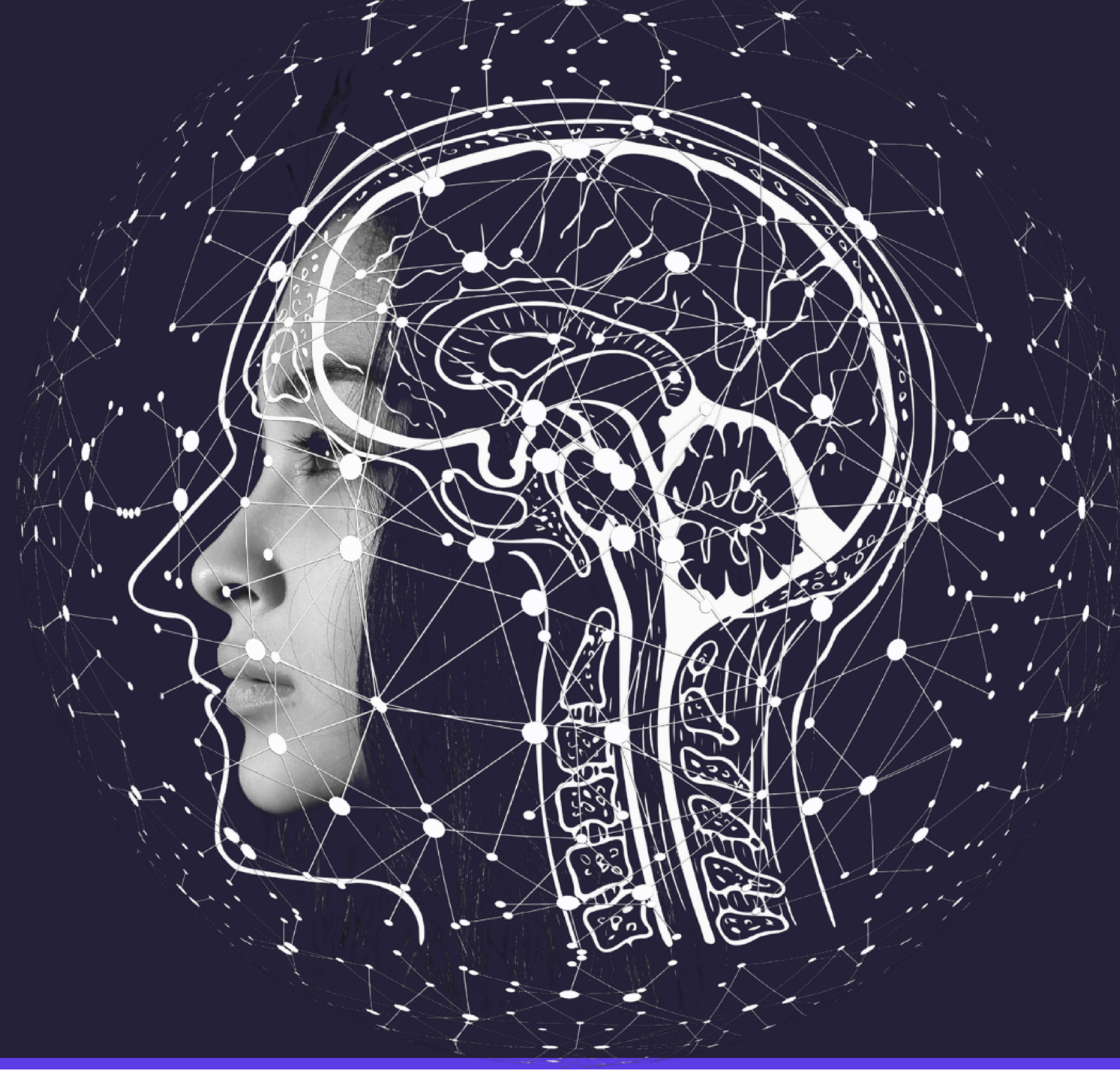
LIBRA A.I. Technologies



A foundation model for single-cell transcriptomics in Alzheimer's disease

Teaching AI to recognize Alzheimer's disease, one brain cell at a time

A. Balomenos[1], C. Petrou[1], T. Siozos[1], I. Charalampopoulos[2], T. Calogeropoulou[3], Y. Kopsinis[1]

1 Libra AI Technologies, Athens | 2 University of Crete / IMBB-FORTH | 3 NHRF, Athens

01 Introduction

Alzheimer's disease affects over **50 million people worldwide**, slowly damaging memory and thinking as brain cells break down. Every brain cell behaves a little differently, and those differences hold clues to the disease.

Single-cell RNA sequencing lets scientists read which genes are switched on inside each individual cell, like a snapshot of what that cell is doing. But snapshots from different labs are hard to compare, and no AI model had been built specifically to understand Alzheimer's biology, until now.

02 Our approach

96 public studies were combined into one harmonized dataset, then pretrained ADPT, a transformer AI (the same kind of technology behind large language models), to learn the language of genes in Alzheimer's-affected brains.
The gene vocabulary was extended with 2,000 additional genes selected for relevance to Alzheimer's disease.

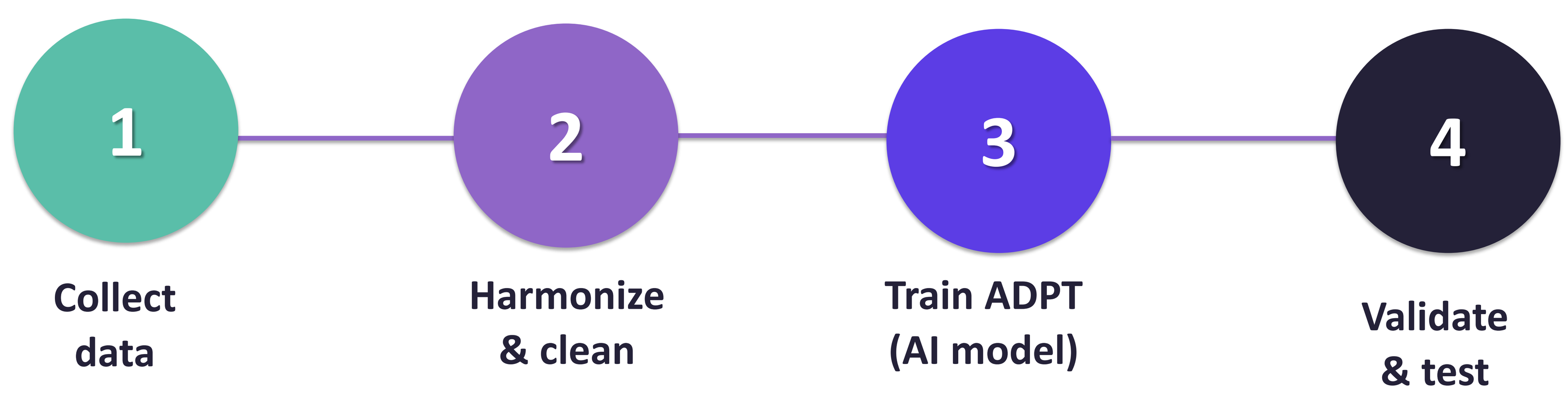


Fig. 1. *ADPT pipeline: harmonize data from 96 studies, pretrain the model, then validate on real Alzheimer's cells.*

03 Dataset at a glance

Summary statistics of the harmonized dataset (Alzheimer's and control donors):

1.7M Cells	**96** Studies	**166** Donors
60k Unique genes	**761k** Braak-staged cells	**4** Data sources

04 Key results

Classification accuracy on two downstream tasks, compared to a non-specialized baseline:

83% AD vs. healthy cell classification accuracy	**83%** Braak stage (severity) classification accuracy

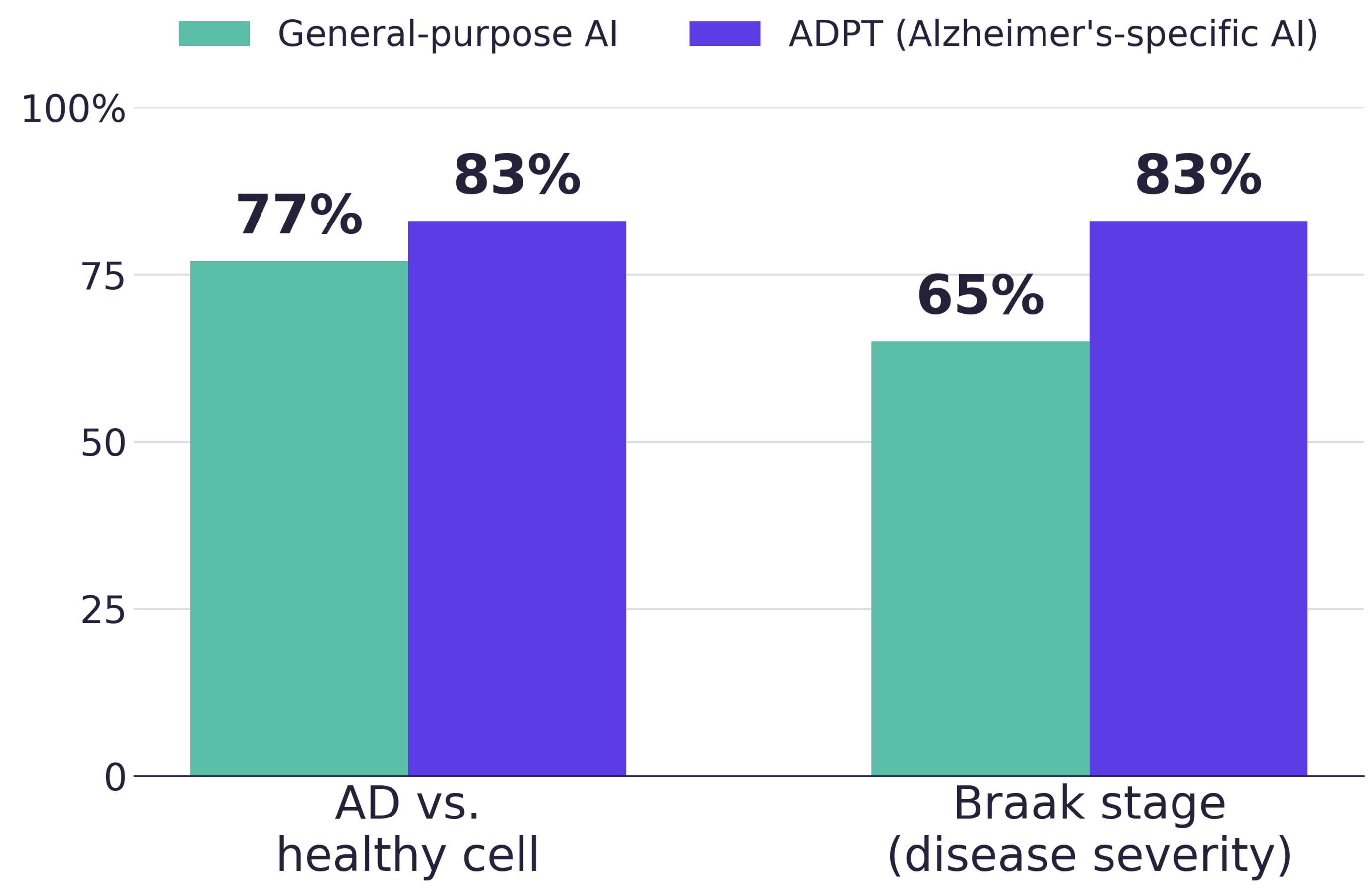


Fig. 2. *Classification accuracy, general-purpose AI vs. ADPT, on both diagnostic tasks.*

The accuracy gain is attributed to AD-specific pretraining rather than dataset size alone.

05 What we discovered

Without being told the answer, ADPT organized genes into clusters on its own, revealing genes linked to Alzheimer's biology. Several clusters were especially active in the entorhinal cortex and hippocampus; the brain regions hit earliest and hardest by the disease.

141 Gene clusters found	**550** AD-linked genes identified

06 Why it matters

- Disease-specific AI training beats general-purpose models for Alzheimer's research.
- ADPT uncovers brain-region-specific gene activity without human supervision.
- ADPT uncovers brain-region-specific gene activity without human supervision.

References & data

[1] Cui et al. "scGPT: a foundation model for single-cell multi-omics." Nat. Methods, 2024.

[2] Dataset: Siozos et al., Zenodo (2025), zenodo.org/records/15526235

SCAN FOR FULL PAPER

Acknowledgement

This work is part of the ENFIELD project. The project is co-funded by the European Union (Grant Agreement No 101120657)

AI NoE ENFIELD

Co-funded by the European Union