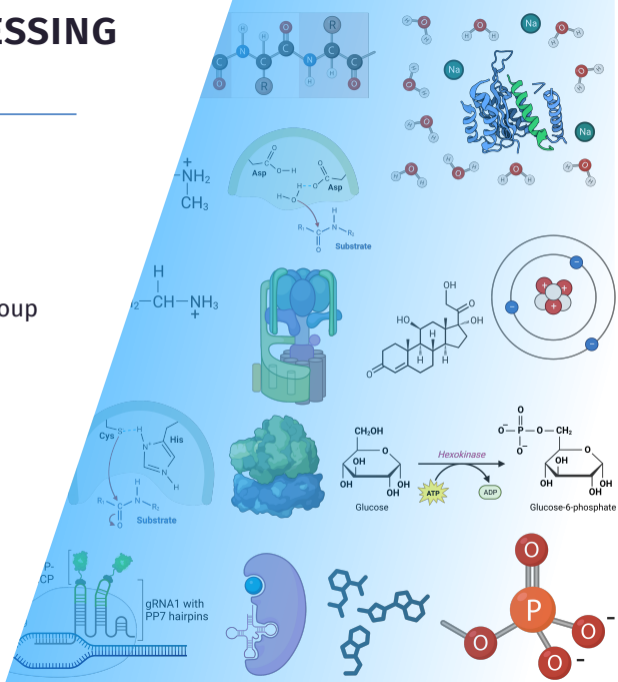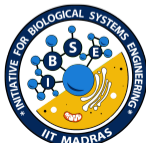# NATURAL LANGUAGE PROCESSING
For Biological Applications

Roshan M S B
bt21d401@smail.iitm.ac.in

BioSystems Engineering and ConTrol (BiSECt) Group
IIT Madras, India

Saturday 9th September, 2023

Agenda

1. What is Natural Language Processing?

2. Why do we Natural Language Processing?

3. Understanding Biological Networks

4. NLP for Biology

BEFORE WE MOVE ON..

How to use ChatGPT as a learning tool

Rather than weaken student effort, artificial intelligence can help prepare students for the real world by encouraging critical thinking—with a few caveats. Here's advice from psychology instructors about how to use ChatGPT and other AI technology

By Ashley Abramson   Last updated: June 1, 2023   Date created: March 27, 2023
Vol. 54 No. 4
Print version: page 67

Editorial ➤ Science. 2023 Jan 27;379(6630):313. doi: 10.1126/science.adg7879.
Epub 2023 Jan 26.

ChatGPT is fun, but not an author

H Holden Thorp [1]

Affiliations + expand
PMID: 36701446   DOI: 10.1126/science.adg7879

How Does ChatGPT Perform on the United States Medical Licensing Examination? The Implications of Large Language Models for Medical Education and Knowledge Assessment

Aidan Gilson [1] [2], Conrad W Safranek [1], Thomas Huang [2], Vimig Socrates [1] [3], Ling Chi [1], Richard Andrew Taylor [# ?] [2], David Chartash [#] [4]

Affiliations + expand
PMID: 36753318   PMCID: PMC9947764   DOI: 10.2196/45312

THE INDUSTRY

Finally, an A.I. Chatbot That Reliably Passes "the Nazi Test"

OpenAI's new ChatGPT is scary-good, crazy-fun, and—so far—not particularly evil.

BY ALEX KANTROWITZ   DEC 02, 2022 · 11:30 AM

Editorial ➤ Cureus. 2023 Feb 19;15(2):e35179. doi: 10.7759/cureus.35179.
eCollection 2023 Feb.

Artificial Hallucinations in ChatGPT: Implications in Scientific Writing

Hussam Alkaissi [1] [2] [3], Samy I McFarlane [3]

Affiliations + expand
PMID: 36811129   PMCID: PMC9939079   DOI: 10.7759/cureus.35179

ChatGPT reaches 100 million users two months after launch

Unprecedented take-up may make AI chatbot the fastest-growing consumer internet app ever, analysts say

Review ➤ Healthcare (Basel). 2023 Mar 19;11(6):887. doi: 10.3390/healthcare11060887.

ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns

Malik Sallam [1] [2]

Affiliations + expand
PMID: 36981544   PMCID: PMC10048148   DOI: 10.3390/healthcare11060887

COMMENTARY

ChatGPT: Educational friend or foe?

Kathy Hirsh-Pasek, Elias Blinkoff
January 9, 2023
Home / Innovation / Artificial Intelligence

How does ChatGPT actually work?

Updated: We take a deep dive into the inner workings of the wildly popular AI chatbot, ChatGPT. If you want to know how its generative AI magic happens, read on.

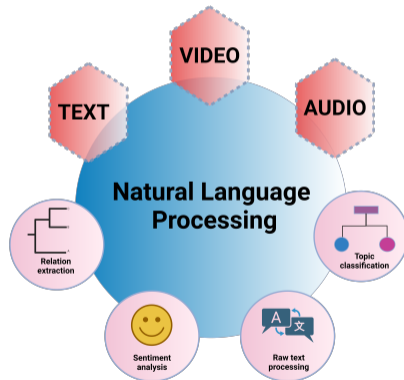Written by **David Gewirtz**, Senior Contributing Editor on July 13, 2023
Reviewed by **Alyson Windsor**

ChatGPT 🟢 is everywhere! Natural Language Processing is the principle behind it!!

1. What is Natural Language Processing?

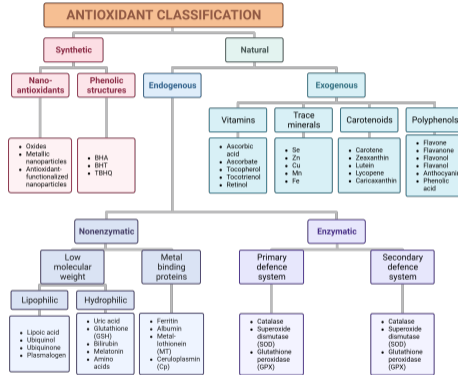# NATURAL LANGUAGE PROCESSING (NLP)

- Branch of AI → Giving computers the ability to understand text
- Computational Linguistics + Statistical and Machine Learning Models
- Very good chance you've interacted with NLP!

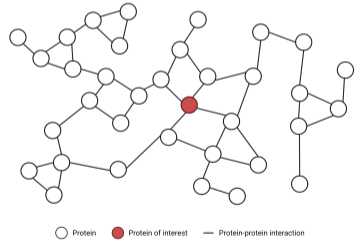# 2. WHY DO WE NATURAL LANGUAGE PROCESSING?

APPLICATIONS

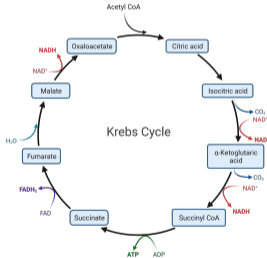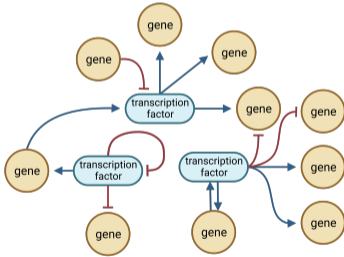- Large volumes of text data!!
- Structuring a highly unstructured data source

3. Understanding Biological Networks

BIOLOGICAL NETWORKS
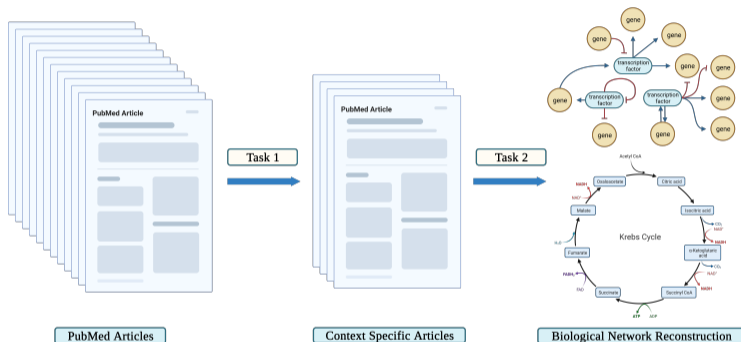
- A holistic understanding of biological systems is important to:
  - Generate hypotheses for biological systems
  - Identify drug targets and biomarkers
- Computational network models can help in describing complex interactions

Gene Regulatory, Metabolic and Protein Interaction Networks

4. NLP for Biology

TRADITIONAL PIPELINE

PubMed Articles → Task 1 → Context Specific Articles → Task 2 → Biological Network Reconstruction
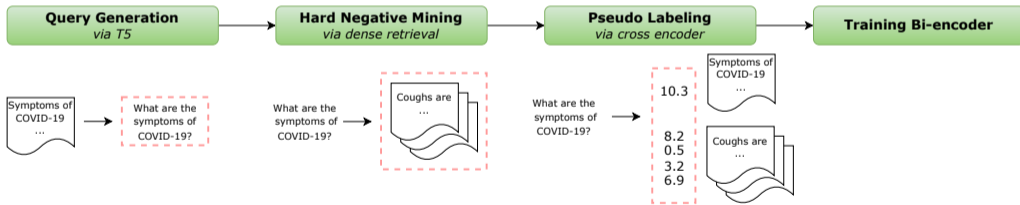
- Exhaustive manual curation 😟 😑
- **The Steps:**
    - Get textual data according to the user's context
    - Look for entities and relations in each PubMed abstract (or full-text)

TASK 1: INFORMATION RETRIEVAL

Biomedical Information Retrieval using Generative Pseudo Labelling

- Science of searching for documents themselves
- Give a query → Get a document

Retrieve → bi-encoders and Re-Rank → cross-encoders

TASK 2: BIOLOGICAL NETWORK RECONSTRUCTION

Named Entity Recognition and Relation Extraction

Thioredoxin-interacting protein (TXNIP) is an α-arrestin family protein that is induced in response to glucose elevation. It has been shown to provide a negative feedback loop to regulate glucose uptake into cells, though the biochemical mechanism of action has been obscure. Here, we report that TXNIP suppresses glucose uptake directly, by binding to the glucose transporter GLUT1 and inducing GLUT1 internalization through clathrin-coated pits, as well as indirectly, by reducing the level of GLUT1 messenger RNA (mRNA). The glucose influx through GLUT1 restores ATP-to-ADP ratios in the short run and ultimately induces TXNIP protein production to suppress glucose uptake once energy homeostasis is reestablished.

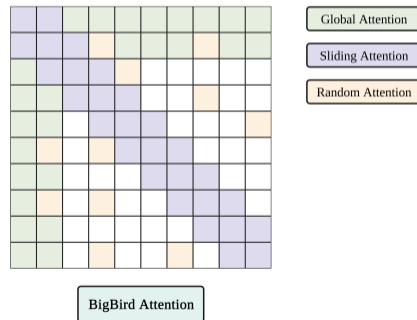Gene/Protein   Chemical   Cell Function

- **Relations Extracted:**
  - Glucose → TXNIP
  - TXNIP → Glucose uptake
  - TXNIP → GLUT1 mRNA → GLUT1

TASK 2: BIOLOGICAL NETWORK RECONSTRUCTION

BioBigBird for Biomedical Language Understanding

- Attention is all you need![a]

- $\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^{\top}}{\sqrt{d_k}}\right).\mathbf{V}$

  - Query ($\mathbf{Q}$): Text in the search bar
  - Key ($\mathbf{K}$): Video title
  - Value ($\mathbf{V}$): Best matched videos

---

[a]Vaswani et al., *Advances in neural information processing systems* (2017).

Global Attention

Sliding Attention

Random Attention

BigBird Attention

Summary and Future work

- The Next Frontier For Large Language Models Is Biology!
- Manual annotation (INCEPTION[1])
- Inclusion of multi-omics data for accurate reconstruction
- Modelling disease conditions

[1] Klie et al., *Proceedings of the 27th international conference on computational linguistics: system demonstrations* (2018).

ACKNOWLEDGEMENTS

- **Dual Degree Program Students:** Vasudev Gupta, Reneeth Krishna M G
- **PhD Scholar:** Pavan Kumar S
- **Advisor:** Dr Nirav P Bhatt

Thank You!

Any Questions?