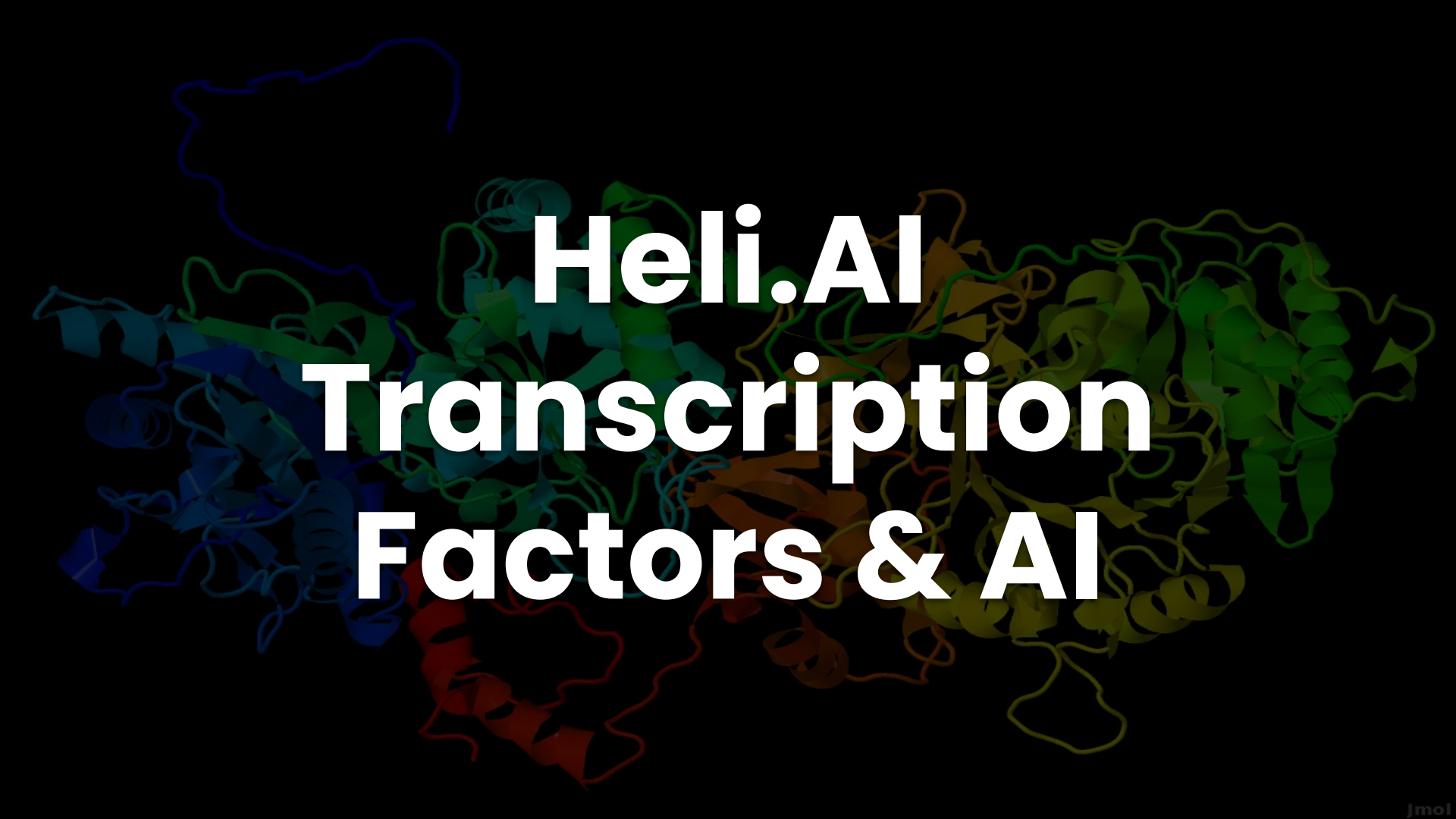
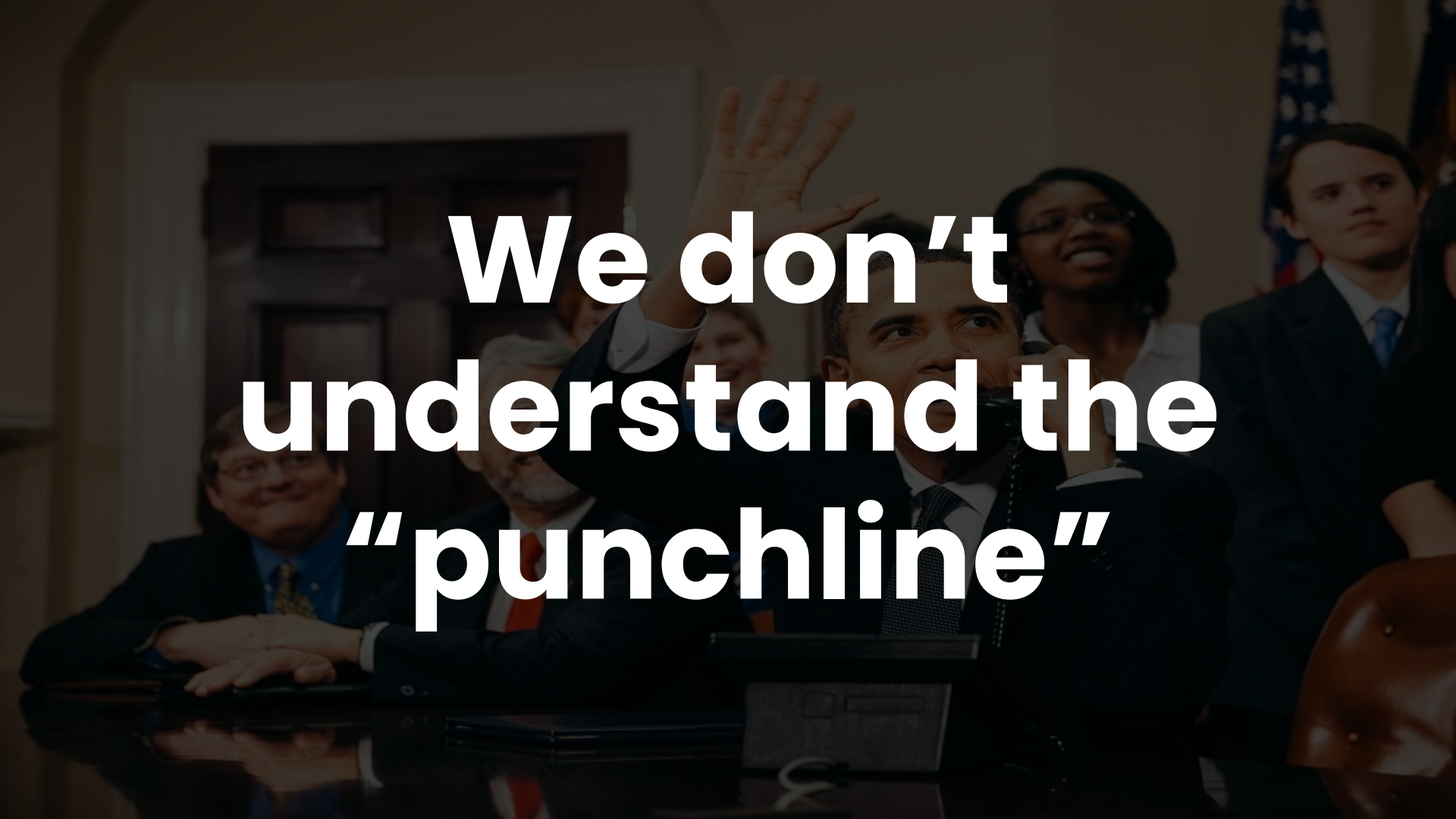
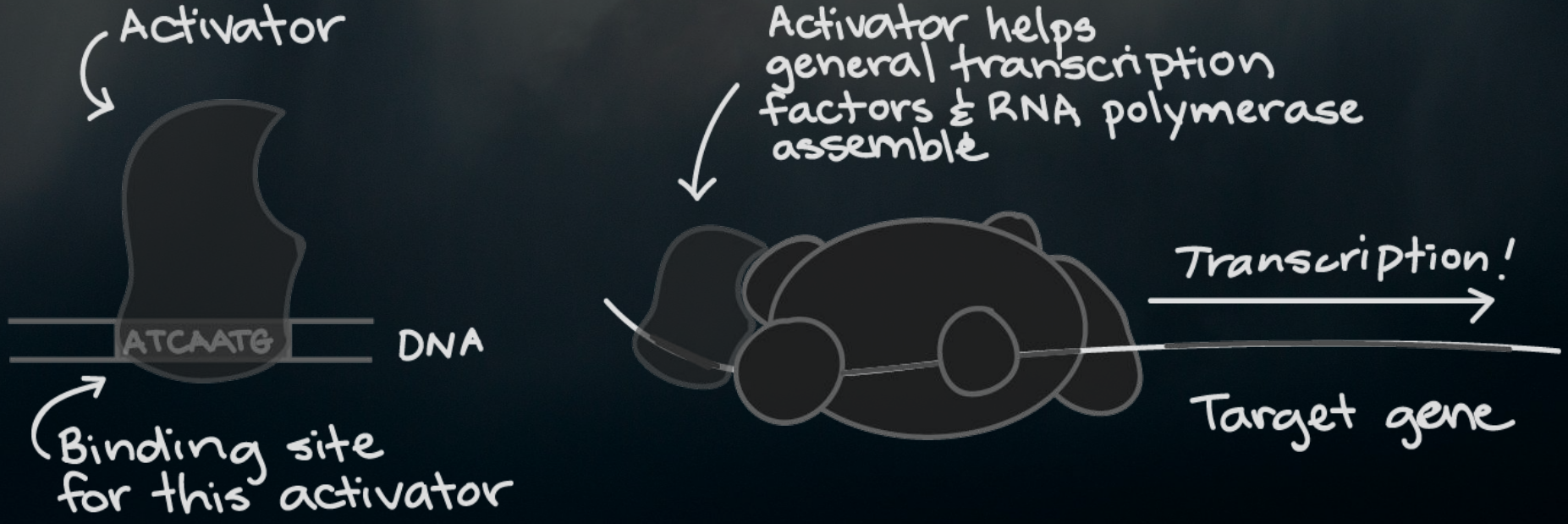
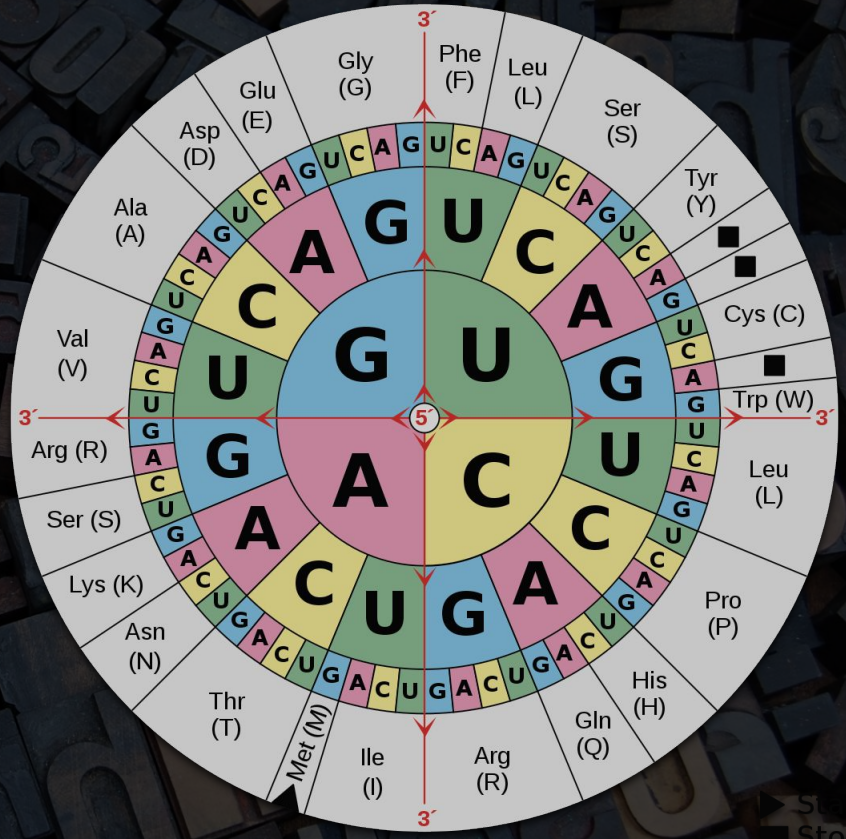

Heli.AI
Transcription
Factors & AI


We don’t
understand the
“punchline”

The "Punchline"



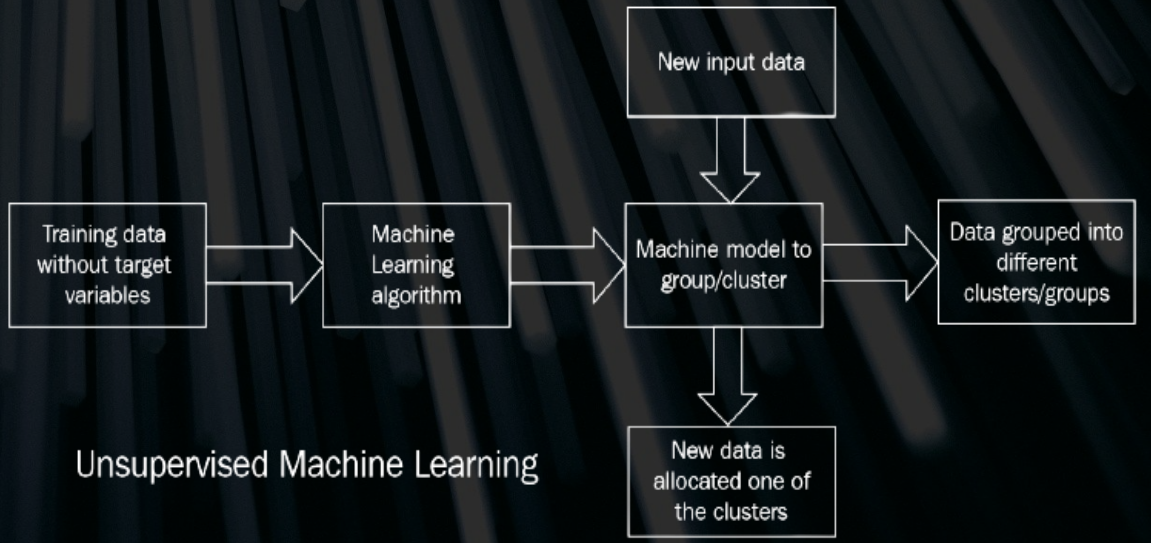
Protein Encoding Regions



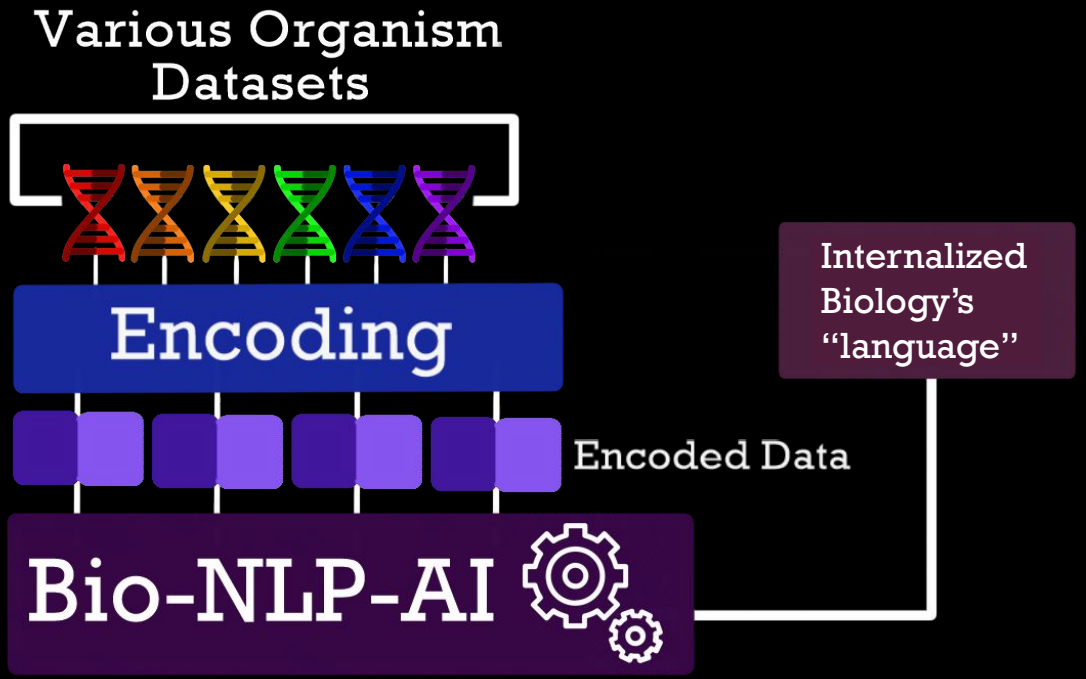

Utilizing Big Data

AI Data Analysis Models

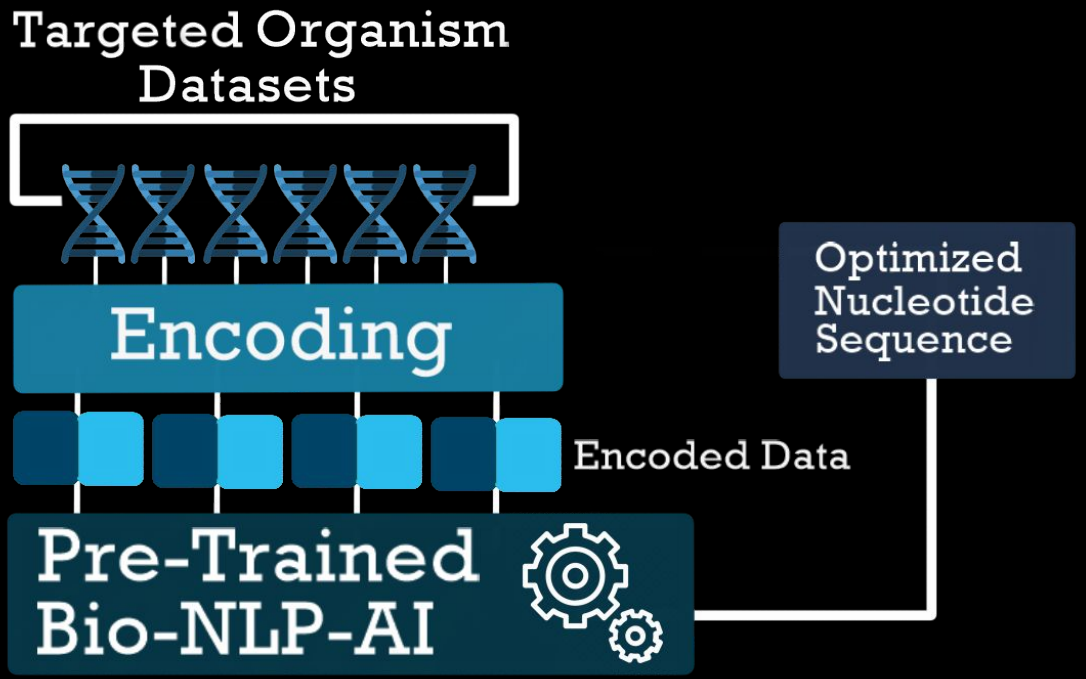
- **Utilizes Bio-NLP algorithms**
- **Unsupervised AI model**
- **Applies Unsupervised Bio-NLP AI algorithm to DNA sequences**



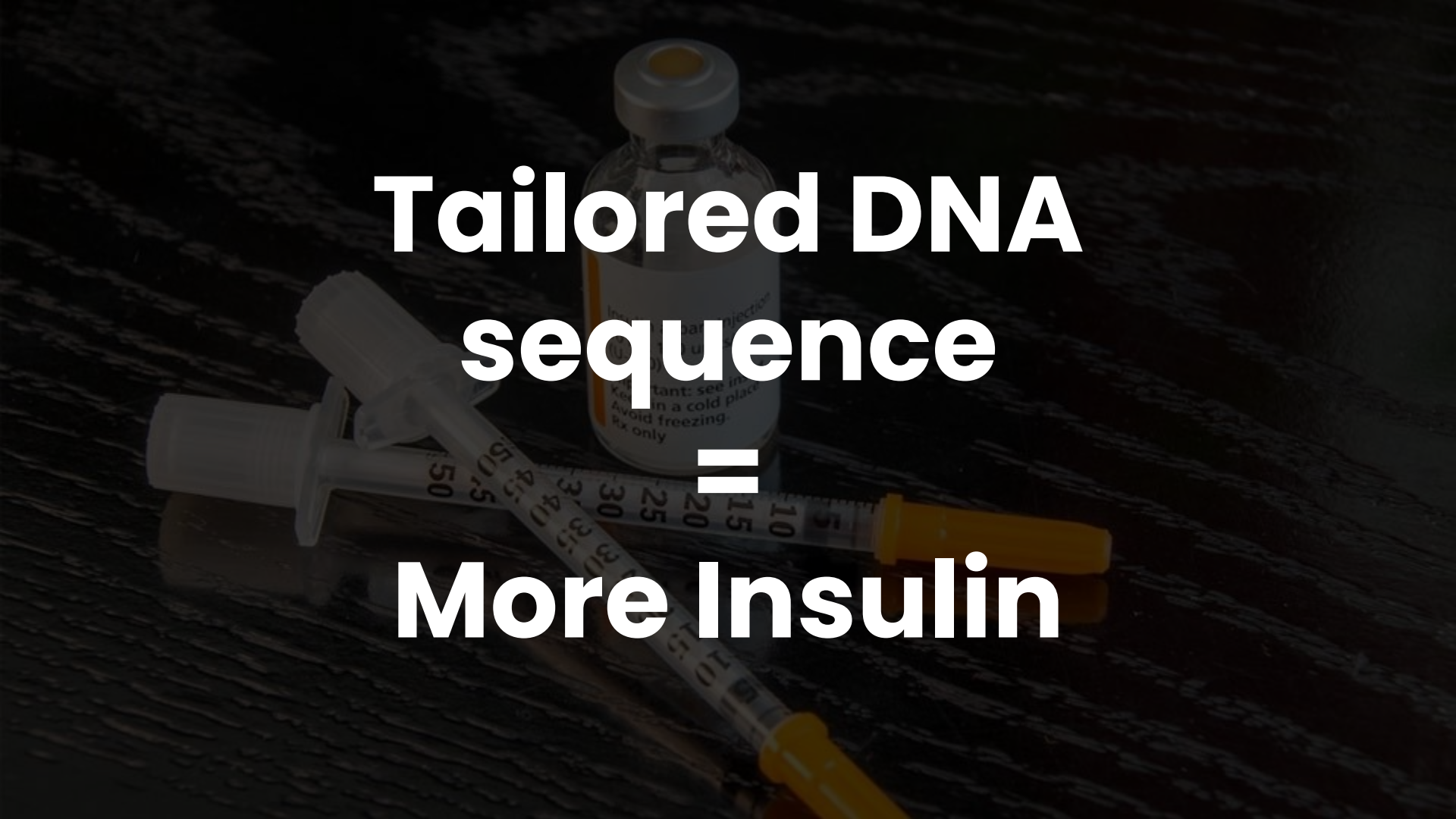
Pre-trained model diagram



Fine-tuned model diagram




Target Market
Biotherapeutics:
$100 Billion


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?

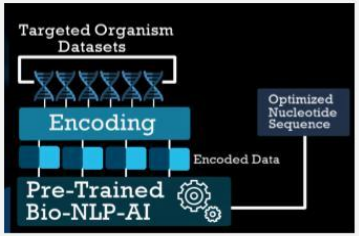
In biology, we only know about the **input** and **outputs** of the protein encoding regions of DNA. We know the basic code and what is being produced, but we don't understand the minute details of **why** it is being produced. If we don't understand, how do we replicate for our own use?

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

Further questions? oran.shai@gmail.com