



NYU

# AI for Scientific Research

Student Leaders: Jiamu, Ishita, Laiba, Lucy, Victor

Ahmad, Bree, Cheyenne, Dan, Garima, Haoyu, Yuchen, Joshua, Juntao, Kartik, Khalid, Lu, Malavika, Kaddy, Ruochen, Seoeun, Shengyao, Shruti, Suriya, Swapnil, Tanishi, Thomas, Tiyas, Varshitha, Xinqi

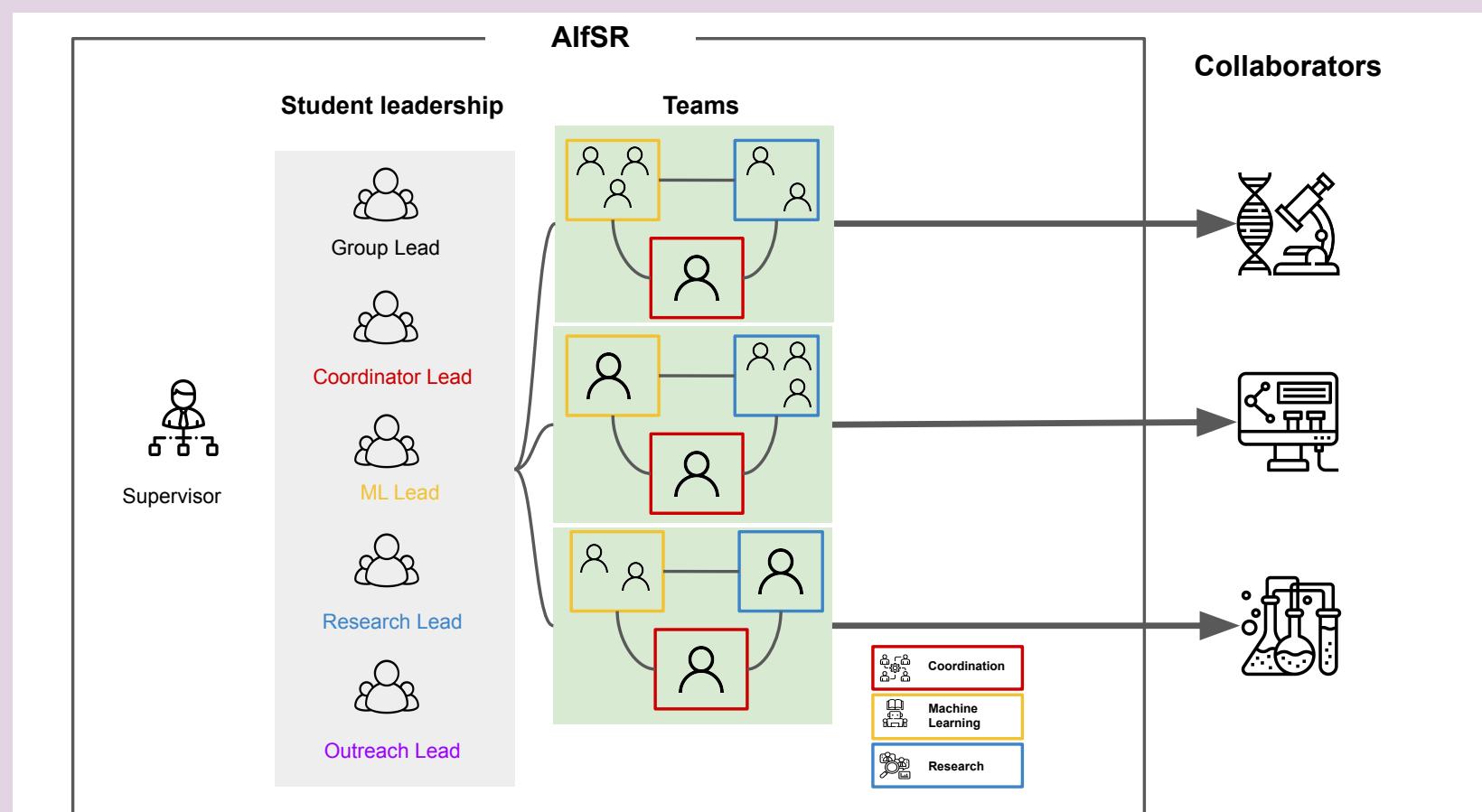
Mentor: Dr. Sergey Samsonau



## AIfSR Structure and Roles

30 participants, 8 projects and 7 teams each working independently.

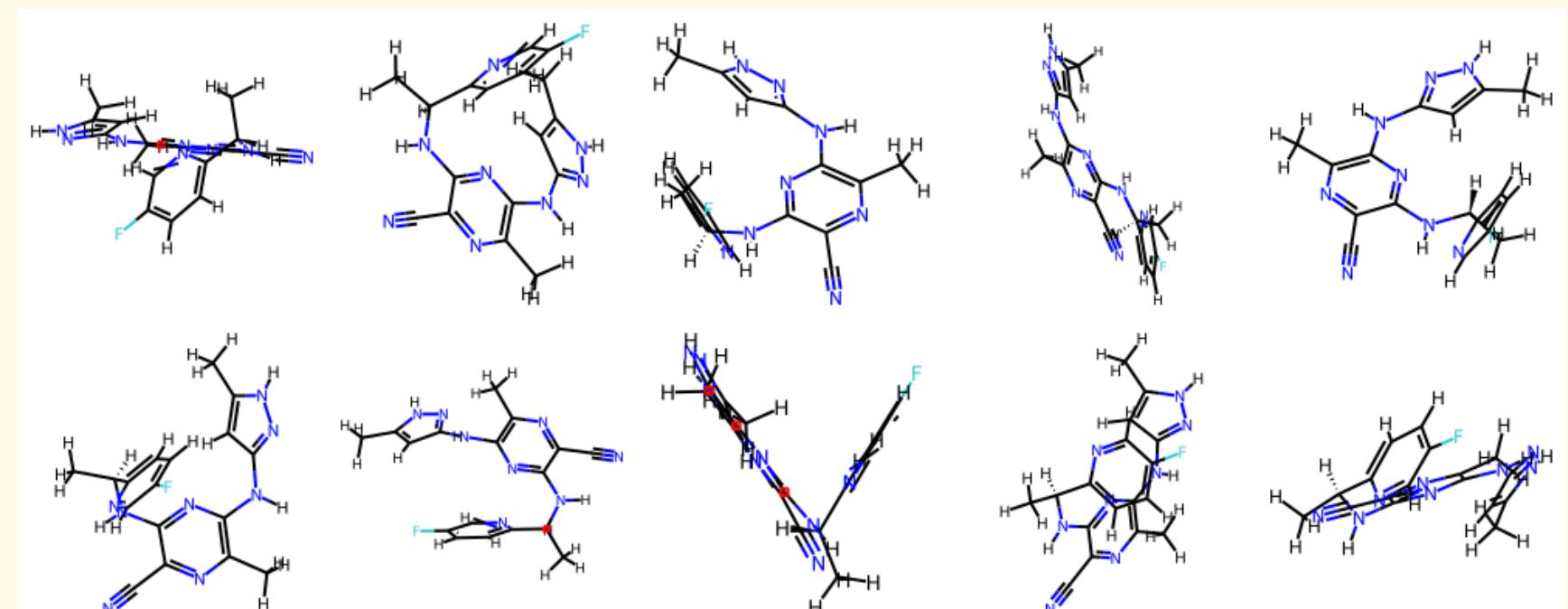
Group structure:



## Material Properties

**Goal:** Explore and discover novel chemical compounds for practical usage through discriminative and generative approaches.

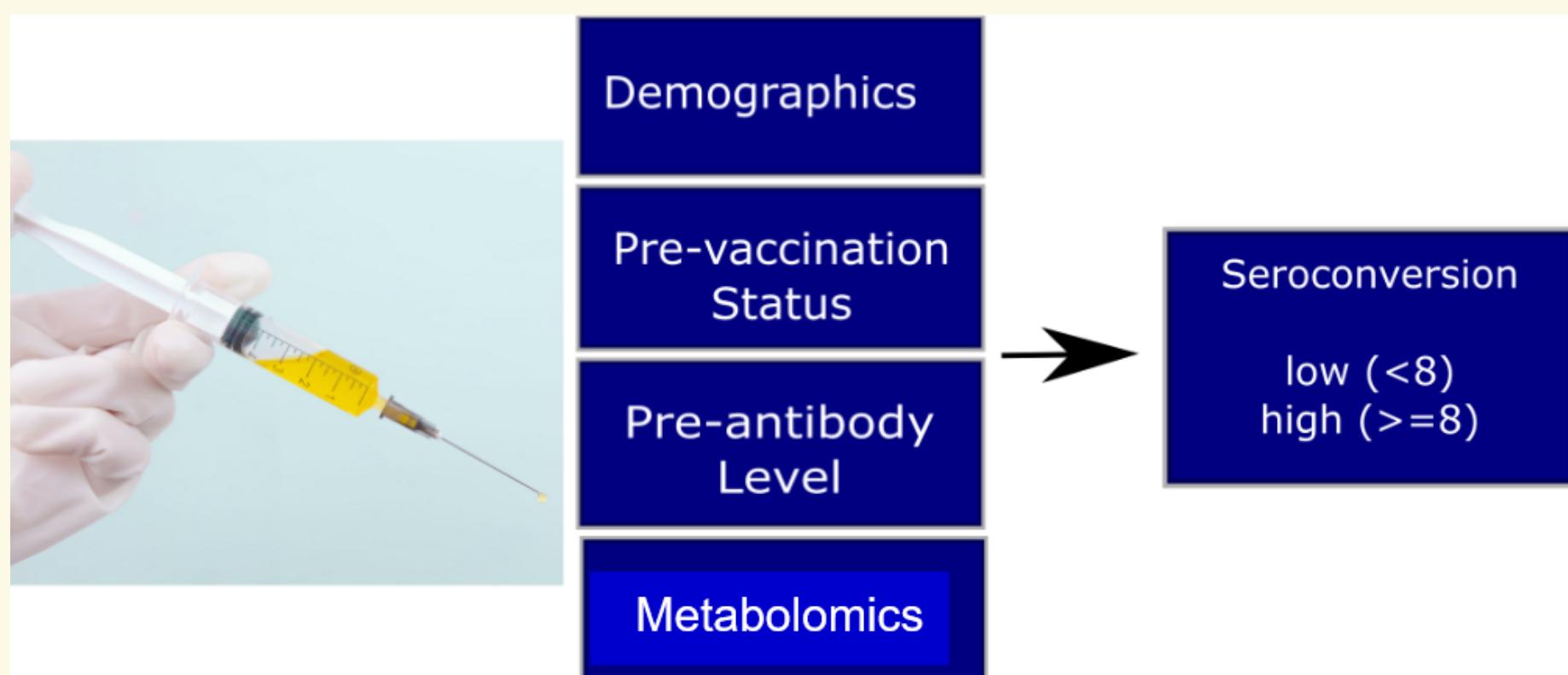
**Collaborator:** Dr. Mark Tuckerman, NYU Arts & Science, NYU Chemistry Department



## Bio Treatment

**Goal:** Develop a model to predict the efficacy of the flu vaccine given multi-folded patient features ranging from demographics to metabolomics.

**Collaborator:** Dr. Christine Vogel, NYU Arts & Science, NYU Center for Genomics and Systems Biology



## Neurolinguistics

**Goal:** Implement a generative pre-trained large language model to produce Multimorpheme English Words, which will aid in the neurolinguistics study of morphology.

**Collaborator:** Dr. Alec Marantz from NYU Arts Science, NYU Psychology Department



## Project Stages

### Formulation

Identify and connect with a collaborator, understand the data and formulate the task.

### Solution Development

Design, develop, validate and test the statistical solution within formulated scope of the task.

### Deployment

Integrate solution into collaborator's research environment, verify usage and collect feedback.

## Project Types

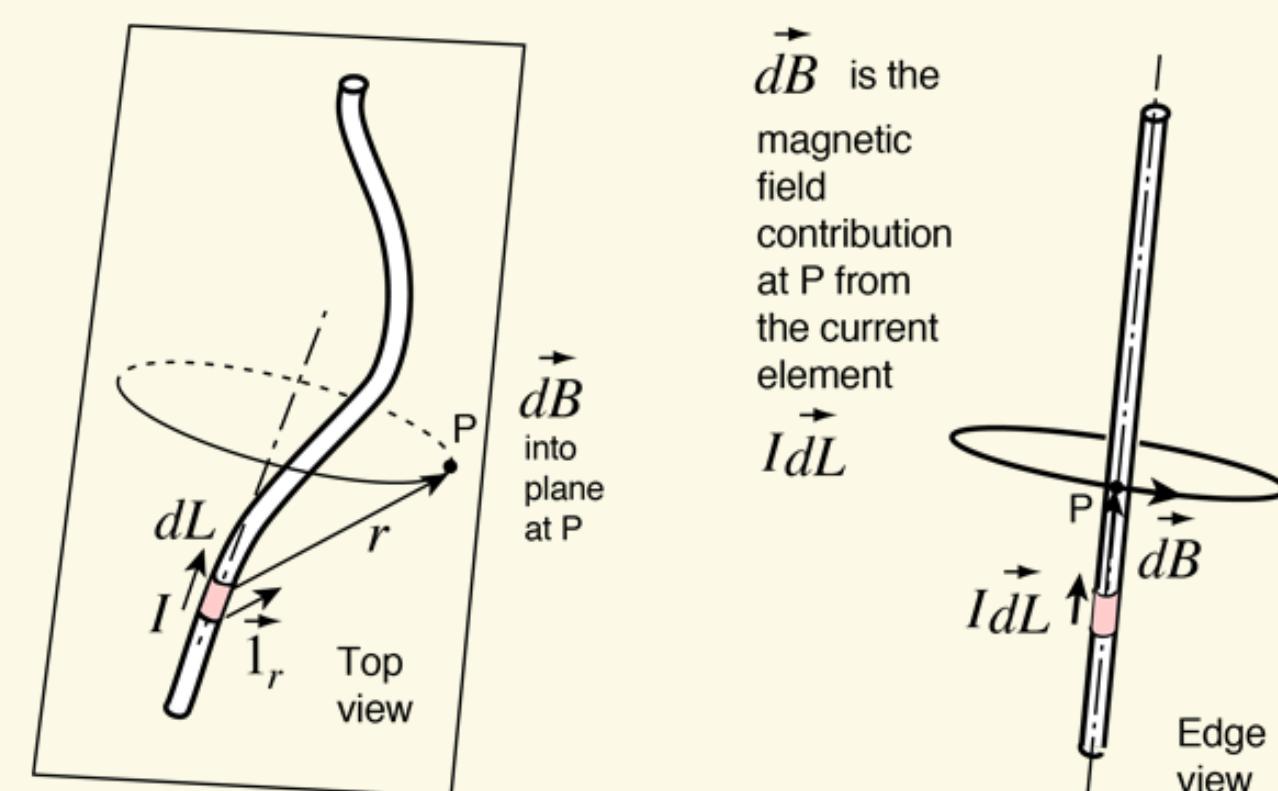
**Type 1** AI solutions on data for specific lab – In collaboration with lab professors, teams work on unique datasets provided by scientific labs

**Type 2** AI-enabled software or software components development and/or evaluation – Teams contribute to the world of scientific software through performance evaluations of already-existing AI software, enrichment of existing python/R packages with AI functionality, or development of new AI-powered packages for scientific applications

## Electromagnetism

**Goal:** Predict the current given the value of magnetic field using machine learning model – inverse task of the Biot-Savart law.

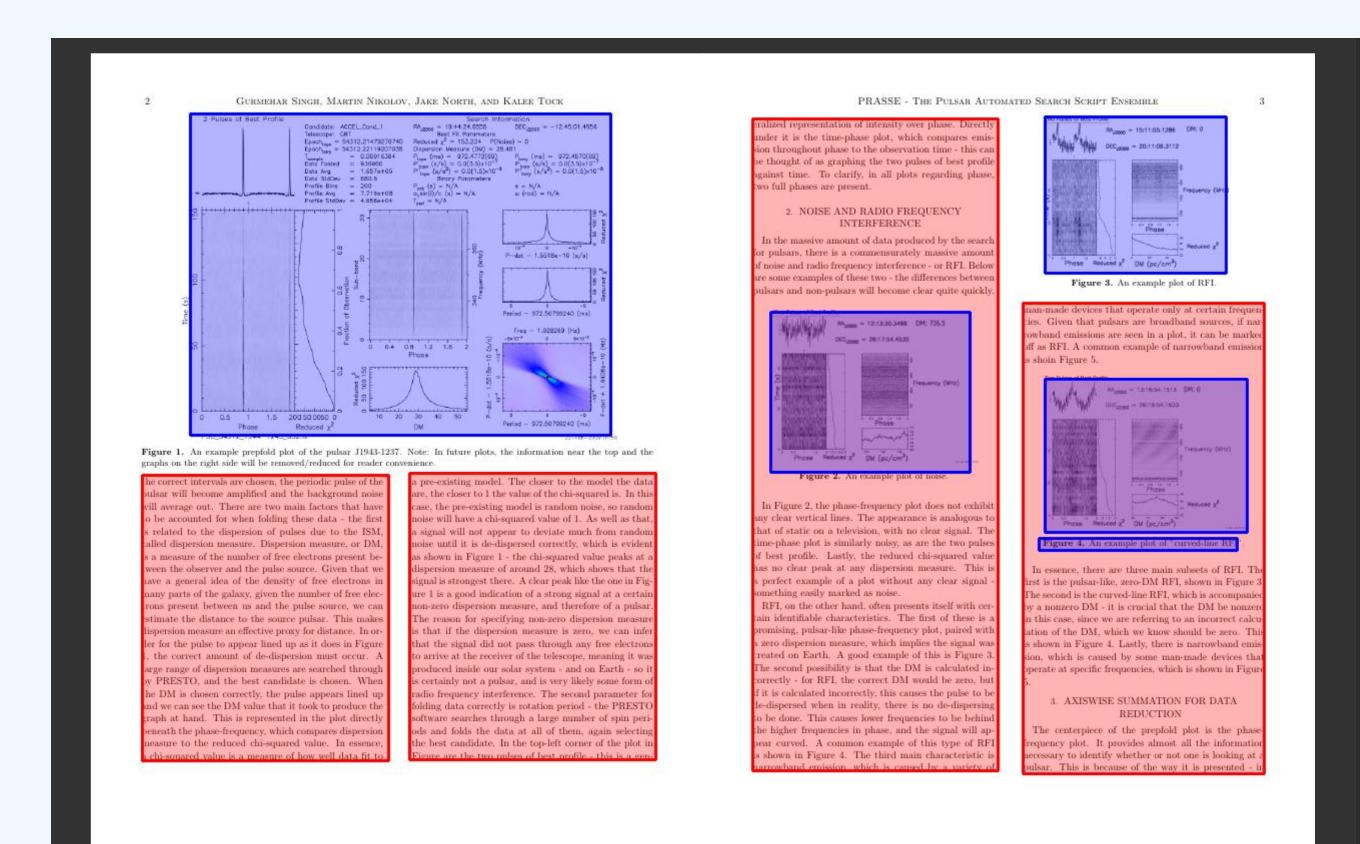
**Collaborator:** Dr. Alexj Jerschow, NYU Arts & Science, NYU Chemistry Department



## Scientific Literature

**Goal:** Establish a pipeline to convert scientific papers in pdf format into html text format. The html text will be converted to audio assisting people with special needs.

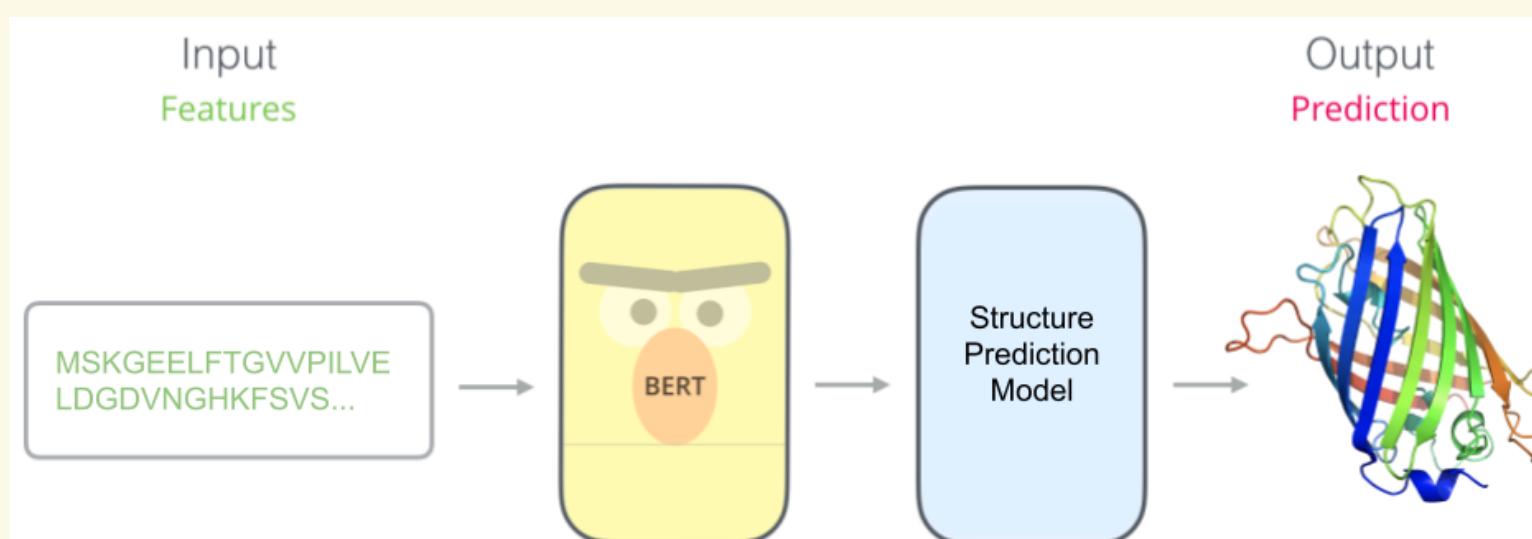
**Collaborator:** Dr. Jonathan Bragg and Dr. Kyle Lo from Semantic Scholar



## Genetics

**Goal:** Applying unsupervised learning techniques by using natural language processing models such as ProtT5, ProtBERT, ProtXlnet, CPCProt to predict the functionality of proteins.

**Collaborator:** Dr. Jane Carlton, NYU Arts & Science, Center for Genomics and Systems Biology



## Benchmarking

**Goal:** Benchmarking team focuses on benchmarking the hardware performance of NYU HPC Greene on cloudmask generation tasks.

**Collaborator:** Dr. Gregor von Laszewski and Dr. Geoffrey Fox working in affiliation with MLCommons.

Benchmarking NYU HPC Greene's performance, time, speed, and efficiency for generating cloud mask for satellite images on MLCommons Science benchmarks. The submission is in Tensorflow and uses the U-net model for the task.



## List of Projects We've Worked On

### Biology

- Vaccination response prediction
- Sleep patterns classification in EEG data
- Microscopy images segmentation: cells
- Tracking of cells moving under microscope
- Prediction for worms embryonic lethality
- Symptoms of Malaria
- Genetics

### Chemistry

- Inverse task: current in batteries from magnetic field
- Material properties prediction from structure

### Physics

- Focused Ion Beam (FIB) profiles
- Trajectory of particles under microscope
- Spectroscopy for microdiamonds

### Psychology

- Neurolinguistics of Morphology Generation
- Generic gender categories labeling in text

### Scientific Software

- Automatic annotation of scientific figures
- Parsing scientific papers from PDF
- CloudMask Benchmarking on NYU HPC