Stanford MEDICINE

Emerging Topics in Immunology Course – Exploring the use of LLMs in Immunology

Teaching team: Nikesh Kotecha, Ananth Ganesan, Daniel Goncharov & Purvesh Khatri

Why this class?
(output from GPT-4)

- As you may already know, a large language model like me, GPT-4, developed by OpenAI, is an artificial intelligence model designed to generate human-like text.

- It's a member of a family of models called "transformer-based models," which are built around the concept of attention mechanisms, allowing the model to weigh the importance of different words or phrases in generating the next word in a sentence.

- This is somewhat akin to how your immune system can "pay more attention" to certain antigens based on past exposure and the creation of memory cells, though it's a rough comparison.

- GPT-4 was trained on a vast range of internet text, and the model aims to predict the next word in a sentence given all of the previous words within some context. However, it's important to note that the model does not understand the text in the way humans do. It doesn't have beliefs or opinions. Instead, it makes predictions based on the patterns it learned during training.

- One key concept that makes these models so powerful is "transfer learning". This means that a model trained on one task can apply its learned knowledge to perform well on a different but related task.

- This is akin to the cross-reactivity we see in immunology, where T-cells or B-cells activated by one antigen can respond to another antigen that has similar molecular structures.

- GPT-4 is an autoregressive model, meaning it generates sentences word by word from left to right. It takes the sequence of words so far as input, and predicts the next word. Each word prediction is conditioned on the previous words, similar to how an immune response is conditioned on the specific antigen presented.

- Models like GPT-4 can be fine-tuned on specific tasks, such as translation, summarization, or answering questions about a text.

Stanford MEDICINE | Technology & Digital Solutions
STANFORD HEALTH CARE & SCHOOL OF MEDICINE

- This is akin to how your immune system "fine-tunes" its response to pathogens through the selection and expansion of specific clones of immune cells.

- However, it's crucial to mention the limitations. GPT-4 has no concept of "truth" or "accuracy" inherently.

- It generates outputs based on the data it was trained on, much like how an immune response's effectiveness is determined by the initial training and exposures of the immune system. It also has a knowledge cutoff, and any events or facts emerging after this cutoff are unknown to it.

- Remember, while GPT-4 is a powerful tool, its responses should be carefully reviewed, especially when it comes to complex scientific subjects.

- The tool can assist in generating ideas and producing drafts, but its output should always be critically evaluated.

Intros & Class Logistics
Nikesh Kotecha

What is the goal of this class?

- To carve out some space and time to learn about large language models and understand how or where they could be applied in immunology

- Provide a forum for others who are interested in a similar conversation and learn from each other

What is our mode of operation

- We are aiming to use the "Socratic" method in this class

- We are not experts in this area and are looking to learn with you
  - Through readings and explorations/topics of interest
  - Through project work

- We can help guide conversations and bring in appropriate speakers based on class interest and participation

What are we expecting from you

- Come to class and participate/ask questions.

- Once we have a sense of the class, we'll assign some readings/videos and ask some of you to present it to the class (~5 min). Please do this if you are assigned.

- One of the best ways to learn in this space will be to try some things. We'll ask you to think of a project/set of explorations. The assignment is open ended and can be done in groups.

- We'll provide some advisors/mentors you can work with once we get a sense of your project

Who are we (I)

- Nikesh Kotecha - Faculty Advisor/Facilitator

  - Adjunct Faculty, Stanford Medicine
  - Head of Data Science, Stanford Healthcare

- Purvesh Khatri – Faculty Advisor/Facilitator

  - Assoc. Professor, Stanford Medicine
  - Institute for Immunity, Transplantation & Infection
  - Center for Biomedical Informatics Resarch

Who are we (II) – Ananth Ganesan

- Ananth is a senior data scientist at Stanford Medicine, where he develops novel computational and machine learning-based methods to study the human immune system.

- His methods have been used to characterize the epigenetic landscape of healthy humans, and the immune response to asthma, IBD, lupus and cancer.

- Ananth obtained his Ph.D. in computational and mathematical engineering from Stanford.

Who are we (III) – Daniel Goncharov

- Daniel Goncharov is an ANA Avatar Xprize finalist and Google Developer Expert in Machine Learning. He leverages his robust knowledge in applied mathematics and computer science to employ advanced AI and machine learning methodologies across diverse fields.

- He has spearheaded significant initiatives such as the AI vs COVID-19 project and the establishment of the AI and Robotics lab at School 42 Silicon Valley.

Stanford MEDICINE | Technology & Digital Solutions
STANFORD HEALTH CARE & SCHOOL OF MEDICINE

Format of the class

- ~30 min overview/presentation

- ~30 min presentations from class members
  - Ideally 5 people present some assigned readings/topics of interest (~5 min each)

- ~30 min hands on component/time for project work

- Registration not required but encouraged

Current Class Syllabus (I)

Date	Overview (Working Titles)	Class Presentations	Hands On
6/26	Class Intro - Nikesh Overview of Transformers - Otavio	Comparing scBERT, scGPT and Geneformers - Ananth	Prompt Engineering - Dan
7/11	Fine tuning/adapting an LLM. Immunology Example – Maxim Zaslavsky	Class presentations of assigned readings	Get together in groups/think about projects
7/18	Enriching LLM prompts via information retrieval. Building your own LLM. - Daniel Goncharov	Class presentations of assigned readings Get together in groups/think about projects – 5-10 min	Exercise around enriching prompts – Dan Enriching example
7/25	Outside Speaker (MedPalm2)		

Current Class Syllabus (II)

Date	Overview (Working Titles)	Class Presentations	Hands On
8/1	Work on projects.		
8/8	Evaluating LLMs – Scott Fleming	Class presentations of assigned readings	Evaluating LLMs exercise - TBD
8/15	(If time/interest) – Where are LLMs going/toward AGI?	Project Presentations	

A request ... From the Game of Offices ...
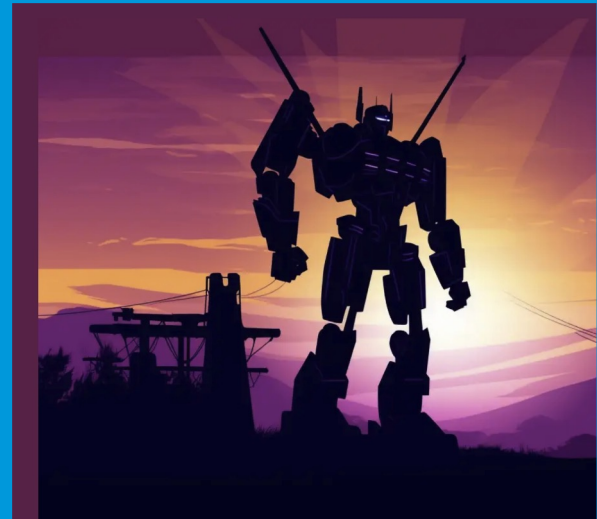

Daenerys Targaryen

Jon Snow

The Hound

r/midjourney Game of Offices

Stanford MEDICINE | Technology & Digital Solutions
STANFORD HEALTH CARE & SCHOOL OF MEDICINE

(Brief) Overview of transformers
Otavio Good

Transformer Based Models for Immunology

Ananth Ganesan

Getting started with LLMs
Daniel Goncharov

Thank you and see you on July 11

- Please fill out the form at
  http://goto.stanford.edu/imm310lec1
  - QR code on the right