



Practical tips for using GenAI in Biomedicine

Tuesday, October 22, 2024

Priyata Kalra

Senior Scientist
Simulations Plus

Image: DALLE claymation prompt

Are You GenAI-Fatigued?

nature

SPOTLIGHT | 30 May 2018

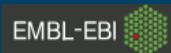
How artificial intelligence is changing drug discovery

Machine learning and other technologies are expected to make the hunt for new pharmaceuticals quicker, cheaper and more effective.

FAST COMPANY

How techbio is shaking up pharma

Change is coming through AI, with the potential to better decipher human biology and disease.



AI for drug discovery: what's the hold up?



Better Molecules, Faster: NVIDIA NIM Agent Blueprint Redefines Hit Identification With Generative AI-Based Virtual Screening

Benchling, Dotmatics, Terray, TetraScience and Cadence Molecular Sciences to use NVIDIA NIM microservices and NIM Agent Blueprints to push the boundaries of drug discovery.

August 27, 2024 by Anthony Costa

Benevolent^{AI}

We used our BenAI Engine to identify a leading COVID-19 treatment, which is now FDA approved.

Our Relationships with GenAI

1. **GenAI-Hesitant**
2. **GenAI-Harmonist**
3. **GenAI-Maximalist**

Our Relationships with GenAI

1. GenAI-Hesitant
2. GenAI-Harmonist
3. GenAI-Maximalist

What is the buzz?

- No one is talking about how humans adapt to AI
- Most employers don't have the answers

... And best practices for working with GenAI responsibly

Our Relationships with GenAI

1. GenAI-Hesitant
2. GenAI-Harmonist
3. GenAI-Maximalist

Augment yourself with AI

- Some creative use-cases and tools that hopefully (?) empower you

Our Relationships with GenAI

1. GenAI-Hesitant
2. GenAI-Harmonist
3. GenAI-Maximalist

The latest and greatest, and why you might care as someone who works with data

GenAI-Hesitant Skills

AI Literacy

Curiosity

Adaptability

Critical Thinking

Communication

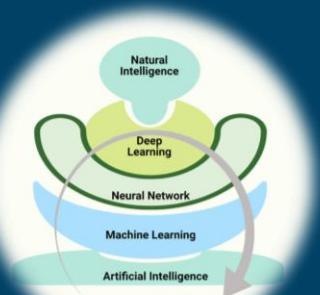
Design thinking

AI Literacy

AI deals with the research on and engineering of technical systems, that are capable of behavior that humans attribute to intelligence

Narrow AI/ Weak AI

Excels in specific tasks within set parameters but lacks general intelligence



GenAI/LLM

Designed to understand and generate human-like text using deep learning.

Trained on vast data, they learn the patterns, structures and nuances of **natural language**.

AGI/ASI

Deals with (future) systems that show a very broad and flexible, human-like intelligence.

Futuristic concept and poses philosophical and ethical questions about control and safety.

Curiosity

Exploration of different user friendly LLM interfaces in the market

- Text response generation(e.g. [ChatGPT](#), [Microsoft Bing](#), [Microsoft Copilot](#), [Perplexity](#))
- Image generation(e.g. [DALL-E3](#), [Midjourney](#))
- Audio generation ([AssemblyAI](#), [ElevenLabs](#))
- Video generation (Sora)
- Gaming (AlphaZero)
- Code generation (Github Co-pilot, Watson Code Assistant)
- Data generation and augmentation (DeepMind, AlphaFold)
- Virtual world (Avatars, Metaverse)

LLM tools showdown



GPT 4o1

ANTHROPIC

Sonnet 3.5



Gemini



Llama 3.2



Grok2



HUGGING FACE



Copilot

Curiosity: What Can LLMs Do for Us?



Prompt category	Content Generation & Writing Assistance	Summarization	Data Analysis & Code Generation	Research & Brainstorming
Prompt Usages	Report generation Email writing / tone refinement Proofreading Translation	Research report / article / earnings transcript summarization (image-text pairing) Email / chat summarization Information / keyword extraction	Excel formula generation VBA Macros, Python, and R code generation Data cleaning and formatting	Concepts / terms explanation Research paper title ideas generation Product name generation Discussion partner

Adaptability: How Can We Be Synergistic?



Skills that are resilient	Design thinking	Creative thinking	Critical thinking	Technical & Math literacy
Communication	Empathize, Define, Ideate, Prototype, Test, Iterate	Novel idea generation	Unstructured problem solving	Data literacy
Story-telling		Associative thinking	Criticizing and questioning information	Coding proficiency
Adaptability		Emotional resonance	Decision making	Debugging skills
Deepening your social ability				System architecture understanding
				Basic mathematics

Resilient Skills

PROMPT: Act as an immunologist. Explain in medium details (150 words), which immunoglobulin (Ig) is sent to the fetus from the placenta and breast milk.



Answer

As an immunologist, I can explain that the primary immunoglobulin (Ig) transferred from mother to fetus through the placenta is IgG ① ②. This transfer begins during the first trimester and increases significantly in the third trimester, with fetal IgG levels often exceeding maternal levels by 37-40 weeks of gestation ①.

After birth, breast milk provides a different array of antibodies. The predominant Ig in breast milk is secretory IgA (SIgA), which coats the infant's mucosal surfaces and provides crucial protection against pathogens ④ ⑥. Breast milk also contains smaller amounts of IgM and IgG ⑥.

While IgG from placental transfer provides systemic immunity to the newborn, the IgA in breast milk primarily offers localized protection in the infant's gastrointestinal tract ④. It's important to note that unlike some animals, human infants do not absorb significant amounts of IgGs from breast milk into their bloodstream ⑥. Instead, these antibodies work locally in the gut to prevent pathogen entry and colonization.



Benefits and Risks of IgG Transplacental ...

· ncbi.nlm.nih.gov

Several studies reported that maternal antibodies transferred across the placenta during pregnancy can decrease infant immunologic response ...



View 5 more

Critical Thinking: Which Image Wasn't AI Generated?

A



B



C



D



Creative Thinking: Brainstorming Fun Ideas

Summarize the [insert pdf] paper so that 10-year-old can understand the science.

CASE REPORT | Open Access | CC BY SA

PK/PD model-informed dose selection for oncology phase I expansion: Case study based on PF-06939999, a PRMT5 inhibitor

Cen Guo, Kai H. Liao, Meng Li, I-Ming Wang, Naveed Shaik, Donghua Yin



Thought for 13 seconds ▾

Understanding How Medicines Work Inside Our Bodies

Scientists are like detectives trying to figure out how medicines help us feel better. To do this, they use special computer programs called **models**. These models are like maps that show where the medicine goes after we take it.

When you swallow a pill or get a shot, the medicine travels through your body. It goes into your blood, reaches different organs, and eventually leaves your body. Scientists want to know:

- How fast does the medicine get into the blood?
- How much of it reaches the parts of the body that need it?
- How long does it stay in the body?
- How does the body break it down and get rid of it?

By using these models, scientists can predict what happens to the medicine inside us. This helps doctors decide the **right amount of medicine** to give to people—enough to help them get better but not so much that it causes problems.

The paper talks about improving these models to make them even more accurate. Better models mean doctors can make safer and more effective medicines for everyone.

Speaker icon, download icon, thumbs up icon, thumbs down icon, refresh icon, close icon



Imagine you have a toy car, and you want to know how fast it can go and how you can make it go even faster or slower. Scientists do something similar but with medicines instead of toy cars. They want to know how much medicine should be given to people to help them get better without making them feel too sick.

This paper is like a guidebook for those scientists. Here's what it says in simple words:

1. **The Medicine Journey:** When you take medicine, it's like the medicine goes on a trip inside your body. It has to travel through your stomach, into your blood, and then to the part of your body that needs help.
2. **How Much Medicine?**: Just like you don't want to put too much fuel in your toy car because it might spill, doctors don't want to give too much medicine because it might make you feel bad. This paper talks about finding the perfect amount, not too little, not too much, but just right.
3. **Using Math and Computers:** Scientists use math problems (like the ones you do at school but a bit harder) and computers to guess the best amount of medicine. They make a kind of video game where the character is the medicine, and they see what happens if they change things like how fast the medicine is given or how it's spread in the body.
4. **Testing and Learning:** After they use their computers, they test the medicine in real life, but not just on anyone. They start with a few people to see if their math was right. If the medicine works well and doesn't make people too sick, they're on the right track!
5. **Why This Matters:** Imagine if your toy car could be made to go exactly where you want at the speed you want just by figuring out the right amount of fuel. That's what doctors want with medicine, to make sure it helps you without any "spills" (side effects).

So, this paper is like a treasure map for doctors and scientists. It helps them find the "X marks the spot" where the medicine does its best work without causing trouble.

This explanation avoids complex terms and focuses on the basic idea of optimizing drug dosages using models, which even at a simple level, conveys the purpose and process described in the scientific paper.

Communication

Supporting empathetic and concise communication

I want this email to be a non-obligatory invite for a virtual coffee break. Create a doodle for the possible days of repeat events. Let the participants tell me if they want to attend it.



Try Draft with Copilot

I want this email to be a non obligatory invite for a coffee break. Create a doodle link inset for the possible dates of the repeat events. Let the participants notify if they are available. Keep the tone of the email semi-formal

Adjust

Generate

Tone

- Casual
- Formal
- ✓ Direct
- Make it a poem

Length

- Short
- ✓ Medium
- Long

Hi all,

I would like to invite you to join me for a virtual coffee break to catch up and take a break from our busy schedules.

I have created a Doodle with possible dates and times for repeat events, and your participation, while not obligatory, would be greatly appreciated.

These informal sessions aim to foster a sense of community, provide a relaxed environment for sharing ideas, and offer a platform for us to connect on a personal level.

If you have any suggestions or ideas for making these coffee breaks more engaging, please feel free to share them.

Thank you for considering this invitation, and I hope to see you at our next virtual coffee break.

Design Thinking

Give me a draft to write the notes and take aways from a conference. Please output a word document as template



I've created the Word document for your conference notes and takeaways. You can download it from the link below:

[Download Conference Notes and Takeaways \[x\]](#)

Conference Notes Template

Conference Name

Location: _____

Date: _____

Organizer: _____

Keynote Speeches

Speaker Name: _____

Title of Speech: _____

Key Points: _____

Personal Reflections: _____

Breakout Sessions

Session Title: _____

Speaker/Panelists: _____

Key Topics Discussed: _____

Insights Gained: _____

Questions Raised: _____

Workshops

Workshop Title: _____

Facilitator: _____

Main Focus: _____

Skills Learned: _____

Application in Current Role/Industry: _____

Networking Opportunities

Contacts Made: _____

Potential Collaborations: _____

Vendor Booths/Exhibitors

Interesting Products/Services: _____

Responsible Use of GenAI

How to build trust in GenAI?

Intentionality

Transparency

Risk Management

Guidelines of Responsible Use

FASTER

Fair
Accountable
Secure
Transparent
Educated
Relevant

EU Guidelines

Government of Canada guidelines

Fair

Determine what was used as training data

Always review outputs; remove biased, non-inclusive or discriminatory content

Don't use Generative AI to make decisions that could legally or materially impact an individual

You must not represent that output was human-generated when it was not.

"OpenAI terms of Use"

You must not use any Output relating to a person for any purpose that could have a legal or material impact on that person, such as making credit, educational, employment, housing, insurance, legal, medical, or other important decisions about them.

Accountable

Accurate

Human oversight on outputs necessary

Do NOT use LLM as a search engine

NEVER use LLMs for skills you don't already possess

Legal

Consult your legal department

Ethical

Determine if training data was obtained legally and ethically

Edit/Improve on what you wrote, rather than generating from scratch

Be inclusive to people using GenAI

September 04 2024

Do LLMs Exhibit Human-like Response Biases? A Case Study in Survey Design

Lindia Tjatja, Valerie Chen, Tongshuang Wu, Ameet Talwalkar, Graham Neubig



Author and Article Information

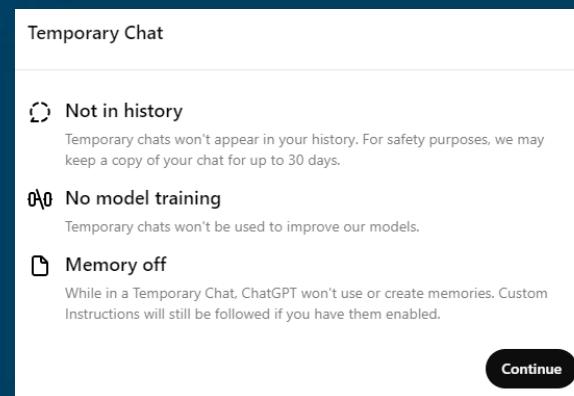
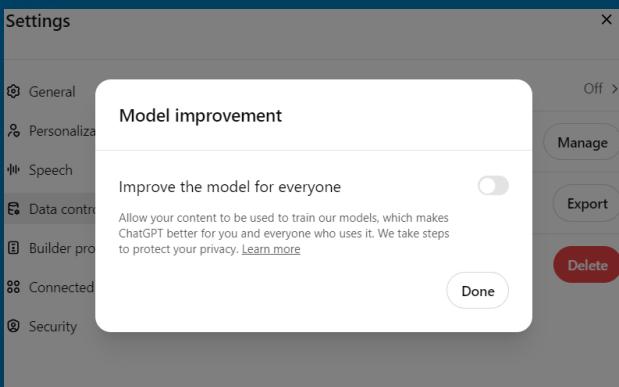
Transactions of the Association for Computational Linguistics (2024) 12: 1011–1026.

https://doi.org/10.1162/tacl_a_00685 Article history 

Secure

Actually read the terms of service

OPT-OUT of using your conversations to train future models, where possible.



Transparent

Keep a detailed record about usage
Inform users when Gen-AI is used

Exact tool used and version

How it works

Why it's being used

To what extent content was reviewed or modified



Authors should disclose in their manuscript the use of generative AI and AI-assisted technologies and a statement will appear in the published work. Declaring the use of these technologies supports transparency and trust between authors, readers, reviewers, editors, and contributors and facilitates compliance with the terms of use of the relevant tool or technology. This policy is intended to cover new content creation only (i.e., new works or new content or chapters added to a revised work). Generative AI and AI-assisted technologies should not be used on previously published material.

Wiley Author Services

Generative Artificial Intelligence tools (GenAI)—such as ChatGPT and others based on large language models (LLMs)—can increase productivity and foster innovation if used appropriately in a safe, ethical and secure manner. STM has general guidance for all stakeholders in scholarly publishing which addresses the role of generative AI technologies. If an author has used a GenAI tool to develop any portion of a manuscript, its use must be described transparently and in detail, in the Methods section (or via a disclosure or within the Acknowledgements section, as applicable). The author is fully responsible for the accuracy of any information provided by the tool and for correctly referencing any supporting work on which that information depends. GenAI tools must not be used to create, alter or manipulate original research data and results. Tools that are used to improve spelling, grammar, and general editing are not included in the scope of these guidelines. The final decision about whether use of a GenAI tool is appropriate or permissible in the circumstances of a submitted manuscript or a published article lies with the journal's editor or other party responsible for the publication's editorial policy.

Does NIH allow the use of generative AI to help write grant applications and/or R&D contract proposals?

NIH guidance only prohibits use of generative AI technology in the peer review of grant applications and contract proposals. In our [Open Mike blog](#) we indicate that if you use an AI tool to help write your application, you do so at your own risk. "This is because when we receive a grant application, it is our understanding that it is the original idea proposed by the institution and their affiliated research team. Using AI tools may introduce several concerns related to research misconduct, like including plagiarized text from someone else's work or fabricated citations. If we identify plagiarized, falsified, or fabricated information in a grant write-up, we will take appropriate actions to address the non-compliance."

Educated

Knowledgeable about GenAI tools before using them

Writing effective prompts

Keeping yourself up-to-date on the newest advancements and challenges

Relevant

Always ask yourself:

Is GenAI the best tool for the task?
Or is it just...
Overkill, time-consuming and messy

GenAI-Harmonist Skills

Prompt Engineering

Communication

Leadership

Critical Thinking

Code-literacy

Systems Design Concepts

Prompt Engineering

PROMPT: A prompt refers to the input to LLMs that guides LLMs to generate high-quality and relevant output

How to write an effective prompt?

C_A_R_E
[Context] [Action] [Result] [Elaboration]

Context

I am delivering a presentation on "Prompting Best Practices".

Action

Create an engaging outline that covers key aspects of prompting best practices.

Result

The outline should include sections on:

1. Introduction to Prompting
2. Best Practices for Prompting
3. Examples of Basic and Enhanced Prompts

Elaboration

The slide deck should also contain content for a 25-minute presentation and ensure that the content is relevant for an audience with mixed levels of experience in AI.

Prompt Engineering

BASIC PROMPT: How many R are there in strawberry



The word "strawberry" contains two "R"s.



There are **two** 'R's in the word "Strawberry."

CARE PROMPT: Find the occurrence of the character 'r' in the word 'Strawberry'



Analyzed ^

python

```
# Let's calculate the occurrence of 'r' in the word 'strawberry'
word = "strawberry"
r_count = word.lower().count('r')
r_count
```

Result

3

The letter "r" occurs 3 times in the word "Strawberry." [2-]



Here are the occurrences of the character 'r' in the word "Strawberry":

1. rawberry - Here's the first 'r'
2. Strawberry - Here's the second 'r'
3. Strawberry - Here's the third 'r'

So, in "Strawberry," there are **three** 'r's.

Critical Thinking: Technical Assistance

PROMPT: I have an excel of time series. I want to do Area under the curve calculation in excel of this time series. Provide me with the formula to do it



To calculate the Area Under the Curve (AUC) for a time series in Excel, you can use the trapezoidal rule. Assuming you have time values in column A and corresponding data values in column B, you can use the following steps:

1. Calculate the differences in time (Δt).
2. Calculate the average of consecutive data points.
3. Multiply these two results to get the area of each trapezoid.
4. Sum these areas to get the total AUC.

Here is the formula to calculate the AUC:

Assuming your time values are in column A (A2:A10) and your data values are in column B (B2:B10):

fx

```
=SUMPRODUCT((A3:A10 - A2:A9), (B3:B10 + B2:B9) / 2)
```

This formula calculates the area under the curve using the trapezoidal rule. Adjust the ranges (A2:A10 and B2:B10) according to your actual data.

AI-generated content may be incorrect



Suppose your data is as follows:

A (Time)	B (Value)
0	5
1	7
2	6
3	8

- Using the single formula approach:

```
excel
=SUMPRODUCT((A3:A5 - A2:A4), (B3:B5 + B2:B4)/2)
Copy code
```

This will calculate:

$$\left(1 \times \frac{7+6}{2}\right) + \left(1 \times \frac{6+8}{2}\right) = 6.5 + 7 = 13.5$$

- The total AUC is 13.5.

Leadership

PROMPT: What is the pKa of ketoconazole. Search the web and give me references from Pubmed

references are made up



Thought for 30 seconds ▾

The pKa of Ketoconazole is approximately 6.51, attributed to the protonation of its imidazole ring. This pKa value significantly influences its solubility and absorption, as Ketoconazole is a weak base with pH-dependent solubility.

Here are three PubMed references that discuss the physicochemical properties and solubility behavior of Ketoconazole:

1. Devarakonda B, Hill RA, de Villiers MM.

Effect of pH and ionic strength on ketoconazole supersaturation and precipitation.

AAPS PharmSciTech. 2003;4(2)

PMID: [12916911](#)

2. Singh S, Rehder S, Diez M, Turnbull N, Forbes RT.

The effect of pH and ionic strength of dissolution media on the solubility and dissolution rate of ketoconazole.

Pharm Dev Technol. 2011 Dec;16(6):612-21.

PMID: [20919969](#)

3. Tønnesen HH, Karlsen J.

Studies on ketoconazole: solubility and stability.

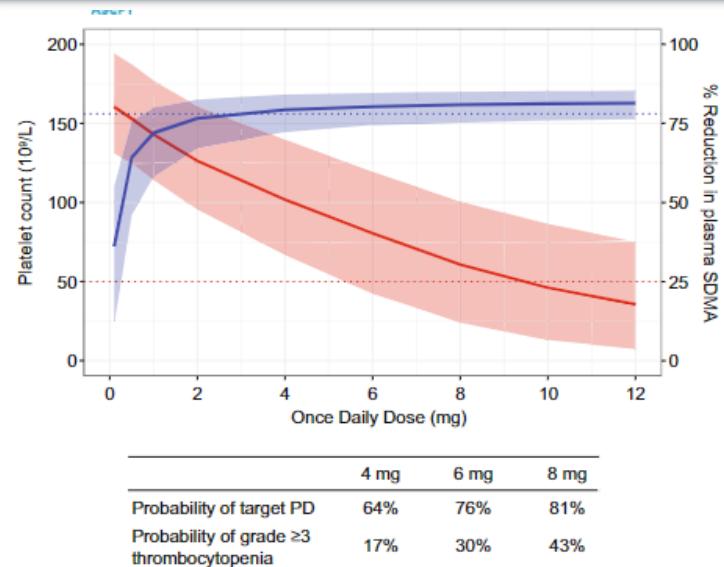
Int J Pharm. 1986 Dec;33(1-3):351-5.

PMID: [2879755](#)

These studies provide insights into how the pKa influences Ketoconazole's solubility and stability under different pH conditions.

Critical Thinking: Discussion Partner

CARE PROMPT: Act as a regulatory scientist who has worked on submitting Clinical Pharmacology reports to medical regulatory agencies. Based on the above figure, from the attached publication in pdf can you help me select a phase 2/3 dose for further investigation and provide justifications. Constrain the response to 300 words



Based on the presented data, the recommended dose for Phase 2/3 investigation is **6 mg once daily (q.d.)**. This selection is based on an optimal balance between efficacy, measured by reduction in plasma symmetrical dimethyl-arginine (SDMA), and safety, specifically the risk of thrombocytopenia.

Justification:

- Efficacy:** The 6 mg q.d. dose achieves a **76% probability of hitting the target pharmacodynamic (PD) response**, which corresponds to a 78% reduction in plasma SDMA. This reduction is associated with substantial PRMT5 inhibition, indicating effective target engagement without further substantial benefits at higher doses. The efficacy only marginally improves to 81% at 8 mg q.d., suggesting that 6 mg q.d. is near the efficacy plateau.
- Safety:** Thrombocytopenia is the main adverse event of concern. The probability of experiencing grade ≥ 3 thrombocytopenia at 6 mg q.d. is 30%, which is within an acceptable range for oncology patients with advanced disease. At 8 mg q.d., the thrombocytopenia risk increases significantly to 43%, posing an unacceptable safety risk relative to the slight improvement in efficacy.
- Risk-Benefit Balance:** The 6 mg q.d. dose offers a favorable risk-benefit profile, achieving effective PRMT5 inhibition while maintaining a manageable safety risk. A lower dose, such as 4 mg q.d., carries the risk that approximately one-third of patients may not reach the PD target, thus potentially underdosing the population.

Considering both efficacy and safety data from the simulations and observed profiles, **6 mg q.d.** is recommended as the optimal dose for Phase 2/3 investigation to maximize patient benefit while minimizing adverse outcomes ([CPT Pharmacom Syst Ph...](#)).

Prompt Engineering: Best Practices



Be clear, concise and logical



Define persona (scientist), tone (eg . profession) or length



Specify an output format (text, bullet points, codes)



Ask follow up questions: curiosity



Context length and relevancy

Learning Resources:

[Anthropic prompt library](#)

[IBM prompt engineering](#)

[Hugging face prompting guide](#)

AI Tools for Scientific Workflow

Writing and text summarization



Teaching content



beautiful.ai

Code assist



GitHub Copilot



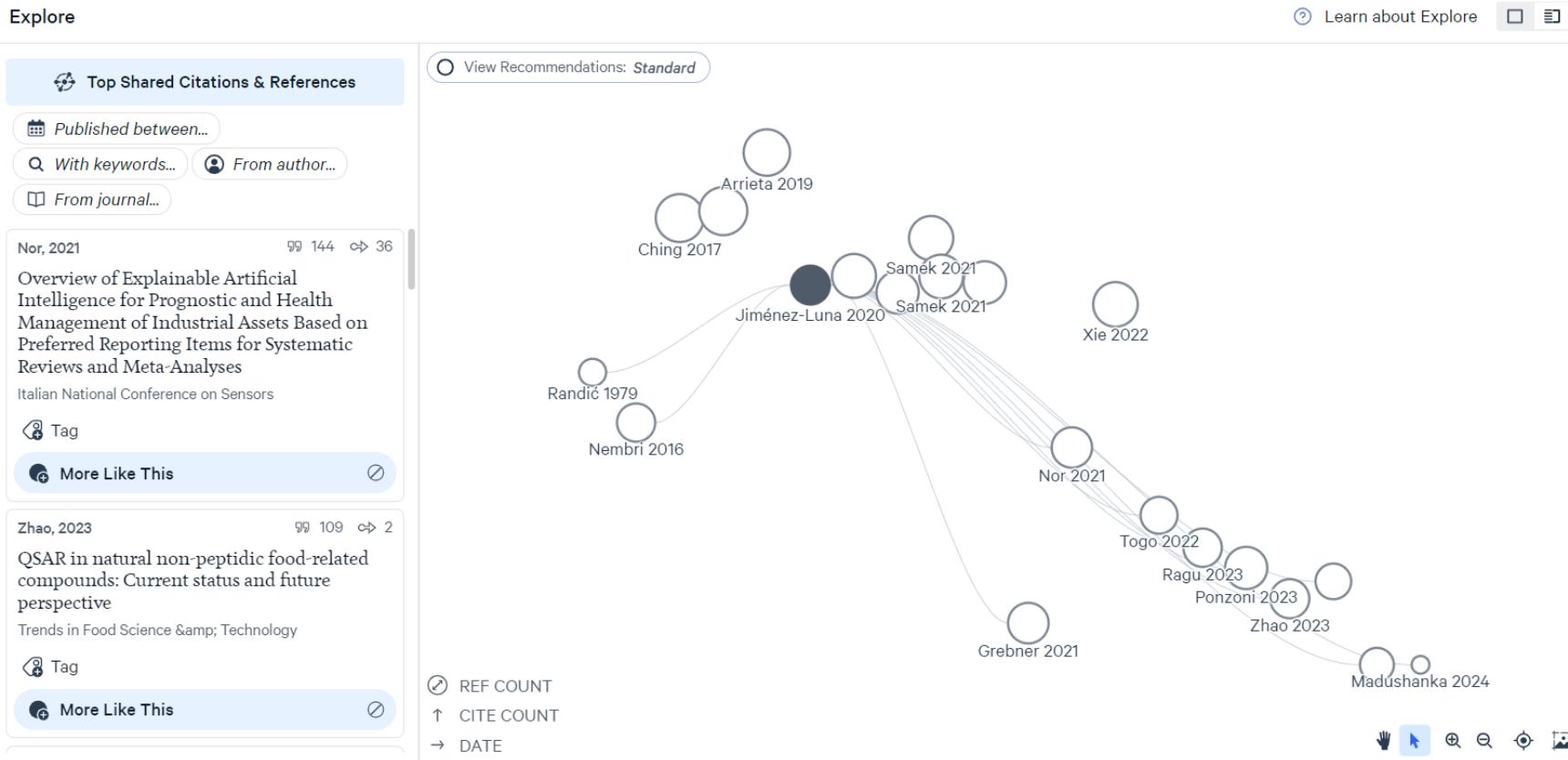
C[x]

Research assistant for biology



Knowledge Graph for Papers

Litmaps



Talk to Pubmed

The screenshot shows a web-based application titled "Generate a report with GEMA". At the top, there's a navigation bar with links for "Report", "Target Enabler", "Trials", "Chemistry", "Companies", and an email address "aj@amasstech.onmicrosoft.com". Below the navigation bar, the main area is titled "Generate a report with GEMA". It features a "Configure GEMA" section with a "Databases to query" field containing "#patents", "#pubmed", and "#research". There are four tabs below this: "Query" (which is selected), "Topics", "Generation", and "Report". In the "Ask GEMA to report on this subject" section, there's a text input field with the query: "Can you give me an overview of atrial fibrillation. Focus on existing treatments and their mode of action.". Below this, there are two checkboxes: "Investment memorandum mode" and "Target assessment mode". A "Report generation options" dropdown menu is open. At the bottom left, there's a "Generate topics" button. On the left side of the screen, there are sections for "Past generated reports by GEMA", "Today's conversations" (listing "Current Treatments and Emerging Perspectives in Heart Failure with Preserved Ejection Fraction (HFpEF)" with a close button), and "Previous 7 days' conversations" (listing "Investigating PI3K Alpha-Specific Inhibitors: A Targeted Approach" with a close button and a circular profile picture of a man).

GenAI-Maximalist Skills

Craft LLM prompts

Experience

Software DevOps

Complex AI Tasks

Collaboration

Systems Level Design

Craft LLM Prompts: Shiny Assistant

R Python Concise < Share

app.R

```
theme_minimal() +  
  labs(title = "Bar Graph",  
       subtitle = "Sankey Diagram")  
}  
}  
  
shinyApp(ui, server)
```

Run app →

Here are the key changes I've made to remove date-related functionality:

1. Removed the `library(lubridate)` import as it's no longer needed.
2. Removed the "Generate Time Series Graph" button from the UI.
3. Removed the Time Series Graph card from the layout.
4. Removed the date column selection from the column mapping UI.
5. Updated the CSV template download to not include a date column.
6. Removed the time series plot generation code from the server function.

This modified version of the app now focuses solely on the Sankey diagram and bar graph visualizations, without any date-related components. Users can still upload CSV files, map columns for source, target, value, and category, and generate Sankey diagrams and bar graphs based on their data.

Enter a message... ↑

Data Visualization App

Sankey Diagram

Bar Graph

Upload CSV File

Browse... Zolpide Upload complete

Download CSV Template

Source Column RECORDC

Target Column VAR

Value Column TIME

Category Column UNIT

Generate Sankey Diagram

Generate Bar Graph

Bar Graph

Total Value

1e+05

5e+04

0e+00

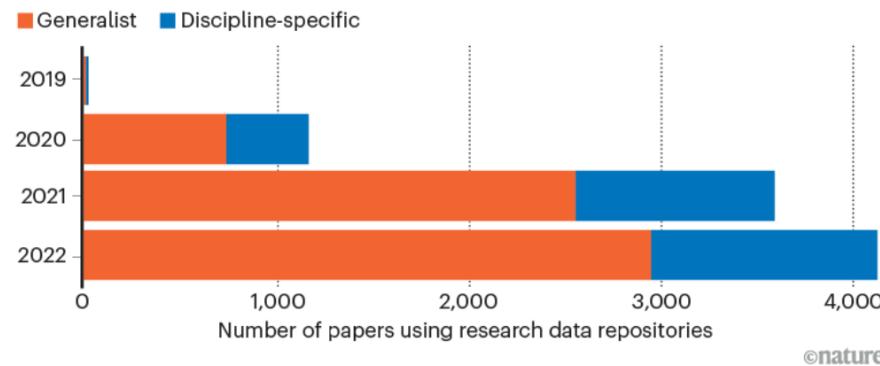
(mg) (ng/mL) (ug/mL)

Category

Data is the New Oil: LLMs are the Combustion Engines

RISE IN DATA ARCHIVING

Researchers are increasingly depositing data in repositories to widen access, but mostly in generalist rather than discipline-specific ones that offer curation. This can be seen in the top 15 repositories used in primary-research papers published in American Geophysical Union journals between 2019 and 2022.



FAIR and OPEN Data for AI and ML:

- Publicly available datasets for benchmark training data are lacking in our field
- Available data sources, types, formats are not interoperable
- Available data are not analysis ready (Labels lacking, structure unsuitable)
- Pretrained models that can be adapted for biological observations are lacking

A call for an industry-led initiative
to critically assess machine learning
for real-world drug discovery
<https://doi.org/10.1038/s42256-024-00911-w>

Kindly provided by Shelly Stall

Hanson et al. Nature 623, 28-31 (2023)

doi: <https://doi.org/10.1038/d41586-023-03316-8>

Maskey, M., Alemohammad, H., Murphy, K. J., and Ramachandran, R. (2020), Advancing AI for Earth science: A data systems perspective, Eos, 101, <https://doi.org/10.1029/2020EO151245>. Published on 06 November 2020

Explainable AI (XAI): Small Steps

As the LLMs have gotten progressively larger, they have begun to surpass some repeatable human performance, however they show plagiarism, bias, are black box like and hallucinate thus area of active research

BioGPT and human annotator have comparable performance in biomedical research test

Selected performances on PubMedQA, which tests biomedical language processing

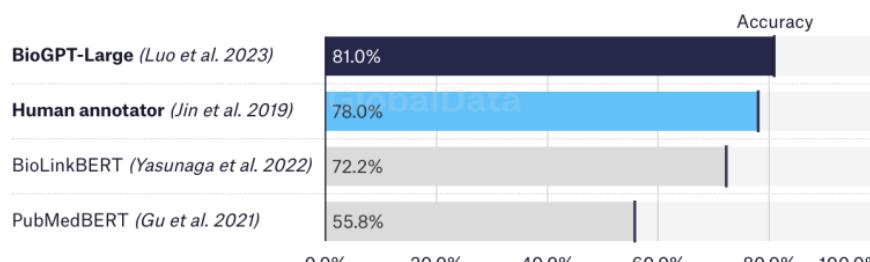


Chart: GlobalData • Source: PubMedQA

Table 2. Application of specialized LLMs in different fields of medical science

Sl. no.	LLM	Year of release	Remarks	Reference
1.	PMC-LLaMA	2023	PMC-LLaMA, an open-source language model that is developed by refinement of an open-source language model on a total of 4.8 million biomedical academic papers for added injecting medical knowledge, improving its capability in the medical domain.	Wu et al. ⁹⁶
2.	ClinicalCamel	2023	It is an open LLM obviously tailored for clinical research. Fine-tuned from LLaMA-2 using QLoRA, Clinical Camel achieves state-of-the-art performance across medical benchmarks among openly available medical LLMs. Leveraging efficient single-GPU training.	Toma et al. ⁹⁷
3.	MedAlpaca	2023	MedAlpaca was developed by instruction fine-tuning of the LLaMA 13B and 7B models on Medical Meadow data. It is also the assembly of reformatted instruction-response pairs with datasets for medical NLP tasks and data derived from various internet sources.	Han et al. ⁹⁸
4.	BioGPT	2023	BioGPT is a domain-specific GPT language model for biomedical text generation and mining. BioGPT follows the transformer language model backbone, and is pre-trained on 15M PubMed abstracts from scratch.	Luo et al. ⁹⁹
5.	BioMedLM	2022	BioMedLM is based on a HuggingFace GPT model (decoder-only transformer) with 2.7B parameters and a maximum context length of 1,024 tokens. It also uses a custom biomedical tokenizer trained on PubMed abstracts with a vocabulary size of 28,896.	Karkera et al. ¹⁰⁰
6.	Med-PaLM2	2022	Med-PaLM is a large language model (LLM) designed to provide high quality answers to medical questions. It is also available to Google Cloud customers, who are able to explore a range of applications, from basic tasks to complex workflows.	Luo et al. ¹⁰¹
7.	ChatDoctor	2023	It has been aligned to the medical domain and evaluated using medical exams, medical research, and consumer queries.	Li et al. ¹⁰²
			This is a specified language model with improved accuracy in medical advice refining the large language model meta-AI (LLaMA) by a large dataset of patient-doctor dialogs obtained from a widely used online medical consultation platform.	

Bhattacharya, M., Pal, S., Chatterjee, S., Lee, S., & Chakraborty, C. (2024). Large Language Model (LLM) to Multimodal Large Language Model (MLLM): A journey to shape the biological macromolecules to biological sciences and medicine. *Molecular Therapy - Nucleic Acids*.

Human-in-the-Loop: Where We Stand

HITL integrates human input and expertise into the lifecycle of ML and AI systems. This includes approaches like:

Active learning (AL): humans label data at the systems request

Interactive machine learning (IML): Users interact frequently with the learning system, providing information incrementally to guide the learning process.

Machine teaching (MT): Human experts define and transfer specific knowledge to the machine learning model, giving them more direct control over what is learned.

Human-in-the-loop assisted de novo molecular design

Iiris Sundin , Alexey Voronov , Haoping Xiao, Kostas Papadopoulos, Esben Jannik Bjerrum, Markus Heinonen, Atanas Patronov, Samuel Kaski & Ola Engkvist

Journal of Cheminformatics 14, Article number: 86 (2022) | [Cite this article](#)

5529 Accesses | 13 Citations | 25 Altmetric | [Metrics](#)

An algorithmic framework for synthetic cost-aware decision making in molecular design

Jenna C. Fromer & Connor W. Coley 

Nature Computational Science 4, 440–450 (2024) | [Cite this article](#)

The AI-driven Drug Design (AIDD) platform: an interactive multi-parameter optimization system integrating molecular evolution with physiologically based pharmacokinetic simulations

Published: 19 March 2024

Volume 38, article number 14, (2024) | [Cite this article](#)

Jeremy Jones , Robert D. Clark, Michael S. Lawless, David W. Miller & Marvin Waldman

Clinical Pharmacology & Therapeutics

State of the Art Review

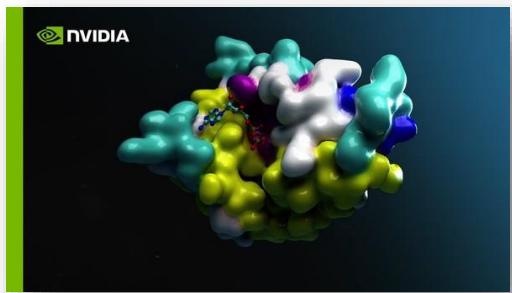
Current Status and Future Directions: The Application of Artificial Intelligence/Machine Learning for Precision Medicine

Kunal Naik, Rahul K. Goyal, Luca Foschini, Choi Wai Chak, Christian Thielscher, Hao Zhu, James Lu, Joseph Lehár, Michael A. Pacanoswki, Nadia Terranova, Neha Mehta, Niklas Korsbo, Tala Fakhouri, Qi Liu, Joga Rao Gobburu  ... See fewer authors ^

Bhattacharya, M., Pal, S., Chatterjee, S., Lee, S., & Chakraborty, C. (2024). Large Language Model (LLM) to Multimodal Large Language Model (MLLM): A journey to shape the biological macromolecules to biological sciences and medicine. *Molecular Therapy - Nucleic Acids*.

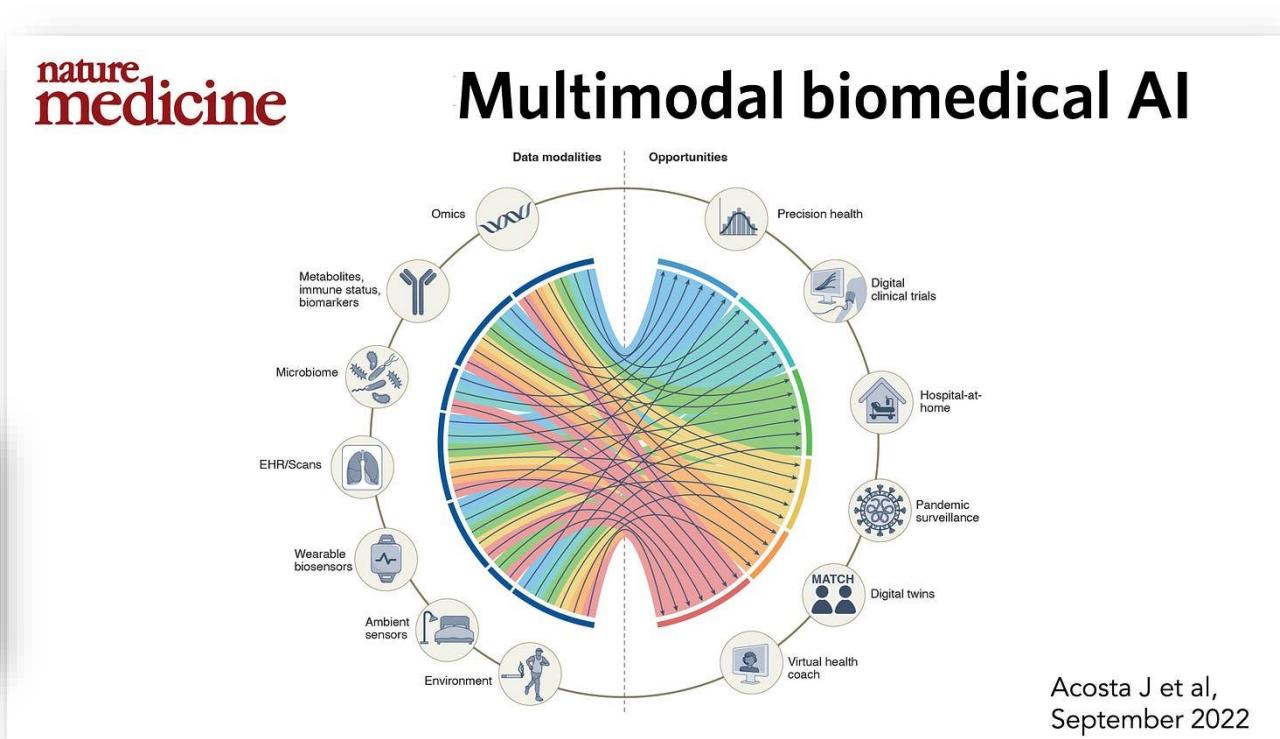
Explainable LLMs: Multi-Modal Information

AI in medicine and drug discovery is a single mode story- But over time that narrowness and constraint may well be alleviated.



Get Started With NVIDIA BioNeMo

NVIDIA BioNeMo™ is an AI platform for drug discovery that simplifies and accelerates the building and training of models using your own data and scaling the deployment of models for drug discovery applications.



<https://nvidianews.nvidia.com/news/nvidia-unveils-large-language-models-and-generative-ai-services-to-advance-life-sciences-r-d>

Acosta, J.N., Falcone, G.J., Rajpurkar, P. et al. Multimodal biomedical AI. Nat Med 28, 1773–1784 (2022). <https://doi.org/10.1038/s41591-022-01981-2>

Key Skills: Context is the Endgame

Understand Market Needs

Attend conferences, webinars, and network with AI experts to identify skill gaps and align expertise.

Continuous Learning

Take courses on LLM methodologies aligned with your interests.

Use AI Strategically

Focus on solving clinical and medical challenges without adding complexity.

Understand Data Quality

Evaluate data quality, regulatory, and ethical aspects.

Select Tools and Techniques

Identify suitable algorithms, frameworks, and relevant tools (e.g., Python/R) for healthcare applications.

Collaborate and Iterate

Work with interdisciplinary teams, test prototypes, and refine solutions for better usability. Balance critical thinking, adaptability, communication, curiosity, and technical proficiency

Learn APIs

Understand wrappers and APIs to enhance system integration.

Questions & Contact



Priyata Kalra, Ph.D.

Senior Scientist

Physiologically Based Pharmacokinetics Solutions

priyata.kalra@simulations-plus.com

Check out my
generative art:



Seeing truth (blog):



Try out my GPTs:



Dream Weather Artisan

Asian-inspired dreamy weather visuals with literary facts and spiritual depth.
By Priyata Kalra ☺ 20+



PBPK First-in-Human Guide

By Priyata Kalra ☺

A PBPK assistant for first-in-human drug prediction brainstorming.

 **SimulationsPlus**



 **Pro-ficiency**
a Simulations Plus company



info@simulations-plus.com



+1-661-723-7723

Connect with us:    

The Allies of this talk

John DiBella
Haiying Zhou
Christina Hendrickson
Vladimir Chupakhin
Rafal Bachorz
Manas Shah
Andrea Ceja
Nash DelCamp
Melissa Van Bussel (Health Canada)
Thomas Hartung (JHU)
Jurgen Pahle (University of Heidelberg)
Shelly Stall (NASA)
Thomas Leuchtefeld (inSilica)

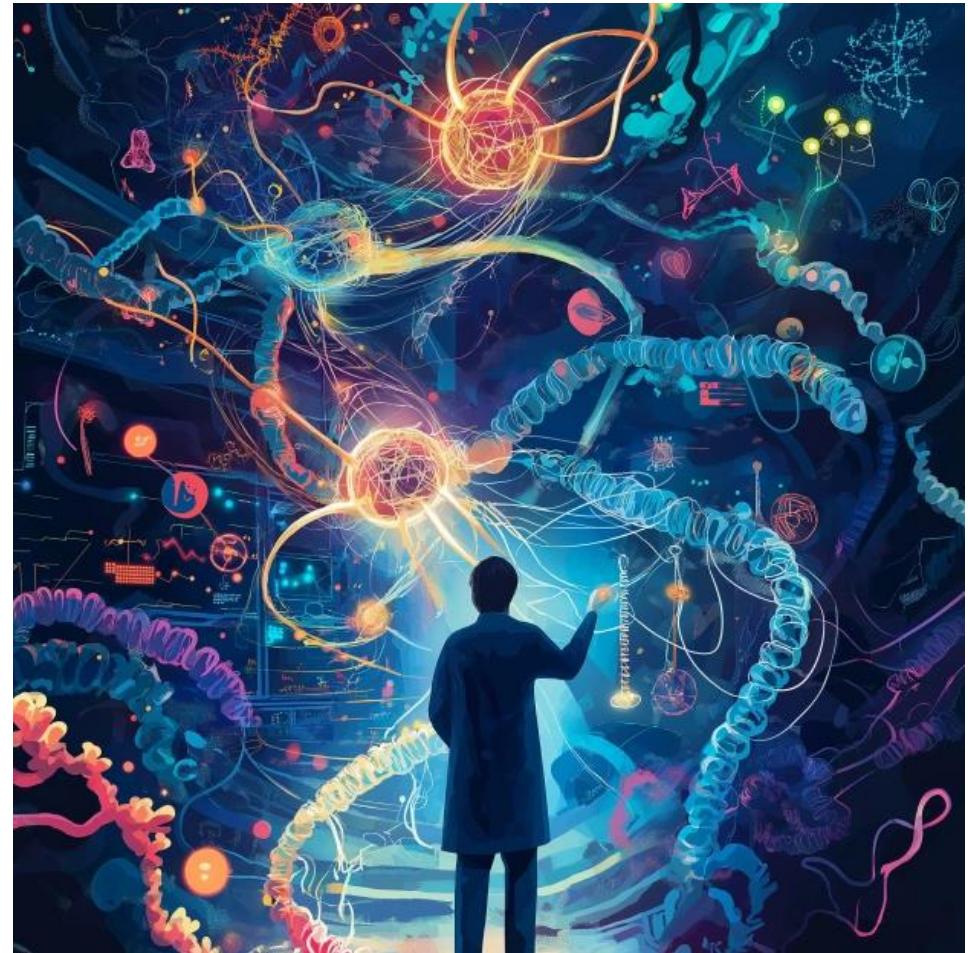


Image: Midjourney

Agenda: Skills for Future of Work?

Humanics

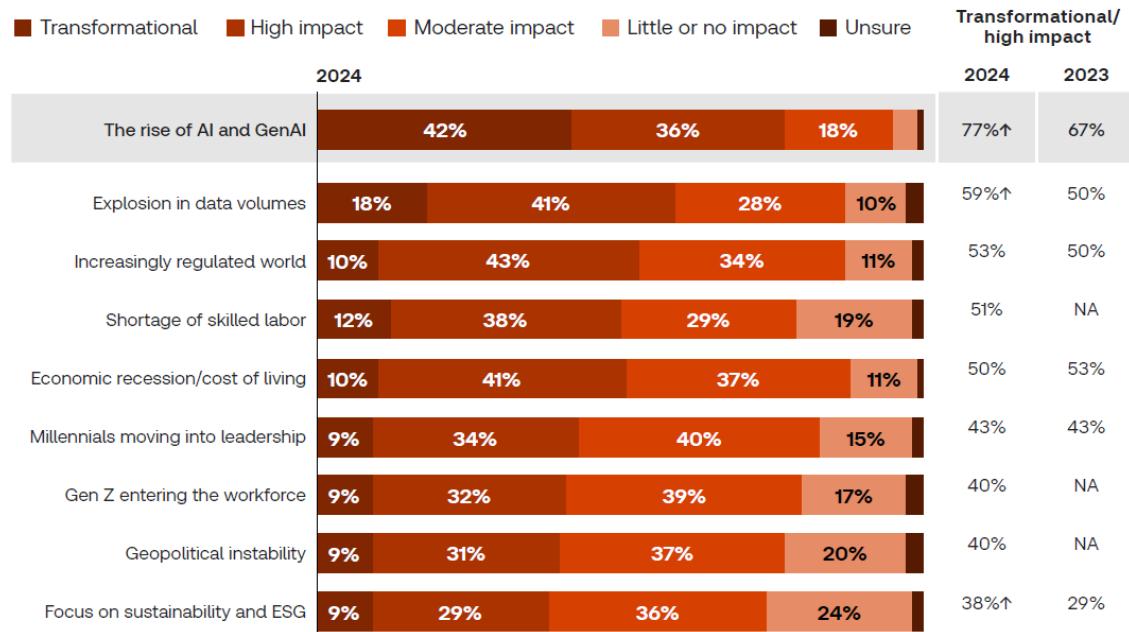
Nurture human nature and affair.

Knowledge

Increase knowledge on the edge of rare problems.

The Rise of GenAI

To what extent do you think the following trends will impact your profession over the next five years?



Source: Thomson Reuters 2024

Generative AI could:

- Produce **\$60 billion to \$110 billion** in annual revenue across the pharmaceutical industry value chain
- Impact **10% of the tasks** carried out daily by 80% of the workers

Human side: 70 percent of digital transformations fail not because of technical issues but because leaders ignored the importance of managing change.

Humanics

Critical Thinking

Analyze problems from multiple angles.

Communication

Articulate ideas clearly to diverse audiences.

Adaptability

Learn to leverage AI for greater productivity.

Curiosity

Embrace continuous learning and exploration.

Experience

Reimagining our education and empathy.

Knowledge

AI Basics

Understand foundational AI concepts and capabilities.

Prompt Engineering

Craft effective prompts to maximize AI output.

AI Tool Familiarity

Gain acquaintance with emerging AI applications

Tech Literacy

Stay current with evolving digital technologies.

Math

Numeracy with statistics and basic math.

Desire to Learn- how I do it?

Permissionless Apprentice

Dive in and learn by doing, then build theory.

Learning Resources

Leverage online courses, podcasts, and social networks.

Learn in Public

Have a habit of creating.

Learn for necessity

Top-down approach for busy work

Creative thinking: training/teaching

AI tool for synthesizing new audio/ avatars for training Teaching and training content generation will be changing

ElevenLabs

English



German



French



Chinese



Play buttons to hear my synthetic voice used to create multilingual educational content on Ozempic

Craft LLM Prompts: 4o canvas

design elements. Here is an updated CSS file that you can add:

Website Styles

```
1  /* styles.css */
2
3  body {
4      font-family: 'Roboto', sans-serif;
5      margin: 0;
6      padding: 0;
7      box-sizing: border-box;
8      background-color: #f4f4f4;
9      color: #333;
10 }
11
12 header {
13     background: #002a4e;
14     color: #fff;
15     padding: 20px 0;
16 }
17
18 .navbar .container {
19     display: flex;
20     align-items: center;
21     justify-content: space-between;
22 }
23
24 .logo {
25     font-size: 1.8em;
26     font-weight: bold;
27 }
28
29 nav ul {
30     list-style: none;
31     display: flex;
32     gap: 20px;
33 }
34
35 nav ul li a {
36     color: #fff;
37     text-decoration: none;
38     font-weight: 500;
39     transition: color 0.3s;
40 }
41
42 nav ul li a:hover {
43     color: #ffd700;
44 }
```

I've added a CSS file to style your HTML similar to the Simulations Plus website. It includes a gradient banner, modern typography, hexagonal images, and a cookie banner. Let me know if you need any further modifications or adjustments!

give me the updated files as .html and .css

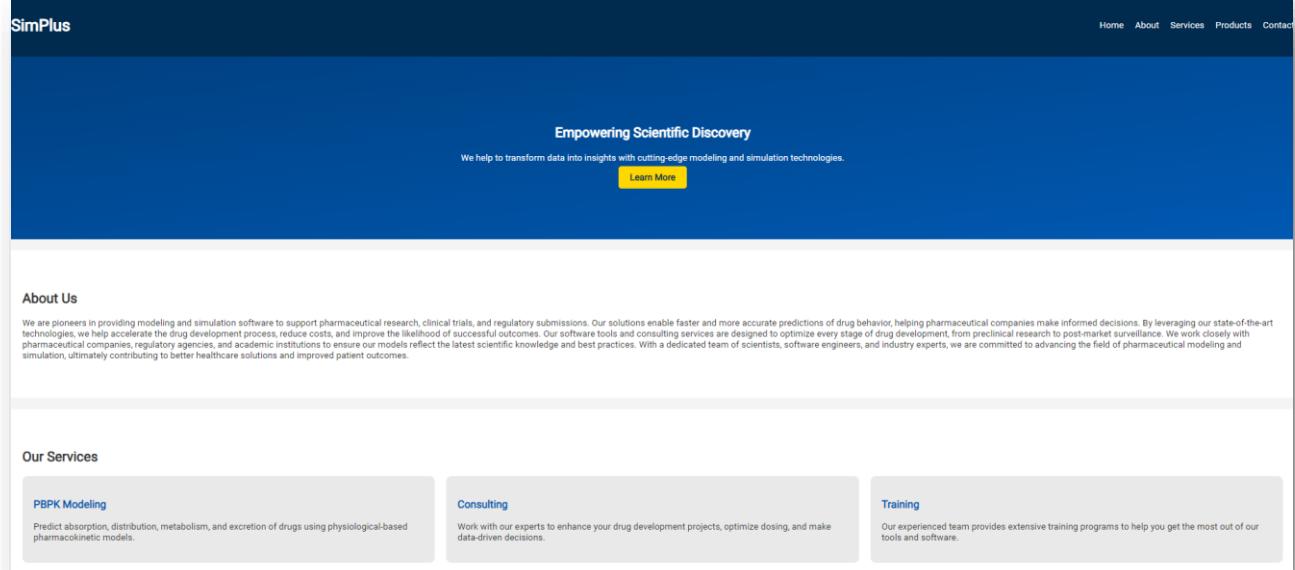
I've saved the updated HTML and CSS files for you. You can download them using the links below:

- index.html
- styles.css

Let me know if you need further adjustments or additional features! [\[?\]](#)

[Download](#) [Download](#) [Download](#) [Download](#)

Message ChatGPT [\[?\]](#)



References

- Add your references. This slide is optional.