

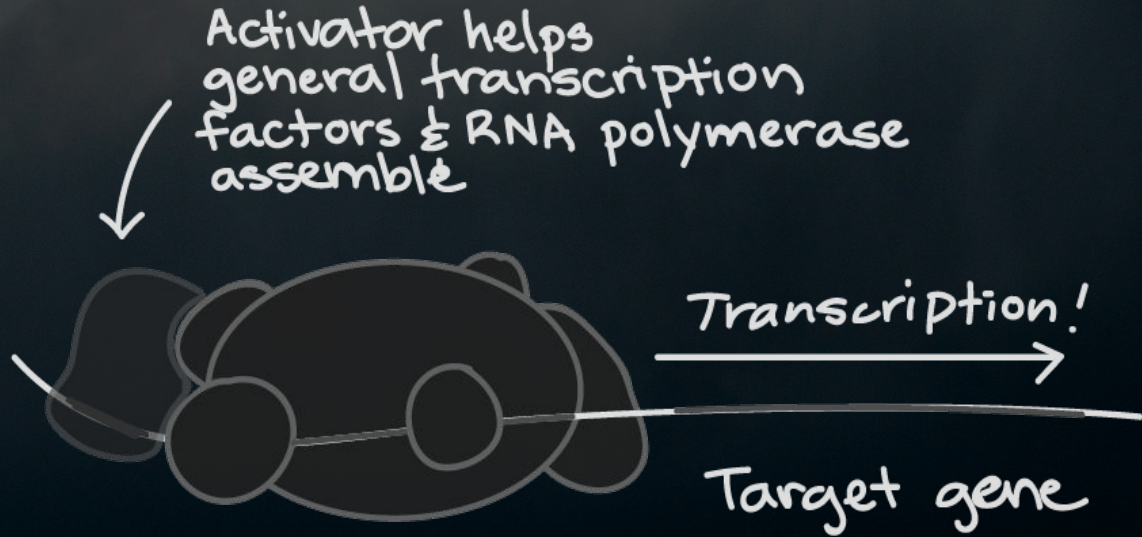


Heli.AI Transcription Factors & AI

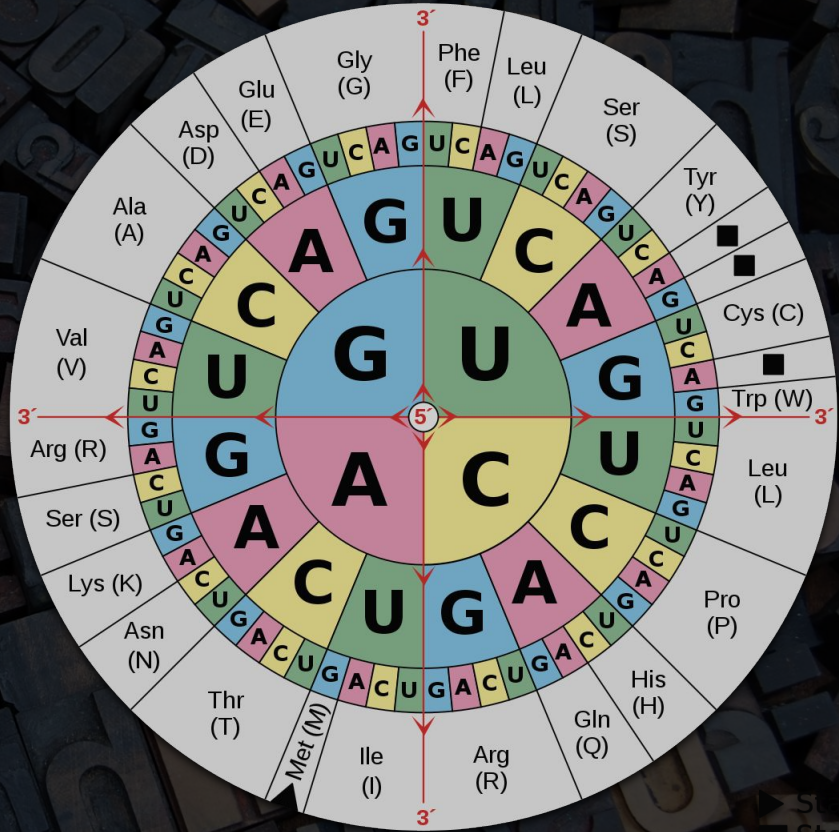
A photograph of Barack Obama speaking at a podium, with his right hand raised. He is wearing a dark suit and a patterned tie. Behind him, a diverse group of people, including a woman with glasses and a young man, are looking towards him. The background features a wooden door and an American flag. The entire image is dimmed, and the text is overlaid in the center.

**We don't
understand the
“punchline”**

The "Punchline"



Protein Encoding Regions

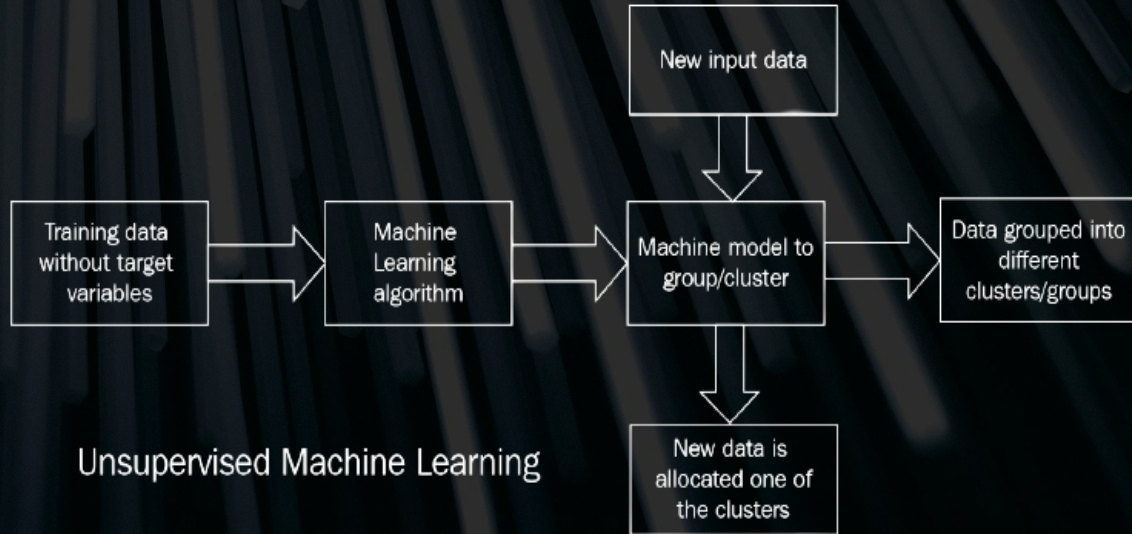




Utilizing Big Data

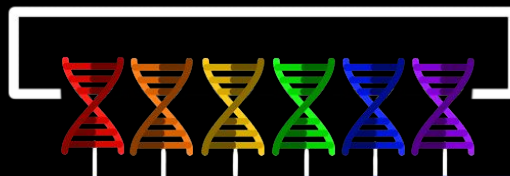
AI Data Analysis Models

- **Utilizes Bio-NLP algorithms**
- **Unsupervised AI model**
- **Applies Unsupervised Bio-NLP AI algorithm to DNA sequences**



Pre-trained model
diagram

Various Organism
Datasets



Encoding



Encoded Data

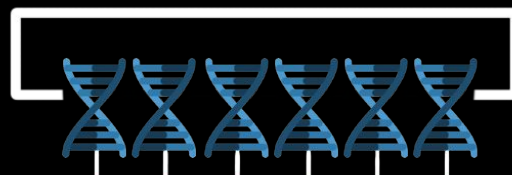
Bio-NLP-AI



Internalized
Biology's
"language"

Fine-tuned model
diagram

Targeted Organism
Datasets



Encoding

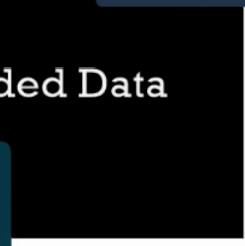


Encoded Data

Pre-Trained
Bio-NLP-AI



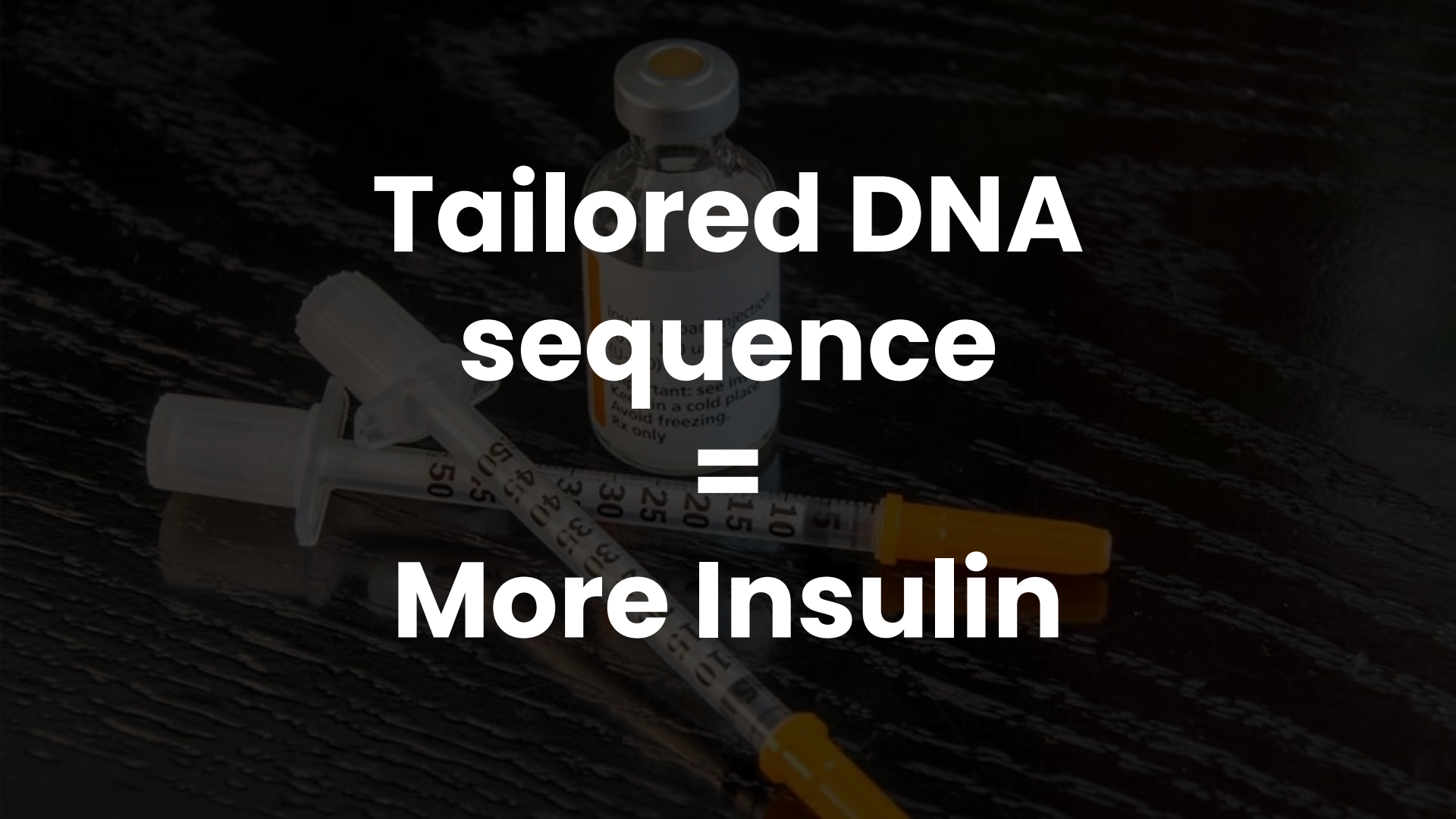
Optimized
Nucleotide
Sequence



A scientist with long dark hair, wearing a white lab coat and blue nitrile gloves, is working in a laboratory. She is holding a multi-well plate and appears to be transferring liquid from it into a larger container. The background is slightly blurred, showing other laboratory equipment and shelves. The overall lighting is dim, with a focus on the scientist and her work.

Target Market

Biotherapeutics: \$100 Billion

The background of the image shows medical supplies on a dark, textured surface. In the center is a small glass vial with a white cap and a label that includes the text "Keep in a cold place" and "Avoid freezing". Two syringes with yellow plungers are positioned diagonally across the frame, one in front of the other. The text "Tailored DNA sequence" is overlaid in large, white, bold font on the upper half of the image.

**Tailored DNA
sequence**

=

More Insulin

Potential

- **Optimize protein synthesis**
- **Expand the market**
- **Impact tens of millions**

Next Steps

- **Reference genomic data from RefSeq**
- **Develop AI models**
- **Test in lab**

Heli.AI: Transcription Factors & AI

Using unsupervised learning to optimize DNA sequences

What is the problem?

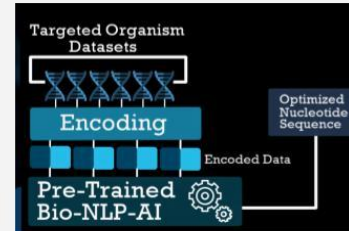
In biology, we only know about the input and outputs of the protein encoding regions of DNA. We know the basic code and what is being produced, but we don't understand the minute details of why it is being produced. If we don't understand, how do we replicate for our own use?

Solution

Transcription factors are activators that attach to the protein sequence DNA and create mRNA strands. Different protein encoding DNA strands for the same protein produce different levels of transcription factors. Our goal is to find the DNA sequence that optimizes the sequence to attract the most transcription factors.

Unsupervised Learning

Unlike supervised learning, unsupervised learning algorithms would take an input of a sequence and has to predict the next part. We would train our model with millions of DNA sequences and it will be forced to learn the patterns in these sequences, ultimately finding the optimized sequence.



Impacts

- Increasing efficiency in drug through Biotherapeutics (\$100 billion industry)
- Tailored DNA sequence for certain protein-based drugs
- Improve optimization of protein synthesis
- Can potentially impact tens of millions of people using protein-based drugs