

# Early Prediction of Alzheimer's Disease Using Social Determinants of Health





# Group Team Members

01. Branley Ope
02. Brian Kipgnetich
03. Cynthia Atieno
04. Geoffrey Mwangi
05. Linet Mazsusa
06. Maureen Wanjeri
07. Mercy Silali

DSF-PT07-Group 12

# Table of Contents

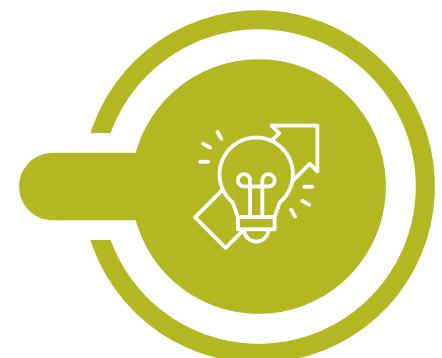
01 Introduction



02 Table Of Contents



03 Project Overview



04 Problem Statement



Modeling 08



Evaluation 09

05 Business Understanding



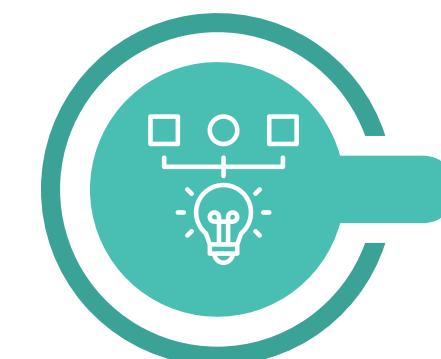
06 Data Understanding



Deployment 11

Challenges 12

07 Data Preparation



Recommendation & Conclusion 13



Thank You & QnA 14

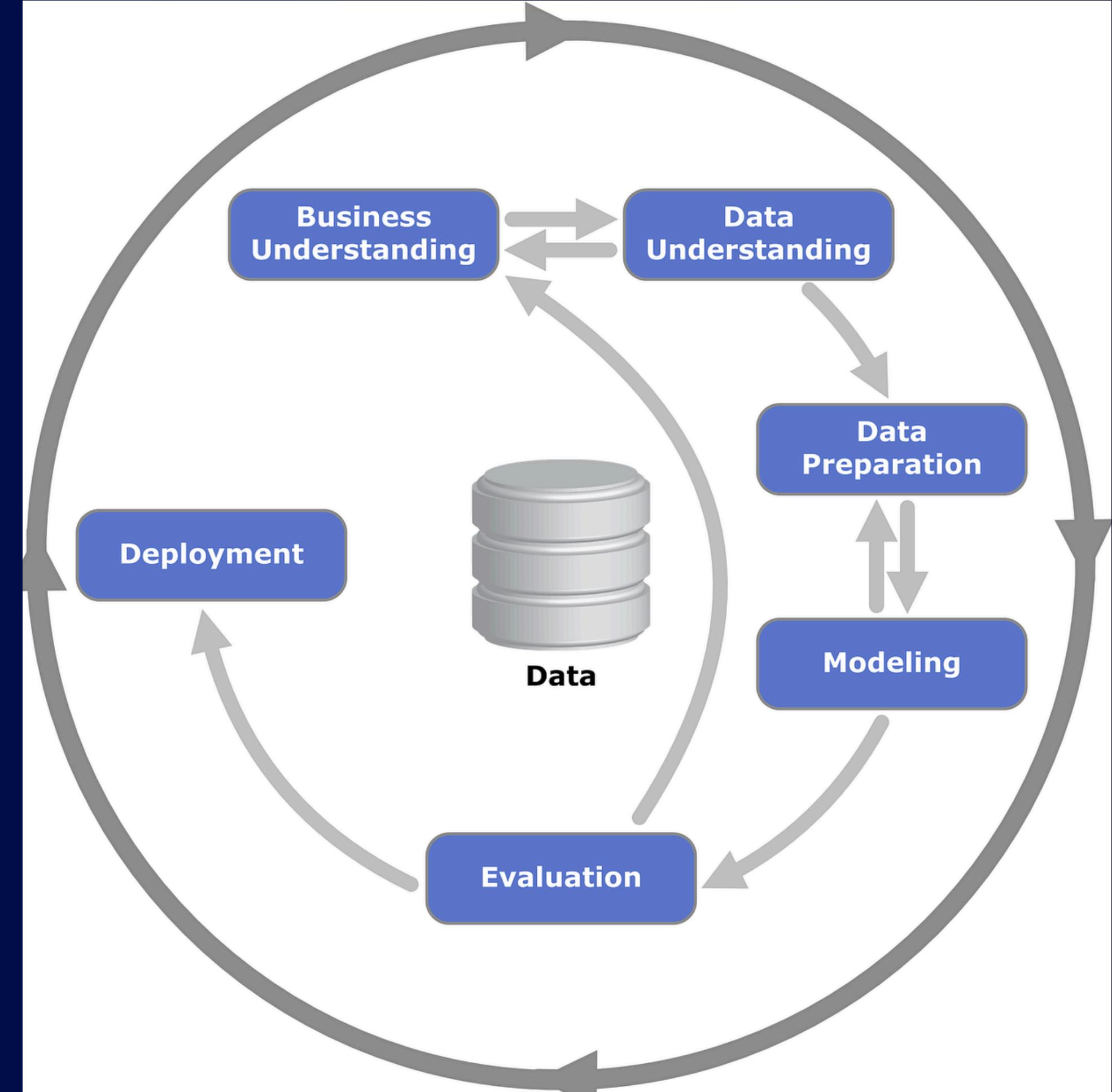
## Introduction Cont..

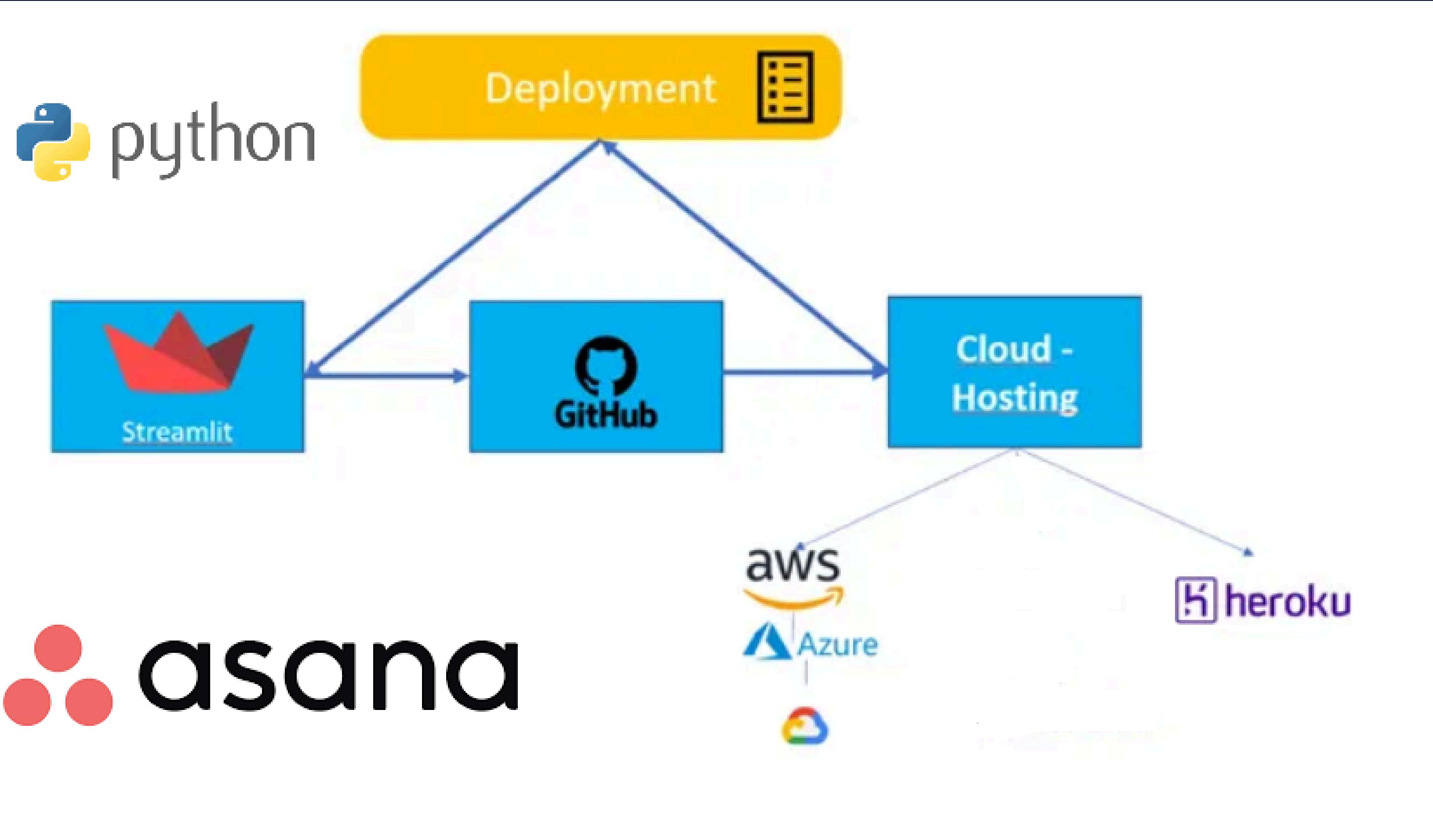
### Methodology - CRISP-DM

Cross Industry Standard Process For  
DATA MINING

### DISCLAIMER!

This project is solely for  
educational purposes and  
should not be used or relied  
upon by medical professionals  
as a diagnostic tool.







# ALZHEIMER'S DISEASE A . D



# Problem Statement

**Alzheimer's Disease (AD) and Alzheimer's-related dementias (ADRD)** are growing global health challenges, with early detection being critical for effective intervention.

However, existing screening methods are complex, time-consuming, and difficult to scale.

This **limits timely diagnosis and contributes to disparities** in care, particularly among marginalized populations.

Our project is inspired by a **competition** aimed at improving early prediction of AD/ADRD using social determinants of health (SDOH).

Sponsored by the **National Institute on Aging (NIA)**, part of the **National Institutes of Health (NIH)**, and supported by **NASA**.

This initiative emphasizes the development of innovative models that can **predict cognitive decline 4 to 9 years in advance**, reducing diagnostic delays and promoting equitable healthcare access.



## WHO Fact Sheets 2023

Alzheimer's Disease (AD) and related dementias (ADRD) affect over **55 million** people worldwide, with nearly **10 million** new cases annually.



## Mexican Health and Aging Study (MHAS)

Our project leverages data from the Mexican Health and Aging Study (MHAS), a nationally representative longitudinal survey of adults in Mexico aged 50 and older.



How can we detect Alzheimer's earlier, using accessible and non-invasive data, to reach a broader population and potentially slow the disease's progression?

# Ethical Considerations



- This project adheres to strict ethical guidelines to ensure the protection of individual privacy and confidentiality.
- All personal identifiers were removed from the dataset, and the data was fully anonymized prior to being accessed for analysis. The anonymization process ensures that no individual can be identified, directly or indirectly, through the dataset used in this project.
- This research complies with data protection regulations and ethical standards, prioritizing respect for the rights and privacy of all individuals whose data contributed to this study.
- The findings and insights derived from this project are presented in aggregate, focusing solely on trends and patterns at the population level to advance scientific understanding without compromising individual confidentiality.

# BUSINESS UNDERSTANDING

## PROJECT OBJECTIVES

### Main Objective

- Improve early prediction of Alzheimer's disease and related dementias (AD/ADRД) using social determinants of health.

### Specific Objectives

- Develop an Innovative Predictive Model
- Identify Key SDOH Indicators
- Enhance Accessibility and Equity
- Provide actionable insights

## SCOPE

- Predict cognitive outcomes for 2016 and 2021 based on data from 2003 and 2012.
- Utilize social, economic, and health-related features for prediction.

## IMPACT

- Enables early diagnosis and preventive care, particularly for underserved populations.
- Promotes proactive interventions and improved health outcomes.

# BUSINESS UNDERSTANDING

## STAKEHOLDERS AND SUCCESS METRICS

### Key Stakeholders

- **Healthcare Providers:** Identify high-risk patients for early care.
- **Public Health Organizations:** Drive policy and resource allocation.
- **Researchers:** Expand insights into SDOH impacts on cognitive health.
- **Patients/Families:** Support proactive health mana

### Success Metrics

- **Accuracy:** RMSE within acceptable thresholds.
- **Equity:** Fair performance across demographic groups.
- **Interpretability:** Transparent outputs for stakeholders.
- **Reliability:** Generalizable predictions across populations

# Data Understanding

## Data Sets

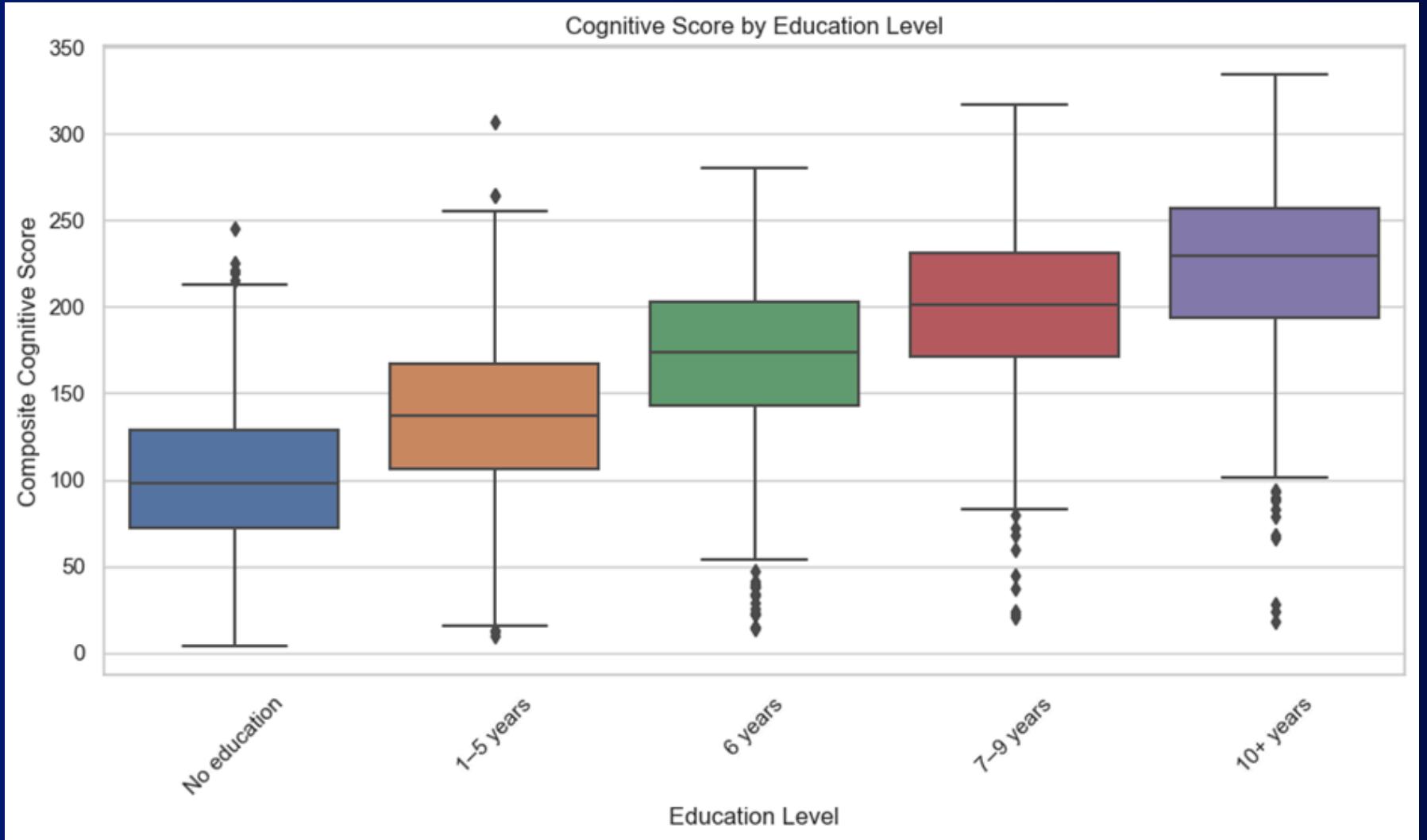
- Train\_features.csv
- Test\_features.csv
- Train\_labels.csv
- submission\_format.csv

## Assessment

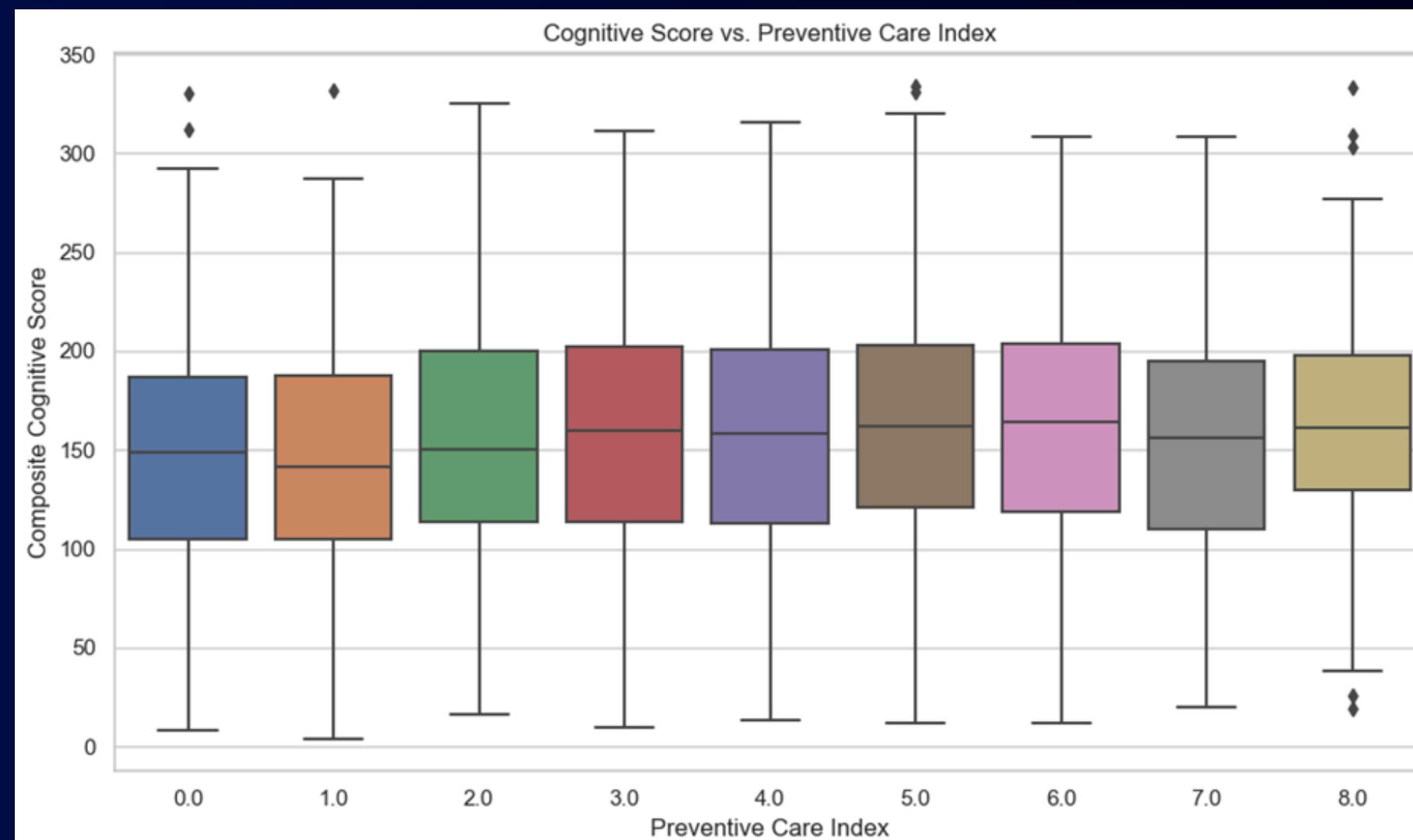
- Data Completeness and Quality
- Feature Diversity and Cardinality
- Temporal Analysis
- Bias and Representation

## Data Description

- The data set includes social determinants of health collected in **2003 and 2012** for **3,276** unique individuals, alongside cognitive assessment scores from **2016 and 2021**, ranging from **4 to 334**.
- It features **demographic, health, lifestyle, and socioeconomic factors**, enabling the development of predictive models to forecast cognitive decline over a **4- to 9-year horizon**.



- Illustrates how **education level** correlates with cognitive function.
- Higher education levels are associated with higher cognitive scores, supporting the recommendation to promote educational initiatives.

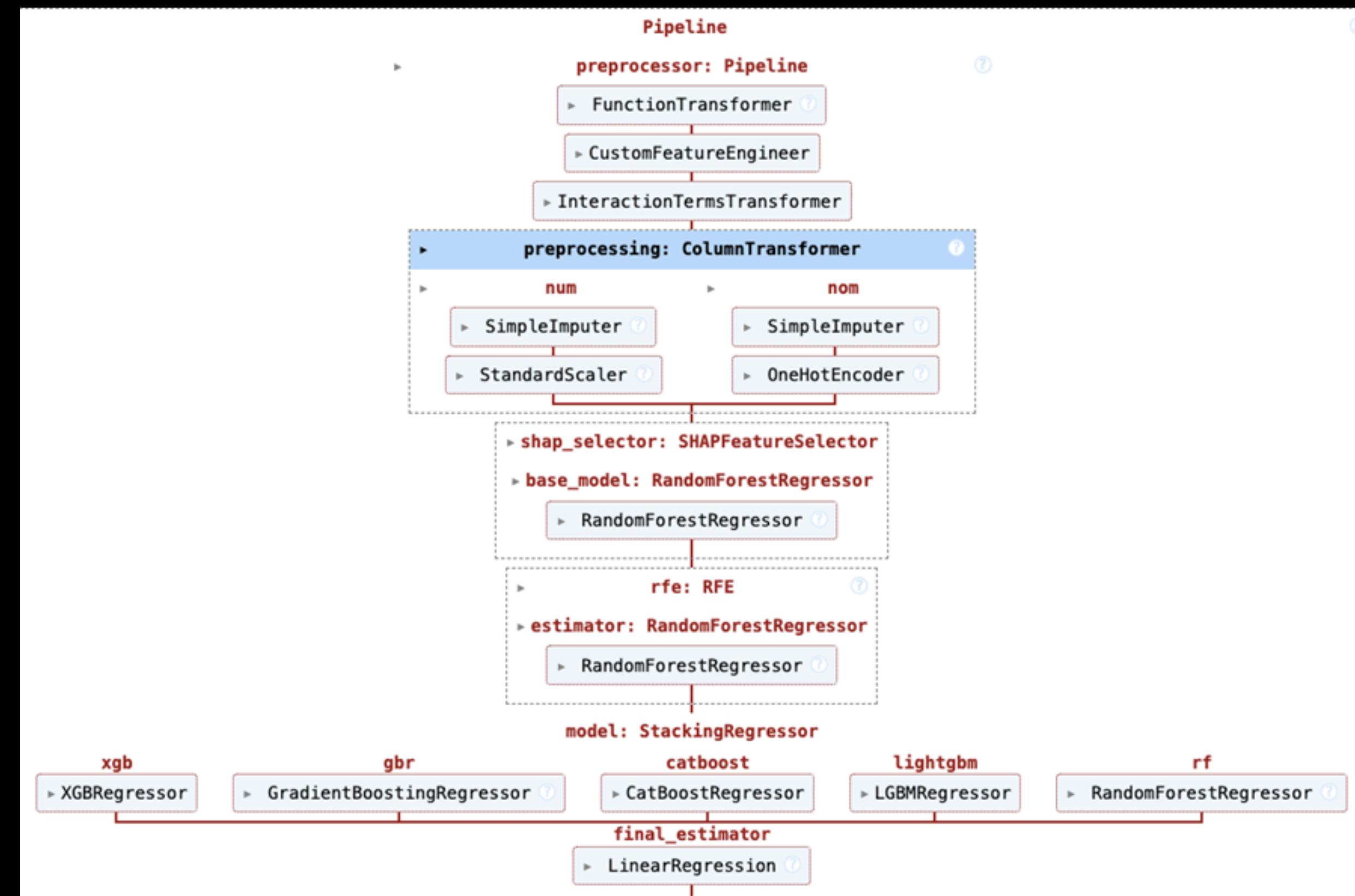
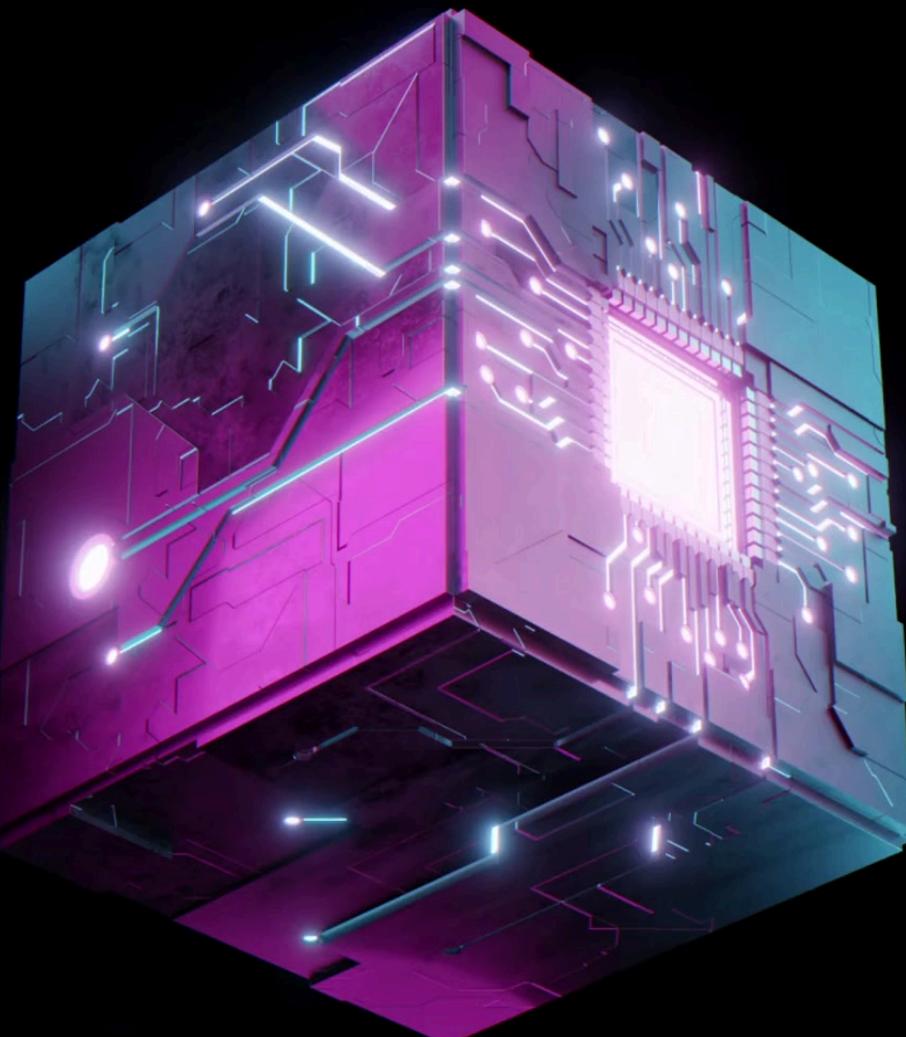


- Illustrates the effect of **preventive care index**, which is a sum of respondents hospital trips to screen for chronic conditions such as diabetes and hypertension, or get vaccines, get dental checkups etc.
- Higher preventive care participation is linked to better cognitive scores, supporting the recommendation to enhance preventive healthcare access.

# DATA PREPARATION



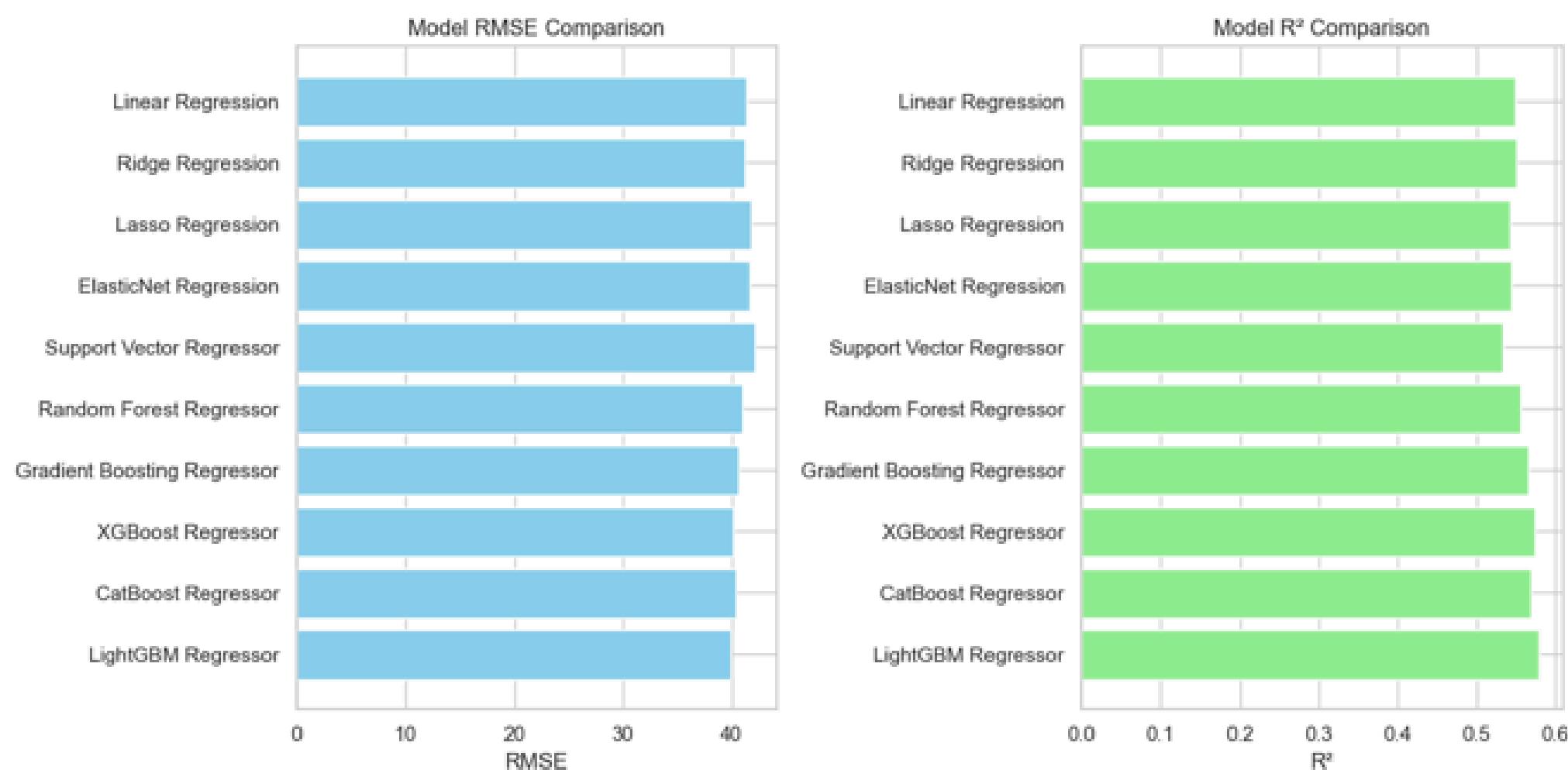
# MODELING



# Model performance

| Model                       | RMSE    | R <sup>2</sup> | MAE     |
|-----------------------------|---------|----------------|---------|
| Linear Regression           | 41.3544 | 0.5492         | 32.6100 |
| Ridge Regression            | 41.2959 | 0.5505         | 32.5292 |
| Lasso Regression            | 41.7082 | 0.5414         | 32.6961 |
| ElasticNet Regression       | 41.5797 | 0.5443         | 32.7976 |
| Support Vector Regressor    | 42.0827 | 0.5332         | 33.0156 |
| Random Forest Regressor     | 41.0365 | 0.5561         | 32.6984 |
| Gradient Boosting Regressor | 40.6148 | 0.5652         | 32.2239 |
| XGBoost Regressor           | 40.1649 | 0.5747         | 31.9913 |
| CatBoost Regressor          | 40.3888 | 0.5700         | 32.1401 |
| LightGBM Regressor          | 39.9189 | 0.5799         | 31.6357 |

# EVALUATION



# Model Interpretability

## Progression in Physical Limitations (adl\_iadl\_progression)



- Impact: Worsening physical abilities predict cognitive decline.
- Actionable Insight: Implement physical health programs.

## Household Income Changes (hincome\_change)

- Impact: Economic stability influences cognitive function.
- Actionable Insight: Support financial well-being initiatives.

## Parental Education Levels (rafeduc\_m, rameduc\_m)



- Impact: Reflects intergenerational effects on cognitive health.
- Actionable Insight: Address educational disparities.



## Social Engagement Activities (social\_engagement\_12)

- Impact: Active social life contributes positively to cognition.
- Actionable Insight: Encourage community participation.

## Preventive Care Index (preventive\_care\_index\_12)



- Impact: Regular health check-ups correlate with better cognitive outcomes.
- Actionable Insight: Promote preventive healthcare access.

# Deployment

## Streamlit

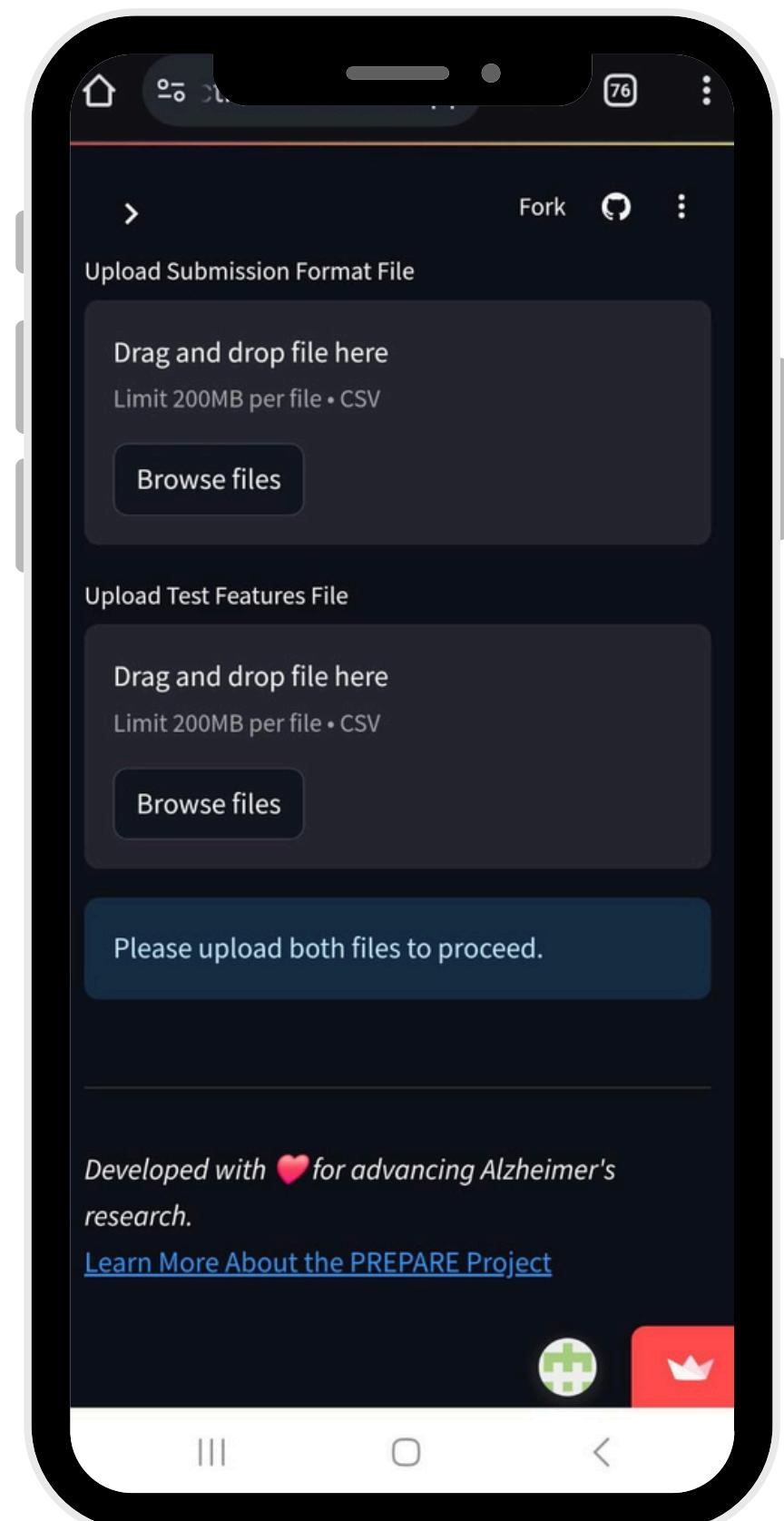
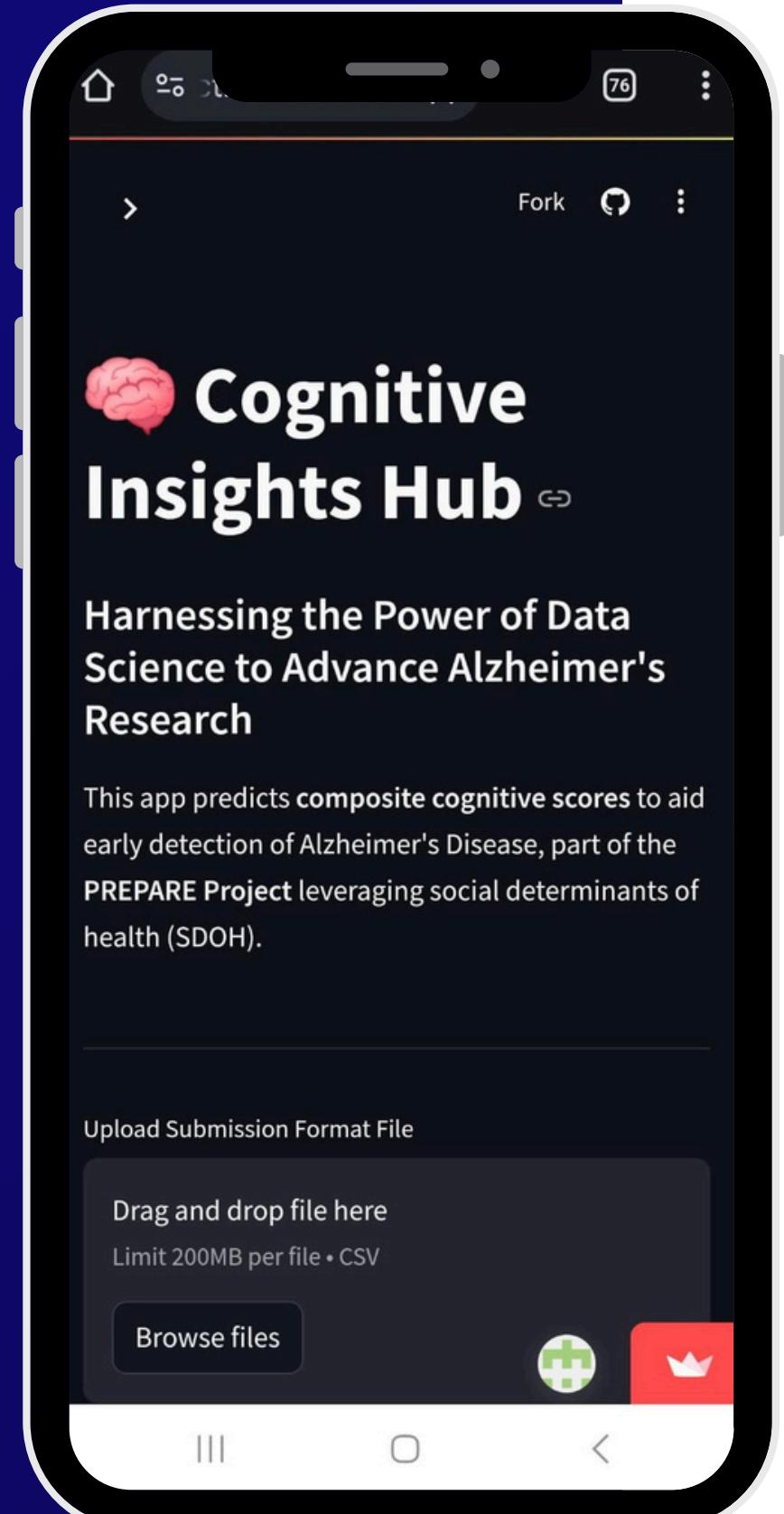
Click below link to access the application

[Cognitive Insights Hub](#)

## Test Data - Google Drive

Click below link to access the test data for uploading in  
the application

[Cognitive Insights Hub test data](#)



**Upload Your Files**

Instructions:

1. Upload the Submission Format file (CSV).
2. Upload the Test Features file (CSV).
3. View detailed predictions and insights.

**Prediction Results**

|    | uid  | year  | composite_score | age_03   | urban_03    | married_03    | n_mar_03 | edu_gru_03    | n_living_child_03 | migration_03 | glob_hlth_03 | adl_dress_03 | adl_walk_03 | adl_bath_03 |
|----|------|-------|-----------------|----------|-------------|---------------|----------|---------------|-------------------|--------------|--------------|--------------|-------------|-------------|
| 4  | afnb | 2,021 | 194             | None     | None        | None          | None     | None          | None              | None         | None         | None         | None        | None        |
| 5  | ajfh | 2,016 | 139             | None     | None        | None          | None     | None          | None              | None         | None         | None         | None        | None        |
| 6  | ajfh | 2,021 | 139             | None     | None        | None          | None     | None          | None              | None         | None         | None         | None        | None        |
| 7  | ajvq | 2,021 | 219             | 2. 60-69 | 1. 100,000+ | 1. Married or | 1        | 4. 10+ years  | 1. 1 or 2         | 0            | None         | None         | 0           |             |
| 8  | akbn | 2,016 | 122             | None     | None        | None          | None     | None          | None              | None         | None         | None         | None        | None        |
| 9  | akcw | 2,021 | 174             | None     | None        | None          | None     | None          | None              | None         | None         | None         | None        | None        |
| 10 | akmb | 2,021 | 190             | 1. 50-59 | 1. 100,000+ | 1. Married or | 1        | 3. 7-9 years  | 1. 1 or 2         | 0            | 4. Fair      | 0            | 0           |             |
| 11 | akow | 2,016 | 116             | 1. 50-59 | 1. 100,000+ | 1. Married or | 1        | 0. No educati | 4. 7+             | 0            | 5. Poor      | 0            | 0           |             |
| 12 | akup | 2,021 | 127             | 1. 50-59 | 1. 100,000+ | 1. Married or | 2        | 0. No educati | 2. 3 or 4         | 0            | 5. Poor      | 1            | 0           |             |
| 13 | albo | 2,016 | 129             | 1. 50-59 | 0. <100,000 | 1. Married or | 1        | 1. 1-5 years  | 2. 3 or 4         | 0            | 3. Good      | 0            | 0           |             |

 [Download Detailed Predictions](#)**Summary Insights**

Average Composite Score

**156.68****Distribution of Predicted Scores**

# CHALLENGES

## Data Limitations

Missing data and incomplete data

## Bias in Data

Potential under-representation of certain populations, leading to biased predictions.

## Ethical Concerns

Balancing data privacy and security while using sensitive health information.

## Interpretability

Ensuring the predictive model's results are transparent and understandable to stakeholders.

## Resource Constraints

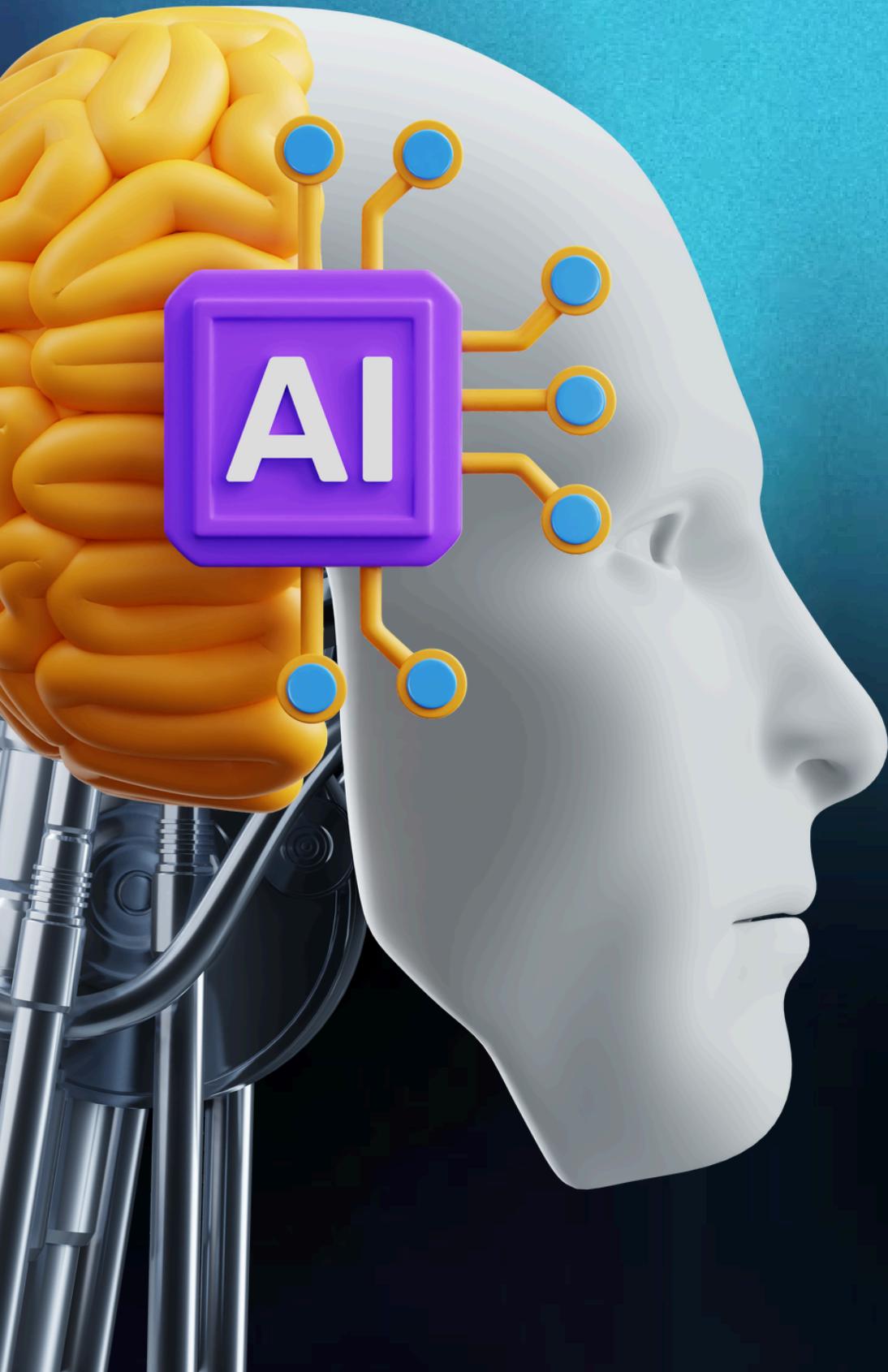
Limited access to computational resources or expertise for complex model development.

## Model Generalization

Ensuring the model performs well across diverse demographics and settings.



# Conclusion



SDOH FACTORS ARE SIGNIFICANT PREDICTORS OF COGNITIVE DECLINE.

UTILIZING SDOH DATA CAN MAKE EARLY DETECTION MORE ACCESSIBLE AND LESS RESOURCE-INTENSIVE.

# Recommendations

## Boost Social Engagement

**Strategy:** Facilitate social clubs and activities.

**Expected Outcome:** Reduce isolation and enhance mental well-being.

## Address Economic Stability

**Strategy:** Implement financial support programs.

**Expected Outcome:** Mitigate stressors affecting cognition.

## Enhance Preventive Healthcare Access

**Strategy:** Increase availability of health screenings.

**Expected Outcome:** Early identification of health issues impacting cognition.

## Promote Physical Activity

**Strategy:** Develop exercise programs for older adults.

**Expected Outcome:** Improve physical and cognitive health.

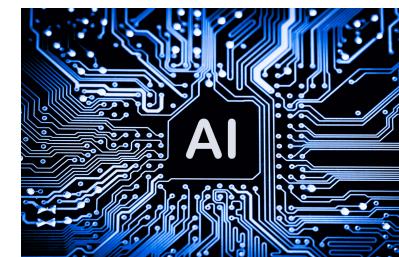
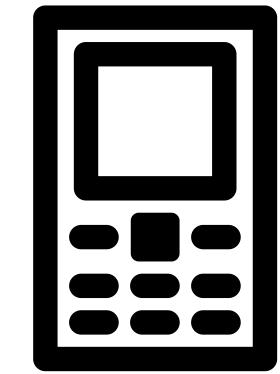
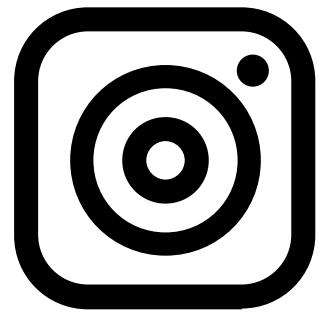
## Support Educational Initiatives

**Strategy:** Provide lifelong learning opportunities.

**Expected Outcome:** Strengthen cognitive reserves.



# Enabling Technology Trends



# Next Steps

Alzheimer's Disease is a complex challenge, but by harnessing the power of data science and focusing on social determinants of health, we can make strides toward early detection and intervention.



01

## Expand the Model

Incorporate additional data sources to improve accuracy and applicability across different populations

02

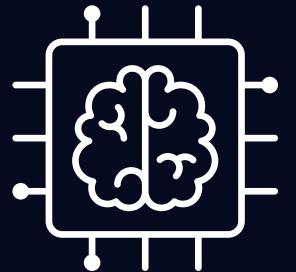
## Pilot Implementation

Test the model in selected communities to validate real-world effectiveness.

03

## Collaborate with Stakeholders

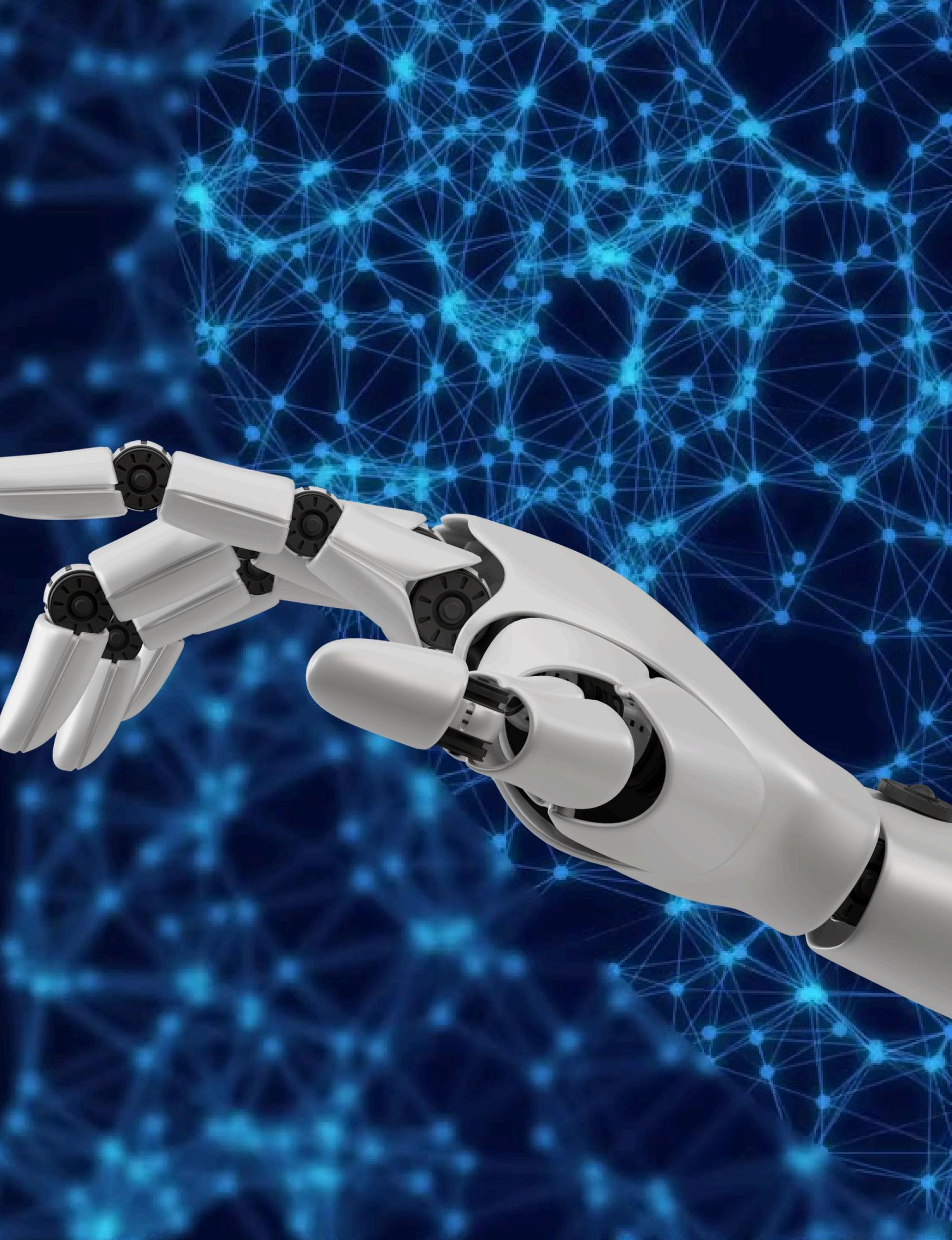
Engage with healthcare providers, policymakers, and community leaders to tailor interventions

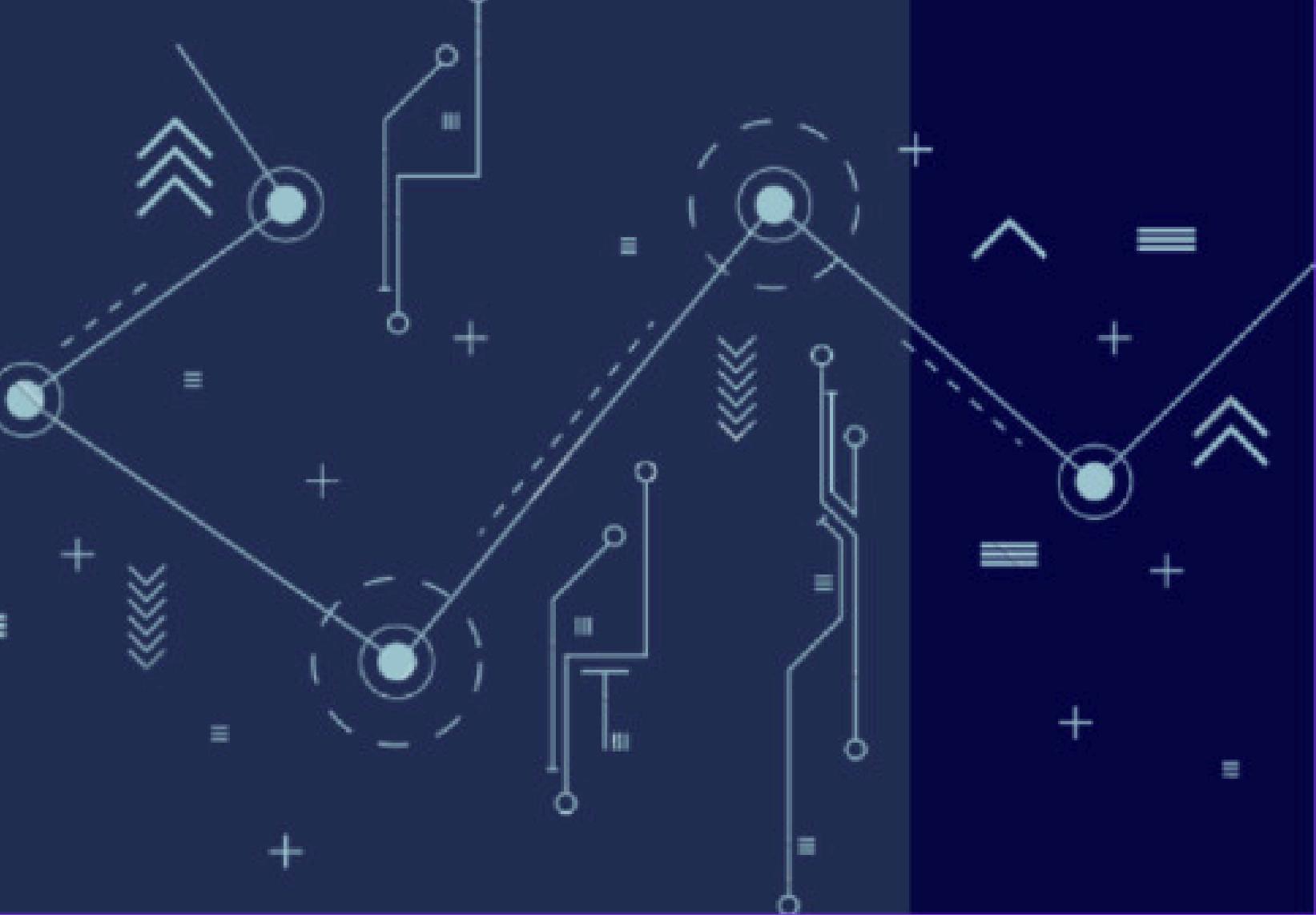


## Acknowledgement

*The MHAS (Mexican Health and Aging Study) is partly sponsored by the National Institutes of Health/National Institute on Aging (grant number NIH R01AG018016) in the United States and the Instituto Nacional de Estadística y Geografía (INEGI) in Mexico.*

*Data files and documentation are public use and available at [www.MHASweb.org](http://www.MHASweb.org).*





## DSF-PT07 PHASE 5- GROUP 12

Branely Ope

Brian Kipngetich

Cynthia Atieno

Geoffrey Mwangi

Linet Maz'susa

Maureen Wanjeri

Mercy Silali



# THANK YOU

# Q&A