

Revolutionizing Clinical Trials: Gen AI Powered Biometrics Breakthroughs

Kevin Lee
BMS

Disclaimer

The views and opinions presented here represent those of the speaker and should not be considered to represent any companies or organizations.

Agenda

- Introduction of Gen AI / ChatGPT
- Gen AI Use Cases in Biometrics
- Gen AI Tools / Systems
- Gen AI Process
- Gen AI powered Biometrics Team
- Future of Gen AI and Us
- Discussion



Do you use “Gen AI” or “ChatGPT”?



Introduction

- Gen AI
- ChatGPT

What is Gen AI?

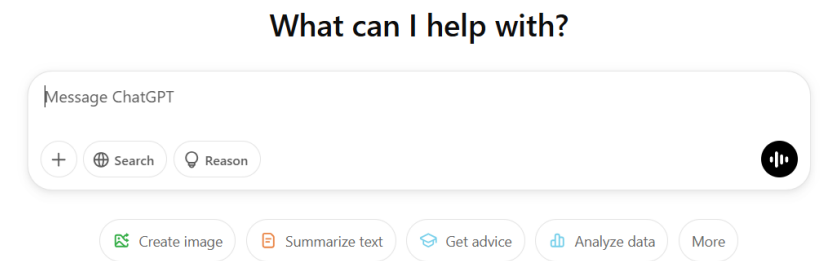
Gen AI – a trained Machine Learning model that generate new contents with a prompt.

- Text : Content Writing, Chatbots, Assistants, Search
- Code : Code Generation, Data Set Generation
- Image : Image Generation, Image Edit
- Audio : Voice Generation/Edit, Sound creation, Audio Translation
- Video : Video Creation/Edit, Voice Translation, Deepfake



What is ChatGPT?

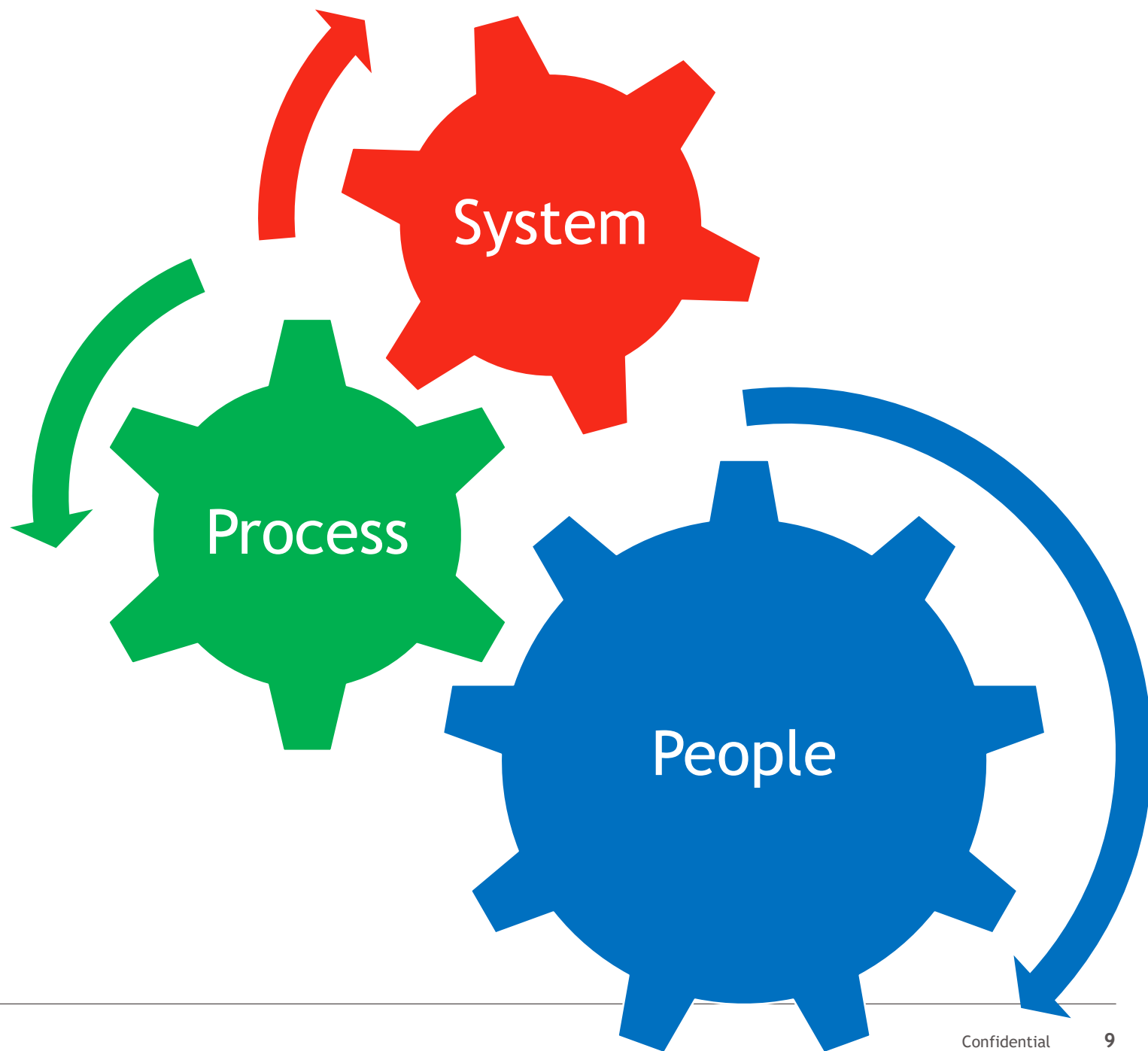
- ChatGPT is an advanced Gen AI developed by OpenAI.
- ChatGPT is trained on large corpus of text (~300B tokens).
- Its main strength
 - The ability to generate human-like responses in various contexts.
 - The ability to understand and generate text in a wide range of domains.



Why Gen AI in Biometrics?

- Enhanced Efficiency and Automation
- Easy Transition to Opensource capabilities
- Improved Data Analysis and Visualization
- Streamlined Clinical Trial Processes
- Super Biometrics Team

Three pillars to maximize Gen AI capability in Biometrics



Gen AI Use Cases / Tools in Biometrics

- Information Inquiry
- Document Development
- Coding
- Data Analysis
- Data Visualization
- Patient Profiles

Information Inquiry

Prompt :

What is CDISC?
Tell me in 200
words.

Output :

The Clinical Data Interchange Standards Consortium (CDISC) is a global, non-profit organization dedicated to creating and promoting standards for the collection, sharing, and analysis of clinical research data. Founded in 1997, CDISC aims to streamline the process of clinical research by enabling more efficient data collection and analysis, ultimately accelerating the development of new medical therapies and improving public health.

CDISC standards encompass various stages of the clinical research process, including study design, data collection, data management, data analysis, and data reporting. The core standards developed by CDISC include the Study Data Tabulation Model (SDTM), the Analysis Data Model (ADaM), and the Clinical Data Acquisition Standards Harmonization (CDASH). These standards ensure that data is collected and presented in a consistent and structured manner, facilitating interoperability between different systems and stakeholders.

Adopting CDISC standards is beneficial for regulatory submissions, as agencies like the U.S. Food and Drug Administration (FDA) and the Japanese Pharmaceuticals and Medical Devices Agency (PMDA) require data in CDISC-compliant formats. This harmonization helps reduce the time and cost associated with data integration and review processes, enhancing the overall efficiency and reliability of clinical trials and research initiatives.

Document Development : Simple protocol synopsis

- Prompt :

Write clinical trial **protocol synopsis** based on below
"title of study".

"title of study" : A phase 3, randomized, double-blind,
placebo controlled study to evaluate the efficacy of drug
A in schizophrenia patients.

• **Output :**

****Clinical Trial Protocol Synopsis: Phase 3 Study on the Efficacy of Drug A in Schizophrenia Patients****

****1. Study Title:****

A phase 3, randomized, double-blind, placebo-controlled study to evaluate the efficacy of drug A in schizophrenia patients.

****2. Objective:****

The primary objective of this clinical trial is to assess the efficacy of drug A in treating schizophrenia in comparison to a placebo. Secondary objectives include evaluating the safety, tolerability, and overall clinical benefit of drug A.

****3. Study Design:****

This study will adopt a randomized, double-blind, placebo-controlled design. Participants will be randomly assigned to either the treatment group receiving drug A or the control group receiving a placebo. The double-blind nature of the study ensures that both investigators and participants are unaware of the assigned treatment.

****4. Participants:****

The study will enroll adult participants diagnosed with schizophrenia according to standardized diagnostic criteria. Inclusion and exclusion criteria will be defined to ensure the homogeneity of the study population.

****5. Interventions:****

- Treatment Group: Participants will receive drug A in the prescribed dosage.
- Control Group: Participants will receive a placebo identical in appearance to drug A.

****6. Primary Outcome Measure:****

The primary efficacy endpoint is the reduction in the severity of schizophrenia symptoms, as measured by a standardized assessment tool, at the end of the treatment period.

****7. Secondary Outcome Measures:****

Secondary endpoints include:

- Changes in cognitive function.
- Improvement in overall functioning and quality of life.
- Incidence and severity of adverse events.
- Assessment of treatment adherence.

****8. Study Duration:****

The trial will be conducted over a specified period, with defined time points for participant recruitment, treatment administration, and follow-up assessments. The duration of the study will be determined based on the anticipated rate of symptom improvement and the need for robust data collection.

****9. Statistical Analysis:****

Statistical methods, including intention-to-treat analysis, will be employed to evaluate the primary and secondary endpoints. Sample size calculations will ensure the study's power to detect meaningful differences between the treatment and control groups.

****10. Ethical Considerations:****

The study will be conducted in accordance with ethical principles outlined in the Declaration of Helsinki and local regulatory requirements. Informed consent will be obtained from all participants, and measures will be taken to ensure participant confidentiality and safety.

****11. Data Monitoring and Safety:****

An independent Data Monitoring Committee (DMC) will oversee the trial's progress, ensuring participant safety and data integrity. Adverse events will be monitored and reported according to regulatory guidelines.

****12. Conclusion:****

This phase 3 clinical trial aims to contribute valuable insights into the efficacy and safety of drug A in the treatment of schizophrenia. The results of this study may have implications for improving the management of schizophrenia and enhancing the overall well-being of affected individuals.

Codes conversion

Prompt :

Convert below SAS codes to R codes:

```
data dm;  
  merge demo rand;  
  by usubjid ;  
run;
```

Output :

You can convert the SAS code to R using the `merge` function in R to perform the equivalent operation. Assuming you have two data frames `demo` and `rand` that you want to merge by the `usubjid` variable, here's the equivalent R code:

```
```\n# Assuming you have loaded your data into two data frames: demo and rand\n# Merge the data frames by usubjid\ndm <- merge(demo, rand, by="usubjid")\n```
```

In this code:

- `demo` and `rand` are your input data frames.
- We use the `merge` function to merge the data frames.
- The `by` parameter specifies the key variable by which to merge the data frames, which in this case is `usubjid`.

After running this code, the merged data frame `dm` will contain the result of the merge operation, similar to what you would achieve with the SAS code you provided.

# Codes Development

## Prompt :

Write a **R script** that groups data by "trta" and "avisitn" columns in "advs.xpt" and calculates the mean of the "aval" column for each group at parmcd = 'WEIGHT'

## Output :

```
```R
# Load necessary libraries
library(haven)
library(dplyr)

# Read the XPT file
advs <- read_xpt("path/to/advs.xpt")

# Filter the dataset for parmcd == 'WEIGHT'
advs_weight <- advs %>% filter(parmcd == 'WEIGHT')

# Group by 'trta' and 'avisitn' and calculate the mean of 'aval' for each group
mean_aval <- advs_weight %>%
  group_by(trta, avisitn) %>%
  summarise(mean_aval = mean(aval, na.rm = TRUE))

# Print the result
print(mean_aval)
```
```

Make sure to replace `"path/to/advs.xpt"` with the actual file path of your `advs.xpt` dataset. This script will output a data frame with the mean `aval` for each combination of `trta` and `avisitn` where `parmcd` is 'WEIGHT'.

# Codes Development

## Prompt :

Write a **SAS script** that groups data by "trta" and "avisitn" columns in "adv.sxpt" and calculates the mean of the "aval" column for each group at parmcd = 'WEIGHT'

## Output :

```
`` `sas
/* Step 1: Import the advs.xpt file */
libname mylib xport "path_to_your_adv.sxpt";

/* Step 2: Create a SAS dataset from the imported data */
data advs;
 set mylib.advs;
run;

/* Step 3: Filter the data for parmcd = 'WEIGHT' */
data advs_weight;
 set advs;
 where parmcd = 'WEIGHT';
run;

/* Step 4: Group the data by trta and avisitn, and calculate the mean of aval */
proc means data=advs_weight noprint;
 class trta avisitn;
 var aval;
 output out=mean_aval (drop=_TYPE_ _FREQ_) mean=mean_aval;
run;

/* Step 5: Display the result */
proc print data=mean_aval;
run;
```

# Codes Development

Output :

Prompt :

Write a **Python script** that groups data by "trta" and "avisitn" columns in "advs.xpt" and calculates the mean of the "aval" column for each group at parmcd = 'WEIGHT'

```
```bash
pip install pandas xport
```
```

```
```python
import pandas as pd
import xport
```

```
# Load the .xpt file
with open("advs.xpt", "rb") as file:
    data = xport.to_dataframe(file)
```

```
# Filter the data where parmcd is 'WEIGHT'
weight_data = data[data['parmcd'] == 'WEIGHT']
```

```
# Group by 'trta' and 'avisitn' and calculate the mean of 'aval'
grouped_data = weight_data.groupby(['trta', 'avisitn'])['aval'].mean().reset_index()
```

```
# Rename the column for clarity
grouped_data = grouped_data.rename(columns={'aval': 'mean_aval'})
```

```
# Display the result
print(grouped_data)
```
```

# Data Analysis (1)

- Prompt :

Select patients whose sex is male. Use below "Patient Data". Provide me the answers and R codes only.

“Patient Data”:

```
{ 'patient' : '01-001', 'sex' : 'male', 'age' : 25, 'race' : 'White'},
{ 'patient' : '01-002', 'sex' : 'female', 'age' : 28, 'race' : 'Asian'},
{ 'patient' : '02-001', 'sex' : 'female', 'age' : 50, 'race' : 'White'},
{ 'patient' : '02-002', 'sex' : 'male', 'age' : 40, 'race' : 'Black'}
```

- Output :

```
1. {'patient': '01-001', 'sex': 'male', 'age': 25, 'race': 'White'}
2. {'patient': '02-002', 'sex': 'male', 'age': 40, 'race': 'Black'}
```

```
Create data frame
```

```
patient_data <- data.frame(
 patient = c("01-001", "01-002", "02-001", "02-002"),
 sex = c("male", "female", "female", "male"),
 age = c(25, 28, 50, 40),
 race = c("White", "Asian", "White", "Black"))
```

```
Filter male patients
```

```
male_patients <- subset(patient_data, sex == "male")
```

# Data Analysis (2)

- Prompt :

Count below “Patient Data” by race

“Patient Data”:

```
{ 'patient' : '01-001', 'sex' : 'male', 'age' : 25, 'race' : 'White'},
{ 'patient' : '01-002', 'sex' : 'female', 'age' : 28, 'race' : 'Asian'},
{ 'patient' : '02-001', 'sex' : 'female', 'age' : 50, 'race' : 'White'},
{ 'patient' : '02-002', 'sex' : 'male', 'age' : 40, 'race' : 'Black'}
```

- Output :

Certainly! Let's count the number of patients by race in the given "Patient Data":

- White: 2

- Asian: 1

- Black: 1

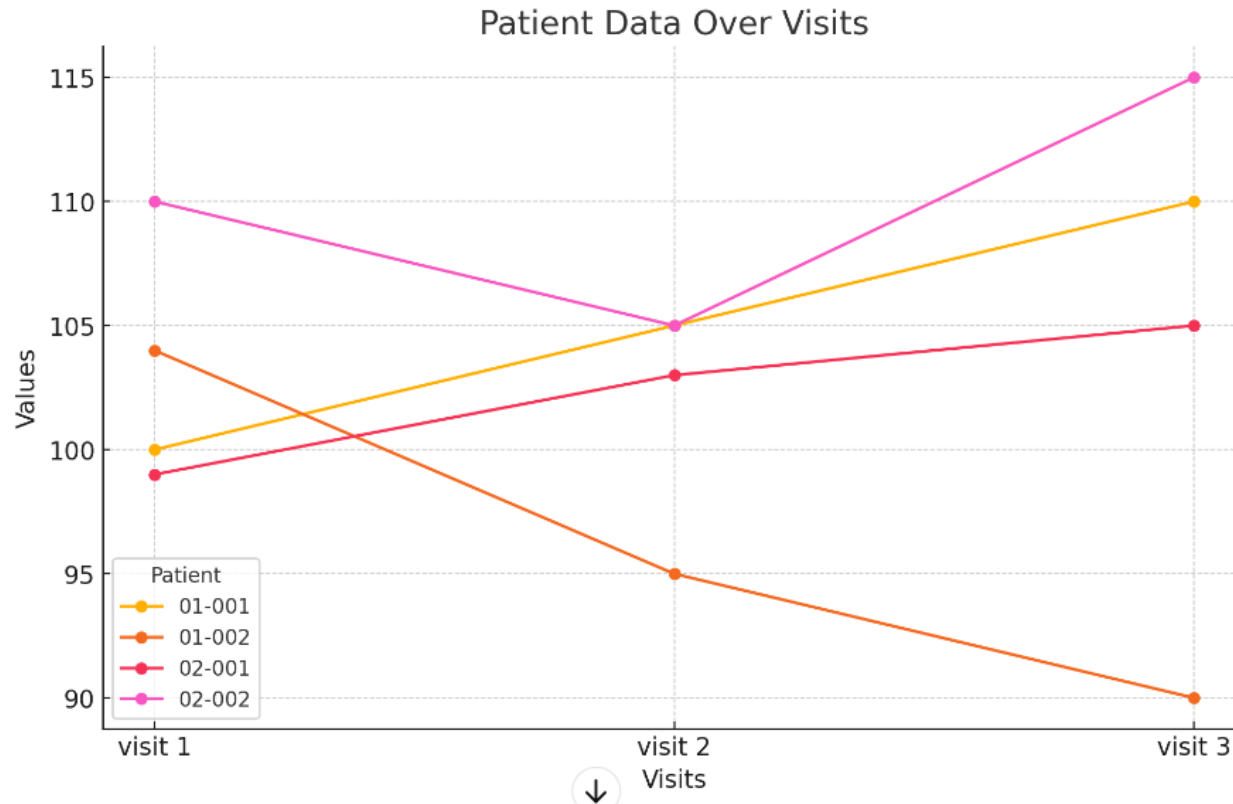
# Data Visualization

- Prompt :  
Create line graph of “Patient Data”

“Patient Data”:

```
{ 'patient' : '01-001', 'visit 1' : 100, 'visit 2' : 105, 'visit 3' : 110},
{ 'patient' : '01-002', 'visit 1' : 104, 'visit 2' : 95, 'visit 3' : 90},
{ 'patient' : '02-001', 'visit 1' : 99, 'visit 2' : 103, 'visit 3' : 105},
{ 'patient' : '02-002', 'visit 1' : 110, 'visit 2' : 105, 'visit 3' : 115}
```

- Output :



# Patient Profile Development

- Prompt :

Use below data to provide patient profiles:

```
{ 'patient' : '01-001', 'sex' : 'male', 'age' : 25, 'race' : 'White'}
```

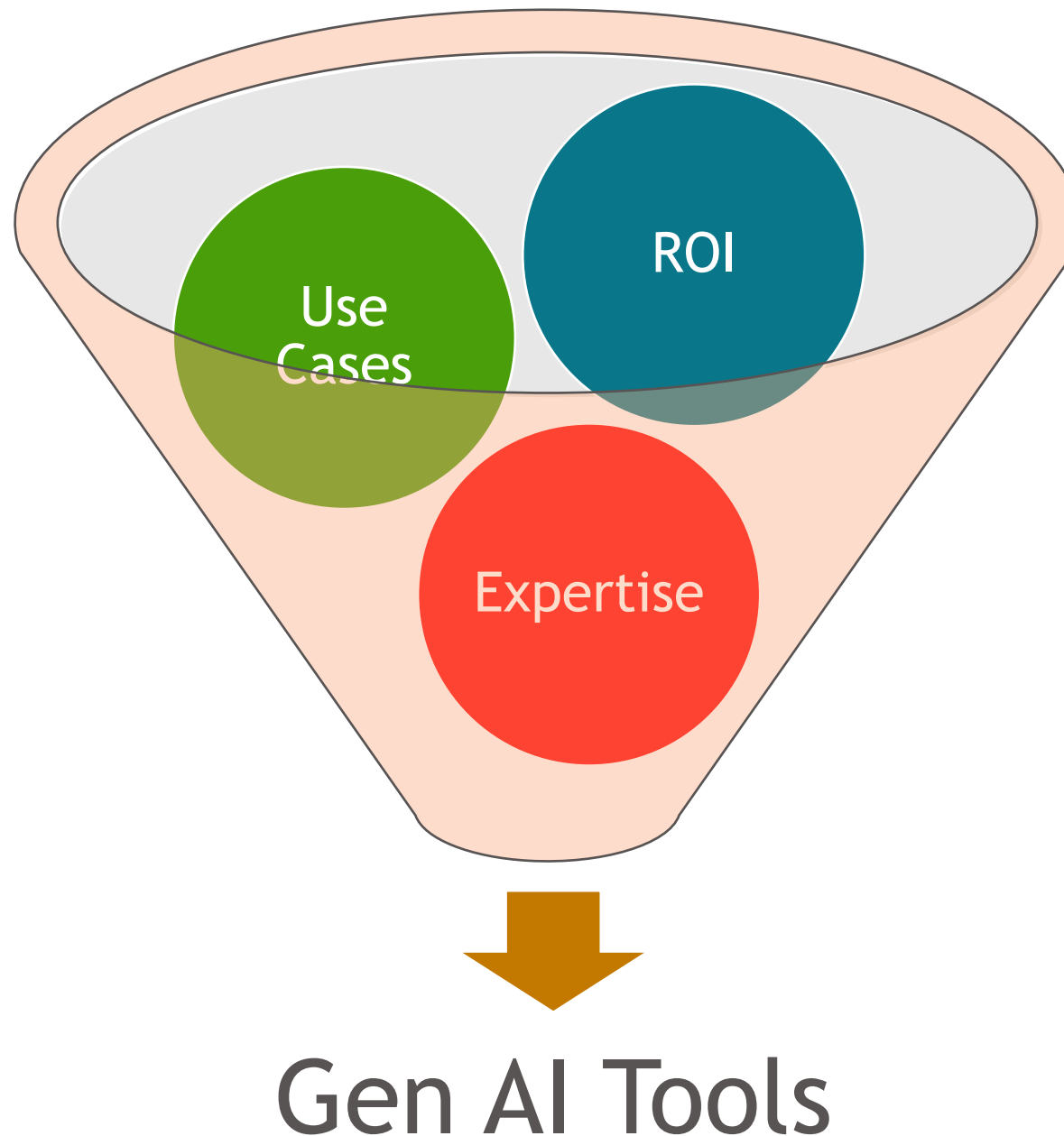
- Output :

Based on the provided data, here is a patient profile for '01-001':

- Patient ID: 01-001
- Sex: Male
- Age: 25
- Race: White

This patient is a 25-year-old male of White race. Please let me know if you need any additional information or if you would like to create profiles for more patients.

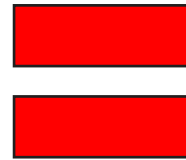
# Use Cases driven Gen AI Tools



# Gen AI Process

- Validation?
- Human intervention
- Final output or source of input?

# ChatGPT is like a very smart, capable, recently graduated intern/new hire.

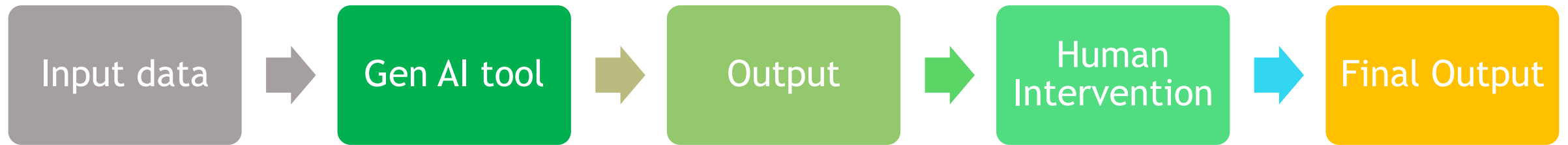


# What will we do with the results from an intern?



**The output of Gen AI is  
not the final result,  
but the beginning of new process.**

# Gen AI driven Process



# People : Biometrics team powered with Gen AI

-Super Biometrics Team

# ChatGPT Users vs Non-Users Comparison in Boston Consulting

- Results of ChatGPT Users
  - finished 12.2% more tasks on average
  - completed tasks 25.1% more quickly
  - produced 40% higher quality results



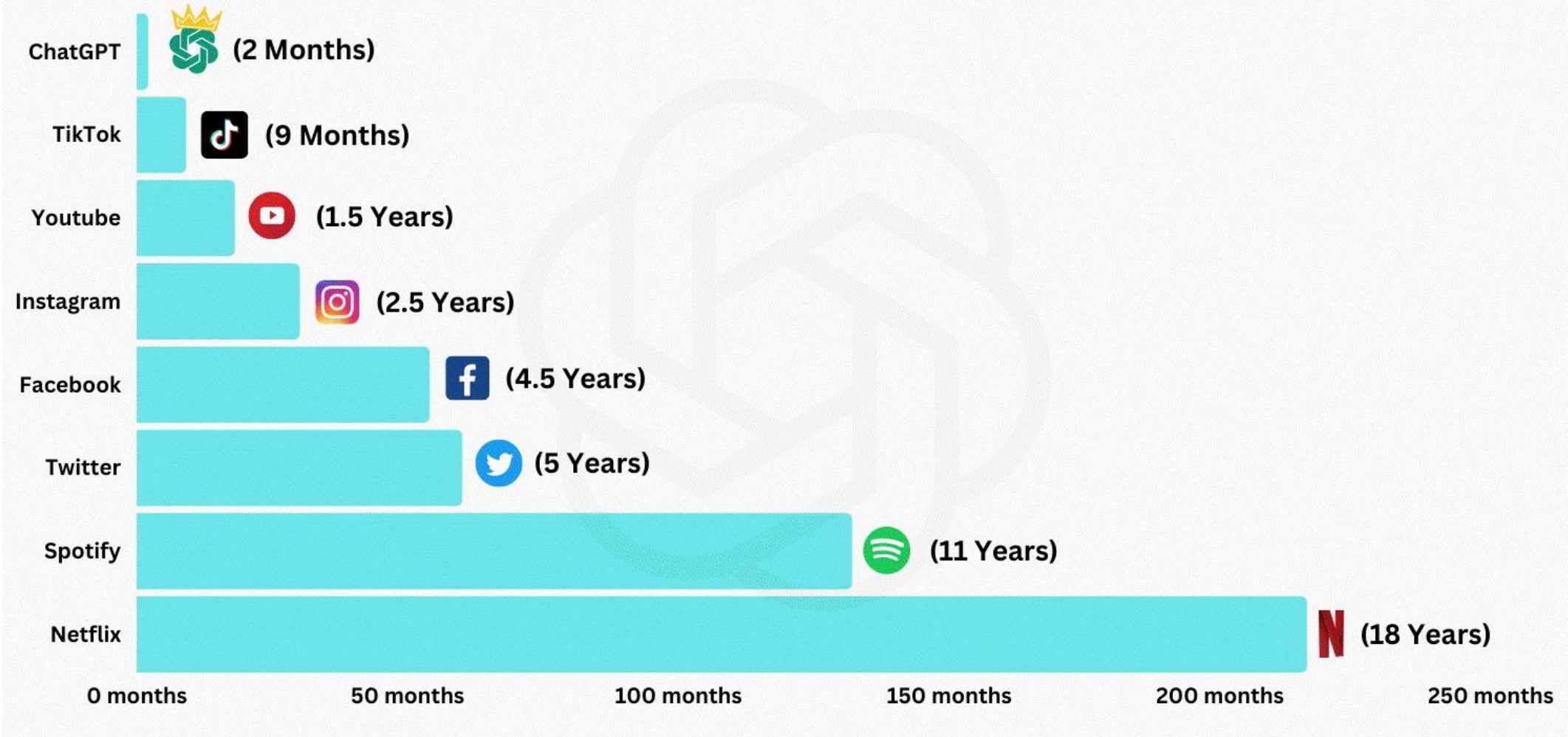
# Gen AI gives us a Head Start. We don't need to start from the blank page.



# Future of Gen AI and US

# ChatGPT Popularity

## Road To 100 Million Users For Various Platforms



ChatGPT gained 100M users within 2 months since its release.

# Gen AI has arrived



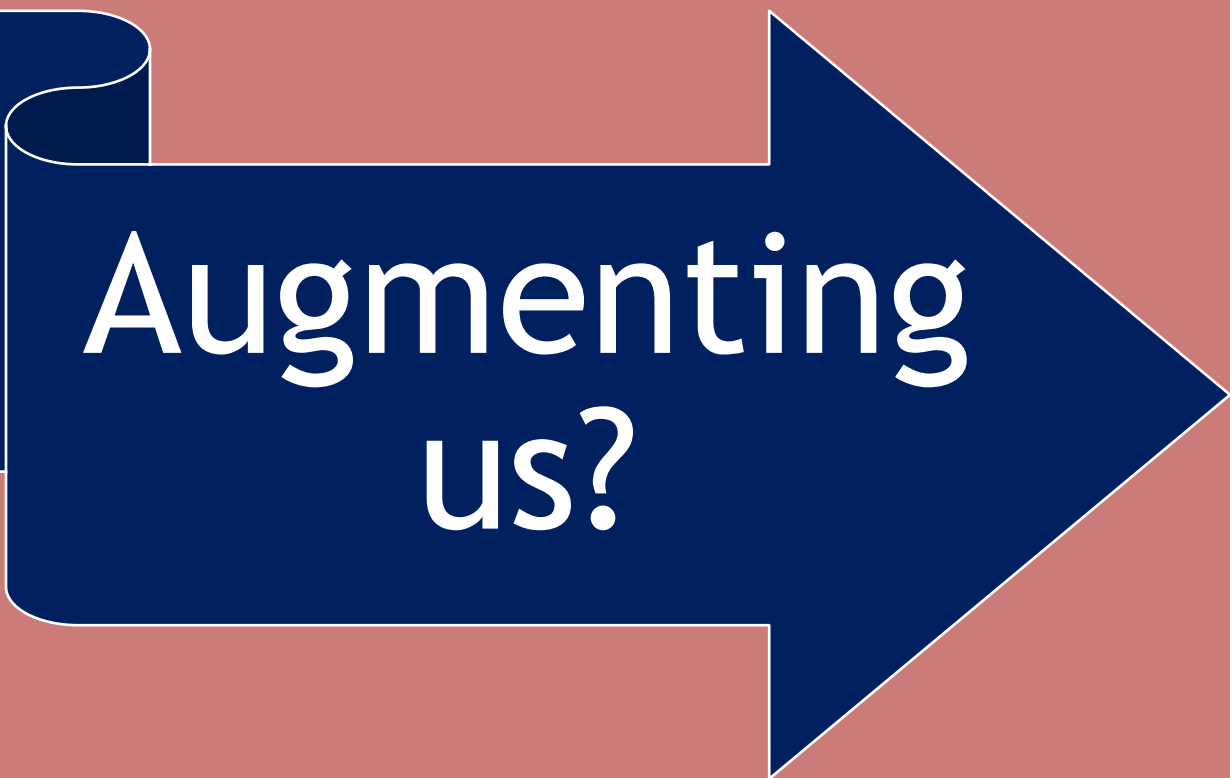
# Attacking vs Understanding



# What is Gen AI for us?

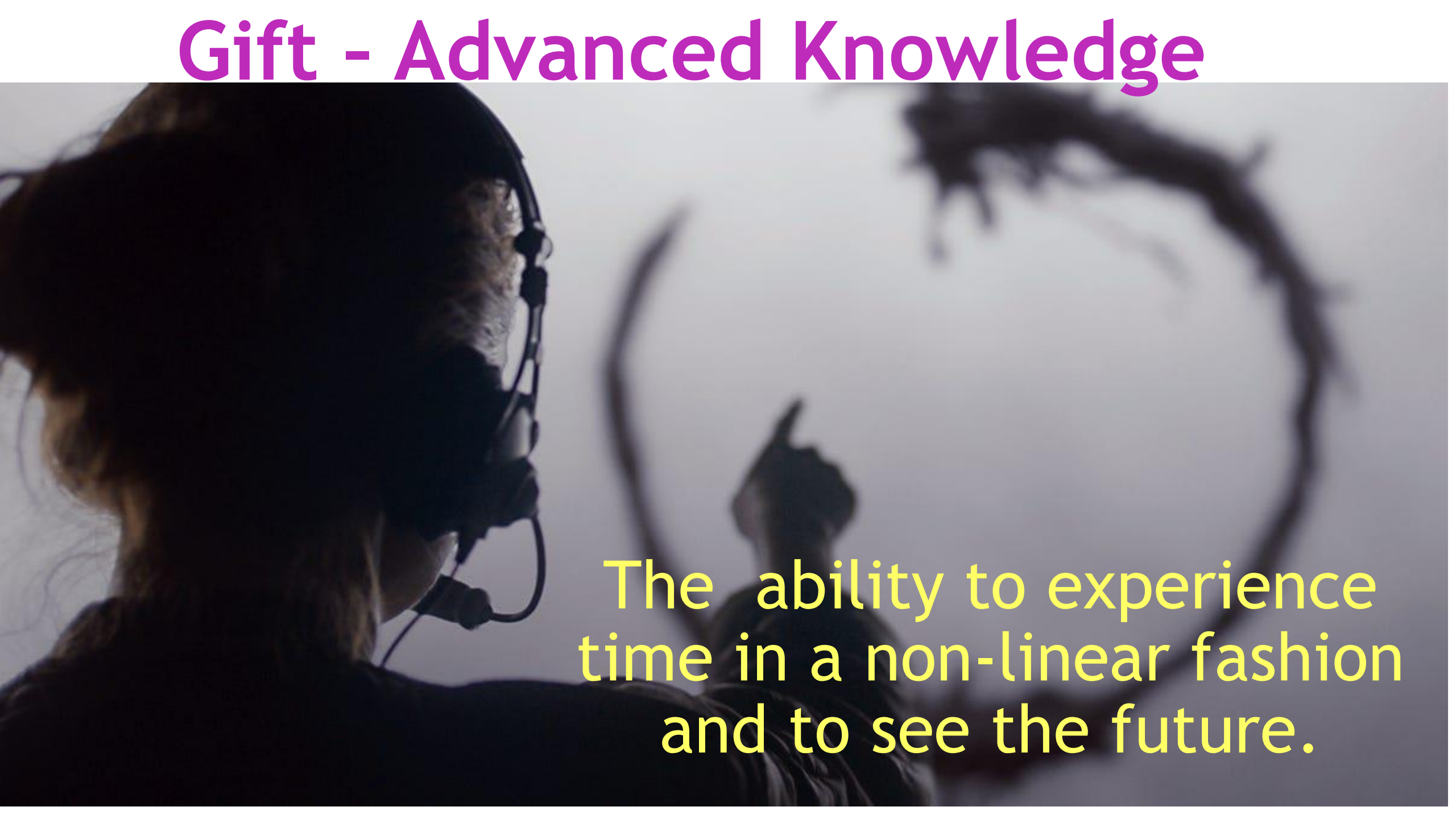


Replacing  
us?



Augmenting  
us?

# Gift - Advanced Knowledge

A person is shown from the side, wearing a VR headset and holding a hand controller. They are looking at a large, glowing, translucent globe that appears to be floating in the air. The person's hand is raised, pointing towards the globe. The background is a soft, out-of-focus light blue.

The ability to experience  
time in a non-linear fashion  
and to see the future.

# Future of Gen AI & Biometric team?

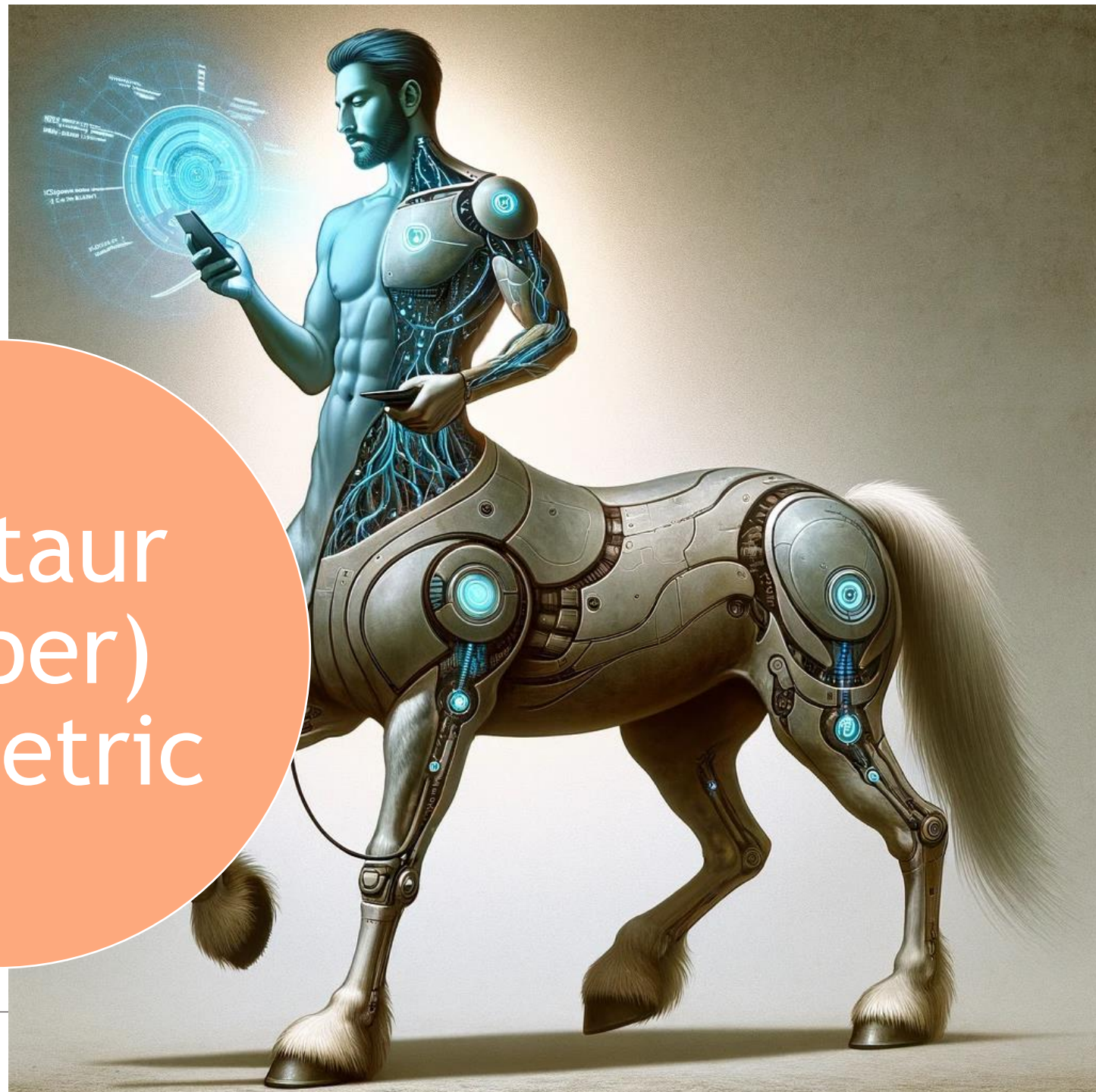
Gen AI



Biometric

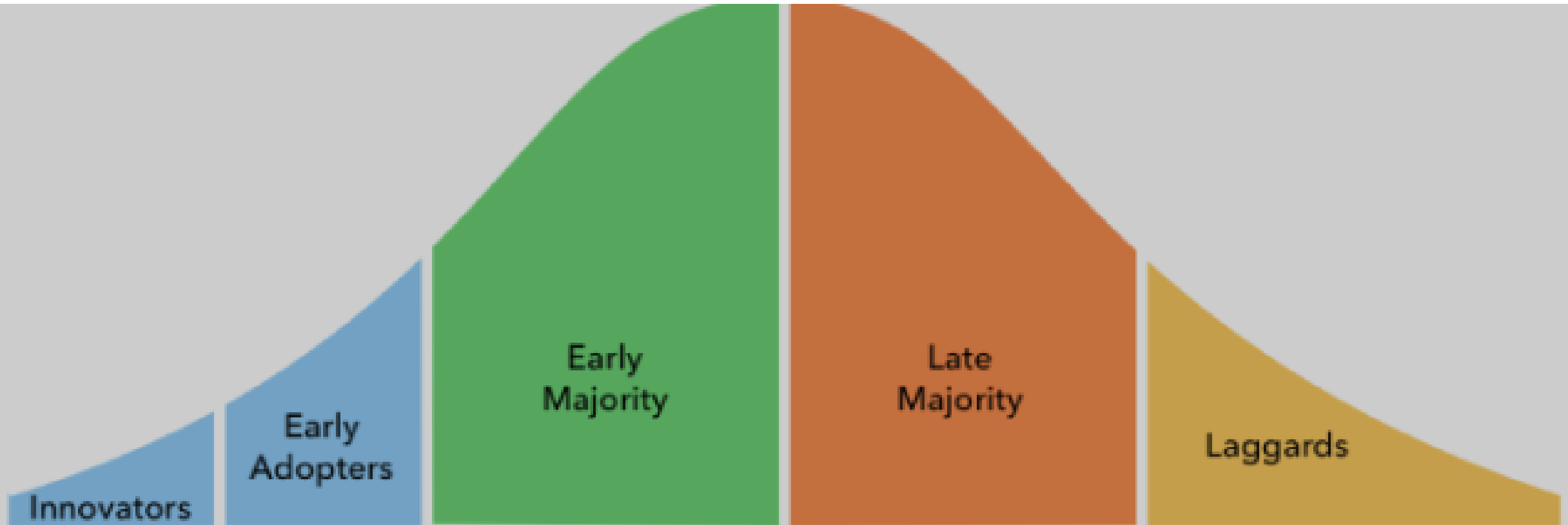


Centaur  
(Super)  
Biometric



# Discussion :

Where do we want our Biometric team to be in Gen AI Adopter?



# Q & A