ALPHAFOLD PROTEIN STRUCTURE DATABASE

CSC 480

ABOUT ALPHAFOLD

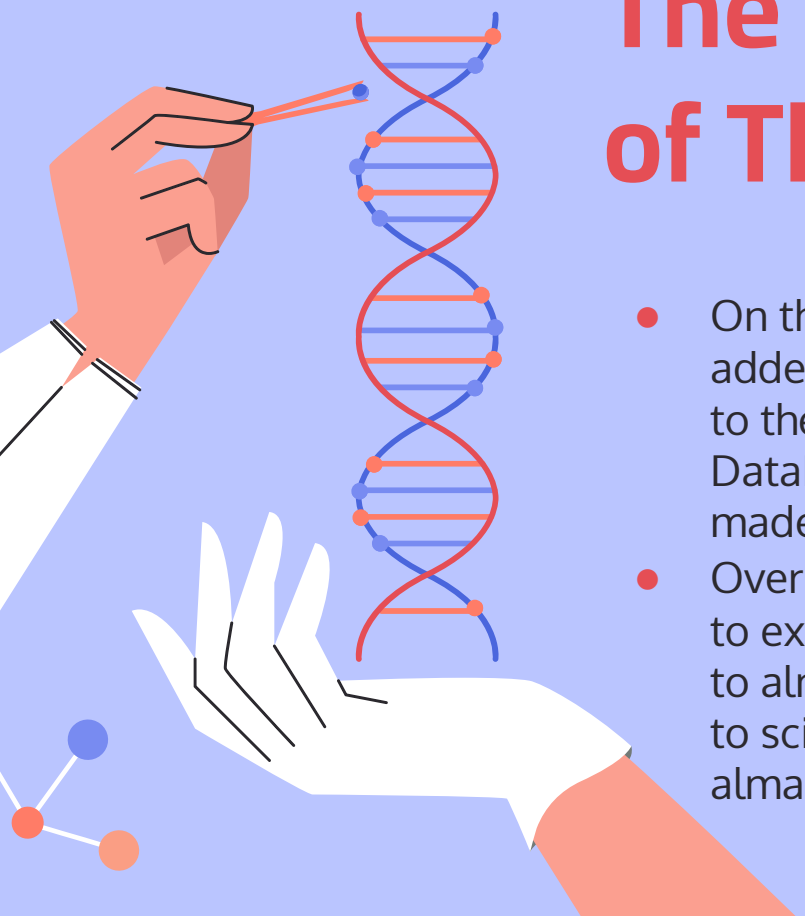
- DeepMind created an AI system called AlphaFold to help scientists understand the forms of a protein from its string of amino acids.
- It's trained on the sequences and structures of about a hundred thousand proteins that scientists around the world carefully mapped out.
- AlphaFold can predict the shape of a protein by its amino acids to a high percentage.
- This could allow us to understand diseases faster and develop drugs to combat them.

OUR TEAM OF "SCIENTISTS"

GIANNA

DANIEL

JASON



The Current State of The Application

- On the 28th of January 2022 DeepMind added 27 new proteomes (190k+ proteins) to the AlphaFold Protein Structure DataBase. Over 300,000 researchers have made use of the database to date.
- Over the coming months DeepMind plans to expand the protein structure database to almost every sequenced protein known to science. This would create a protein almanac of the world.

Who Created AlphaFold ?

- Google’s DeepMind
- In July of 2021, AlphaFold was released for free, to the whole world.

Which AI Techniques Are Used ?

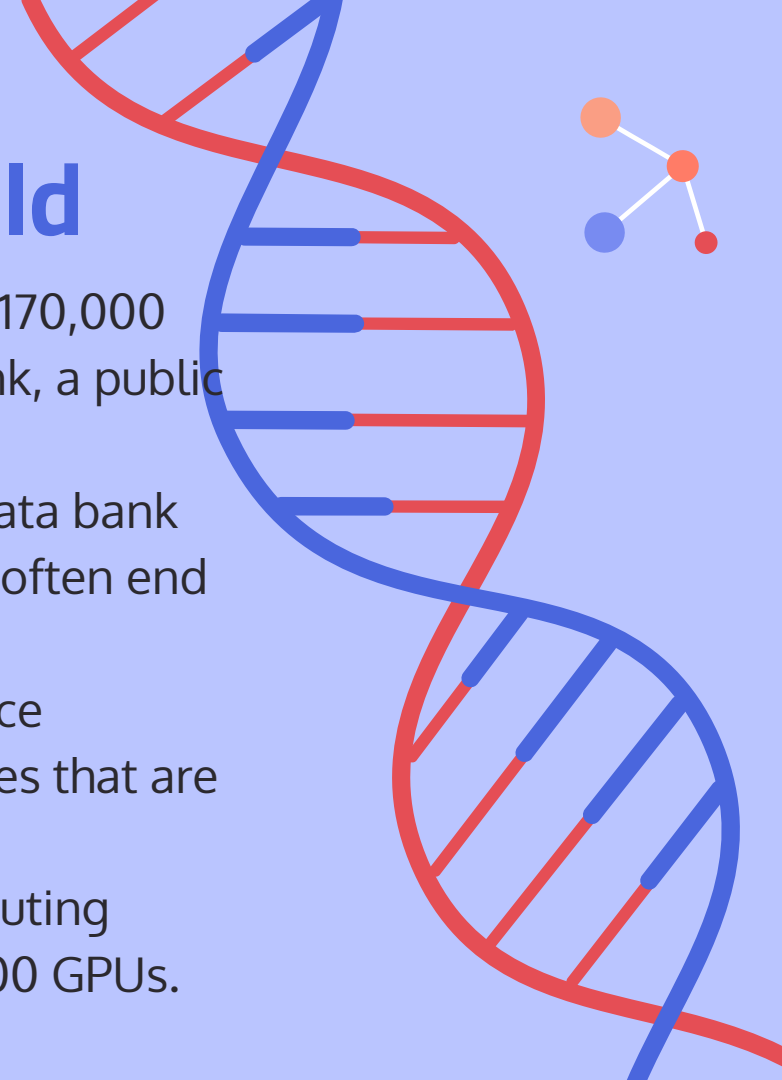
It uses an attention network, a deep-learning technique that allows an AI to train by focusing on parts of a more significant problem.

Example:

- Please think of the approach to assembling a jigsaw: it pieces together local chunks first before fitting these into a whole.

Training AlphaFold

- DeepMind trained AlphaFold on around 170,000 proteins taken from the protein data bank, a public repository of sequences and structures.
- It compared multiple sequences in the data bank and looked for pairs of amino acids that often end up close together in folded structures.
- It then uses this data to guess the distance between pairs of amino acids in structures that are not yet known.
- Training took "a few weeks," using computing power equivalent to between 100 and 200 GPUs.



Current Usage ?

- This year, DeepMind plans to release a total of more than 100 million structure predictions.
- That is nearly half of all known proteins — and hundreds of times more than the number of experimentally determined proteins in the Protein Data Bank (PDB) structure repository.
- Since AlphaFold was released, biologists had a new powerful tool to use that allowed them to work in a new way.
- A tremendous amount of new research papers began coming out about AlphaFold and protein modeling.
- "By now, more than 400,000 people have used the EMBL-EBI's AlphaFold database, according to DeepMind.

STAGES OF DEVELOPMENT

2016

DeepMind establishes a small team to begin work on protein structure prediction.

2018

The first public test of AlphaFold's performance

2020

A solution to a 50-year-old grand challenge in biology

2021

Putting the power of AlphaFold into the world's hands

2022

Over 300,000 researchers worldwide have made use of the database to date.

Issues With AlphaFold

"AlphaFold isn't always that accurate. Of the several dozen GPCR structures his lab has solved but not yet published, about half the time, the AlphaFold structures are fairly good, and half the time they're more or less useless for our purposes."

— Bryan Roth

LSDR :
There's still work to be done on AlphaFold as it does poorly on some protein predictions.

What Are The Benefits Of This Application ?

- Develop vaccines, understand diseases faster and develop drugs to combat them.
- Develop enzymes to break down plastic waste or capture carbon from our atmosphere.
- Predicting protein structure will also help design synthetic proteins, such as enzymes that digest waste or produce biofuels.
- Introduce synthetic proteins that will increase crop yields and make plants more nutritious.

THANK YOU !

ANY QUESTIONS ?