limitations & Risks of LLMs in Medicine



Tian et al. Opportunities and Challenges for ChatGPT and Large Language Models in Biomedicine and Health. *Briefings in Bioinformatics*, 2024.

# Acknowledgments

**Lu lab at NLM/NCBI:** Don Comeau, Rezarta Dogan, Amr Elsayy, Won Kim, Qiao Jin, Robert Leaman, Po-Ting Lai, Ashley Shin, Shubo Tian, Zhizheng Wang, Chih-Hsuan Wei, John Wilbur, Natalie Xie, Yifan Yang, Lana Yeganova, Qingqing Zhu

**NEI:** Emily Chew, Michael Chiang, Tiarnan Keenan

**CC:** Ron Summers, Le Lu, Xiaosong Wang, Ke Yan, etc.

**NCATS:** Tyler Beck, Christine Colvis

**NCI:** Harris Floudas, James Gulley

**PubMed:** Grisha Starchenko, Jessica Chan, Kathi Canese, etc.

**dbSNP:** Lon Phan, Juliana Feltz, Rama Maiti, Tim Hefferon

**ClinVar:** Melissa Landrum

**BioCreative:** Cecilia Arighi, Graciela Gonzalez-Hernandez, Martin Krallinger, Lynette Hirschman

**UniProt:** Alex Bateman, Alan Bridge, Livia Famiglietti, Michele Magrane, Sylvain Poux, Cathy Wu

**GWAS Catalog:** Jackie MacArthur, Fiona Cunningham, Helen Parkinson