

An AI-Driven Clinical Trial & Biomarker Observatory for Alzheimer's Disease Drug Development

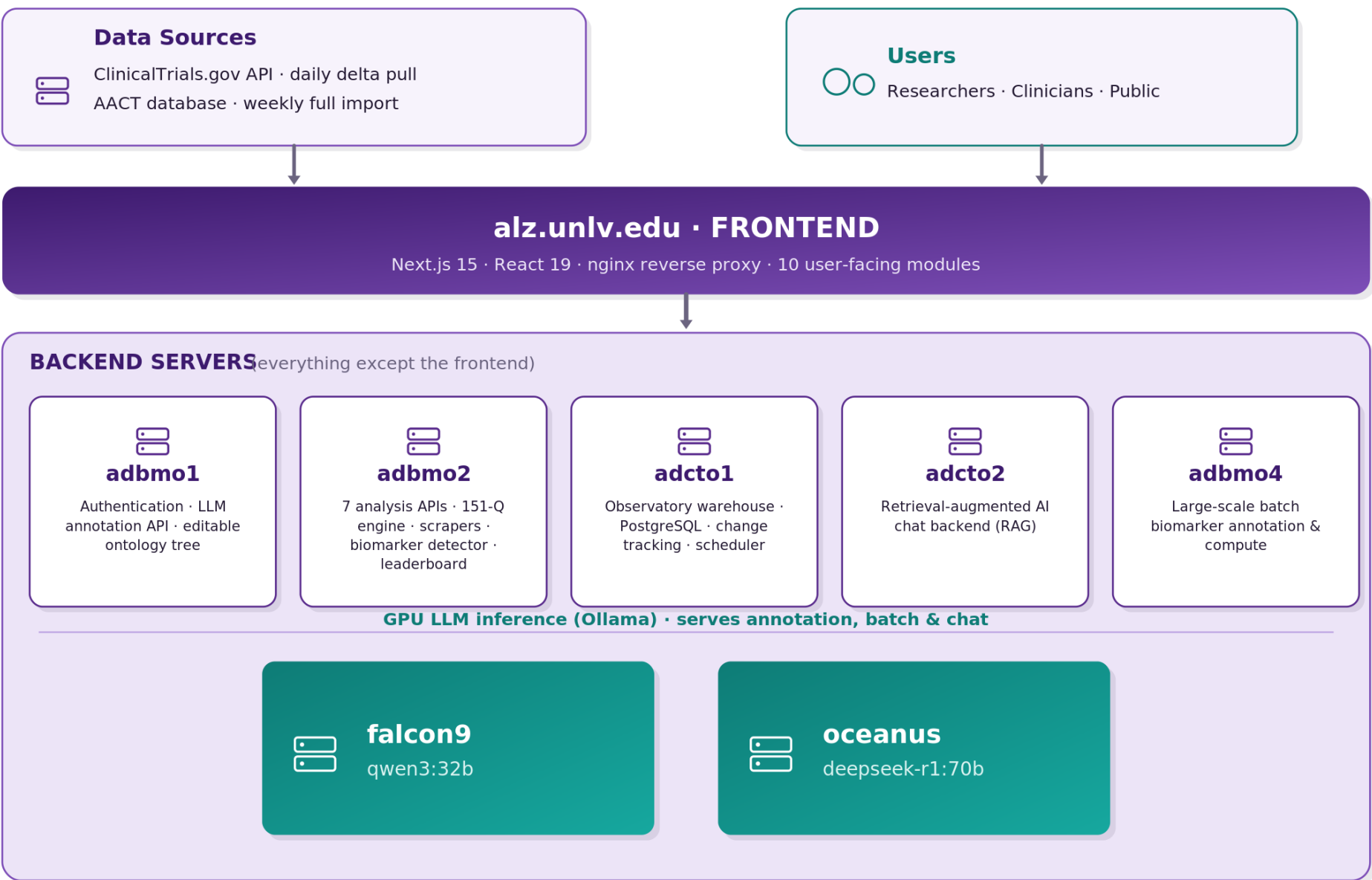
Jorge Fonseca¹; Amanda M. Leisgang Osse²; Andrew Hooyman³; Gaurab Subedi¹; Xiaowei Zhuang⁴; Yadi Zhou⁵; Jefferson W. Kinney²; Feixiong Cheng⁵; Jeffrey L. Cummings²

¹Dept. of Computer Science, Howard R. Hughes College of Engineering, UNLV ²Chambers-Grundy Center for Transformative Neuroscience, Dept. of Brain Health, Kirk Kerkorian School of Medicine, UNLV
³Physical Therapy, Crean College of Health and Behavioral Sciences, Chapman University ⁴Lou Ruvo Center for Brain Health, Cleveland Clinic ⁵Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic

Background: Tracking Alzheimer's disease (AD) clinical trials is important for understanding drug development. We assess this through extraction and interrogation of ClinicalTrials.gov and AACT database data. Manual approaches limit scalability and delay annual reporting. We present a multi-server AI framework automating clinical trial and biomarker tracking, advancing our abilities to evaluate progress in AD drug development.

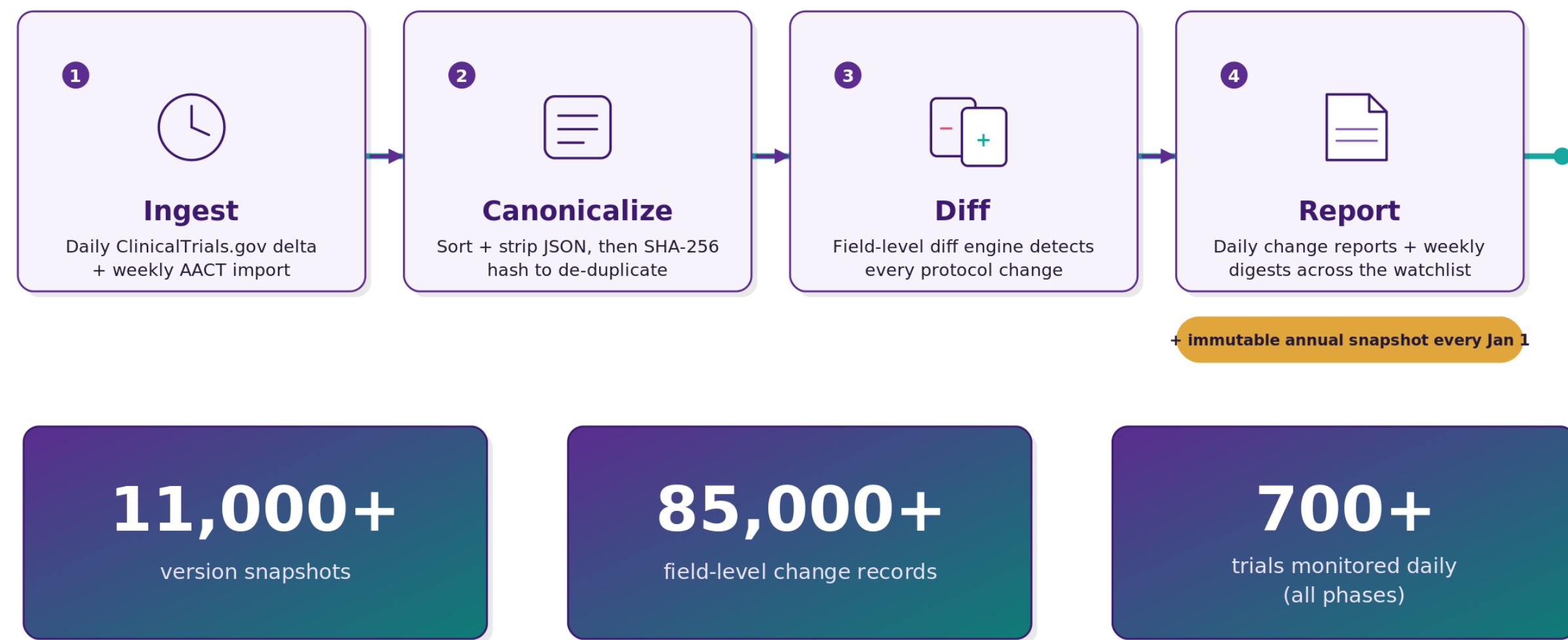
A Distributed, Ten-Server AI Platform

A public web frontend backed by specialized data, AI, and GPU servers, governed by role-based access control



Always-On Trial Surveillance

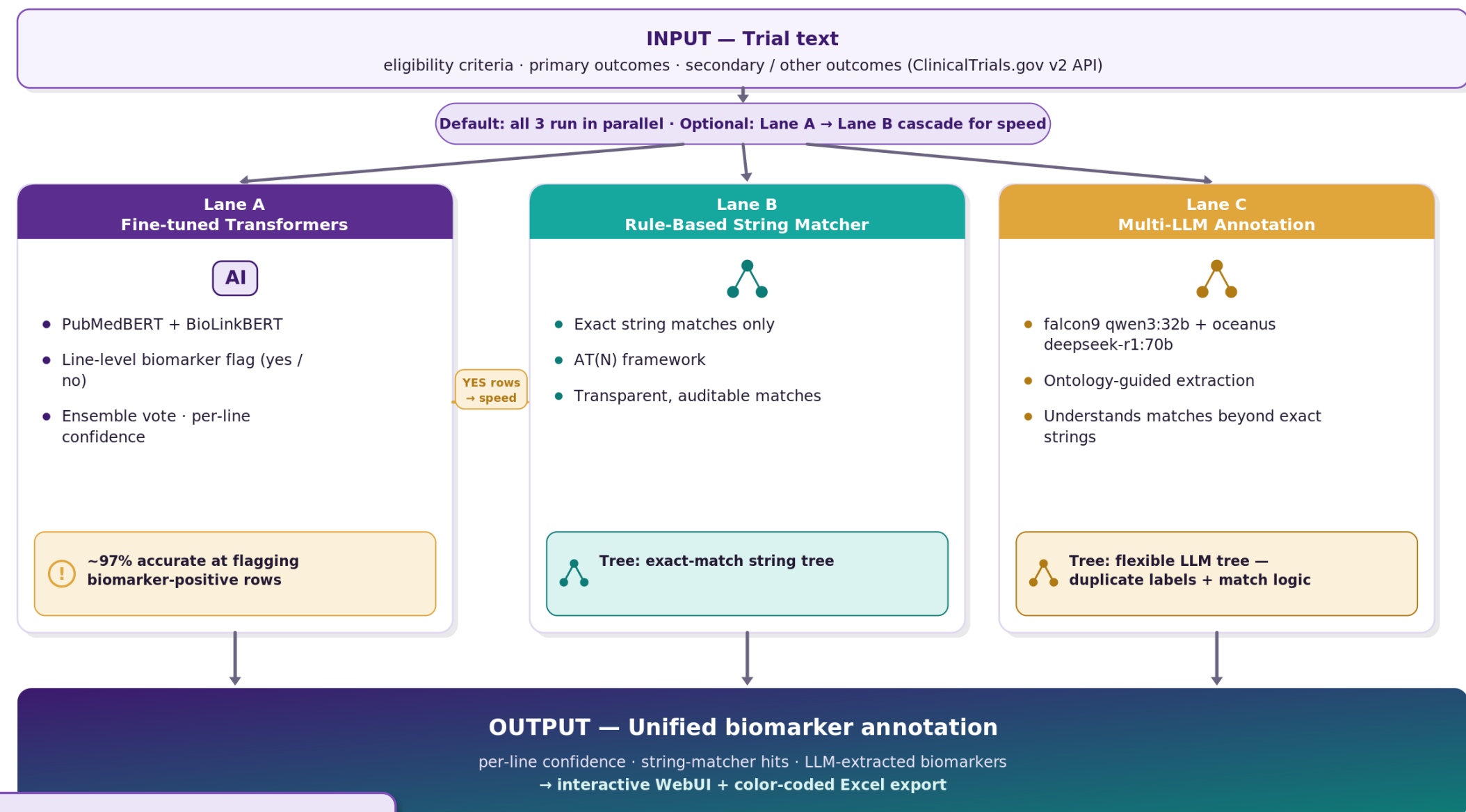
Automated ingestion, versioning, and change tracking — no manual curation



Methods: The platform comprises ten coordinated servers with role-based access control. A Prisma and PostgreSQL-backed observatory ingests trial data through daily ClinicalTrials.gov API updates and weekly AACT imports, automatically classifying new AD trials and generating daily reports of protocol modifications across the active watchlist. Biomarker detection employs a three-tier approach: (1) fine-tuned PubMedBERT and BioLinkBERT classifiers line-level biomarker detection, (2) rule-based string matchers, and (3) concurrent multi-LLM-based annotations guided by researcher-editable ontology trees. The system provides real-time longitudinal analytics, including therapeutic target distributions and category trends.

Three Methods, One Biomarker Annotation

Each method detects biomarker mentions in trial text and produces biomarker annotations. By default all three run in parallel; an optional cascade chains Lane A → Lane B for speed.



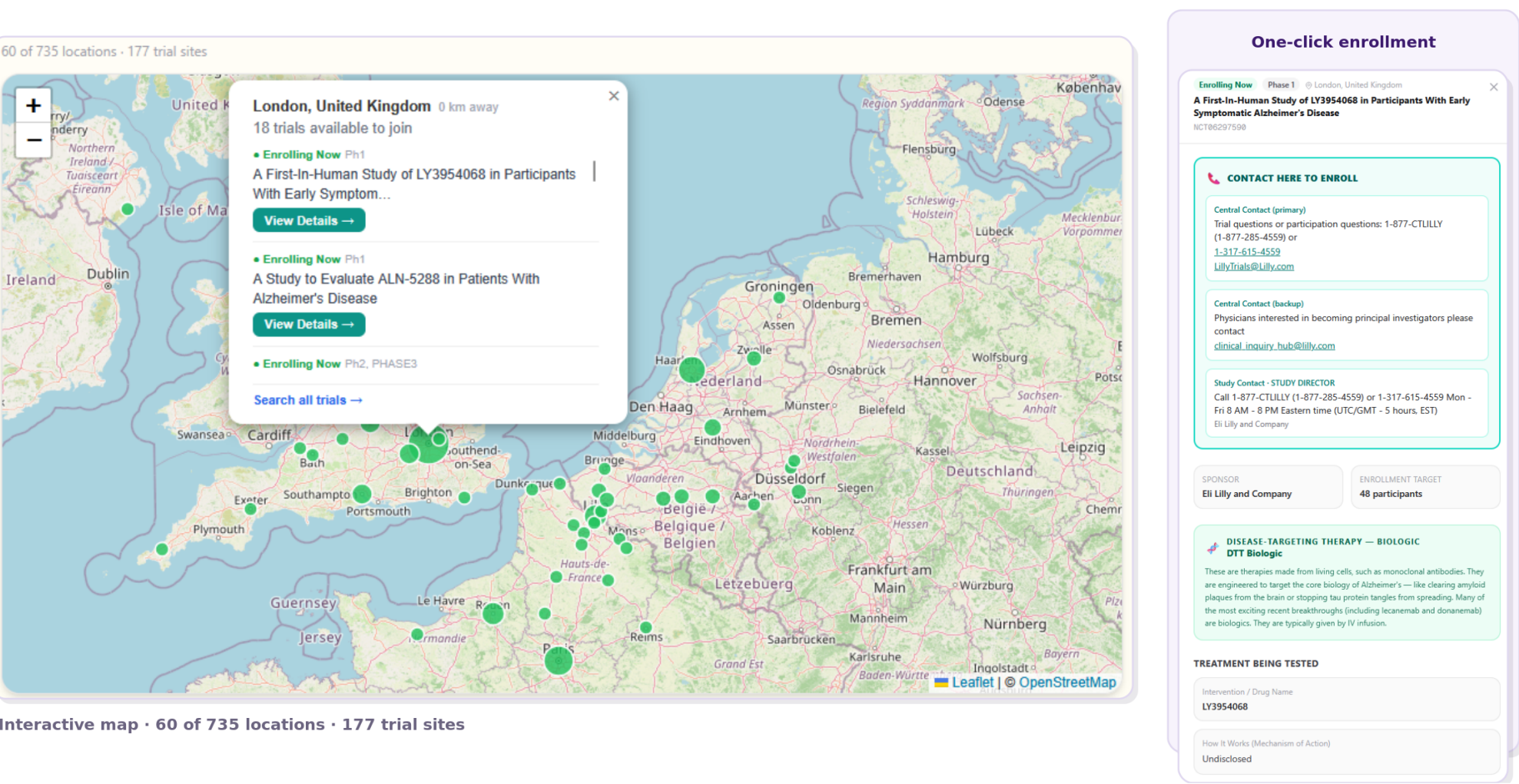
One Platform, Ten Modules

Serving researchers, clinicians, and the public — extensible to other neurodegenerative diseases



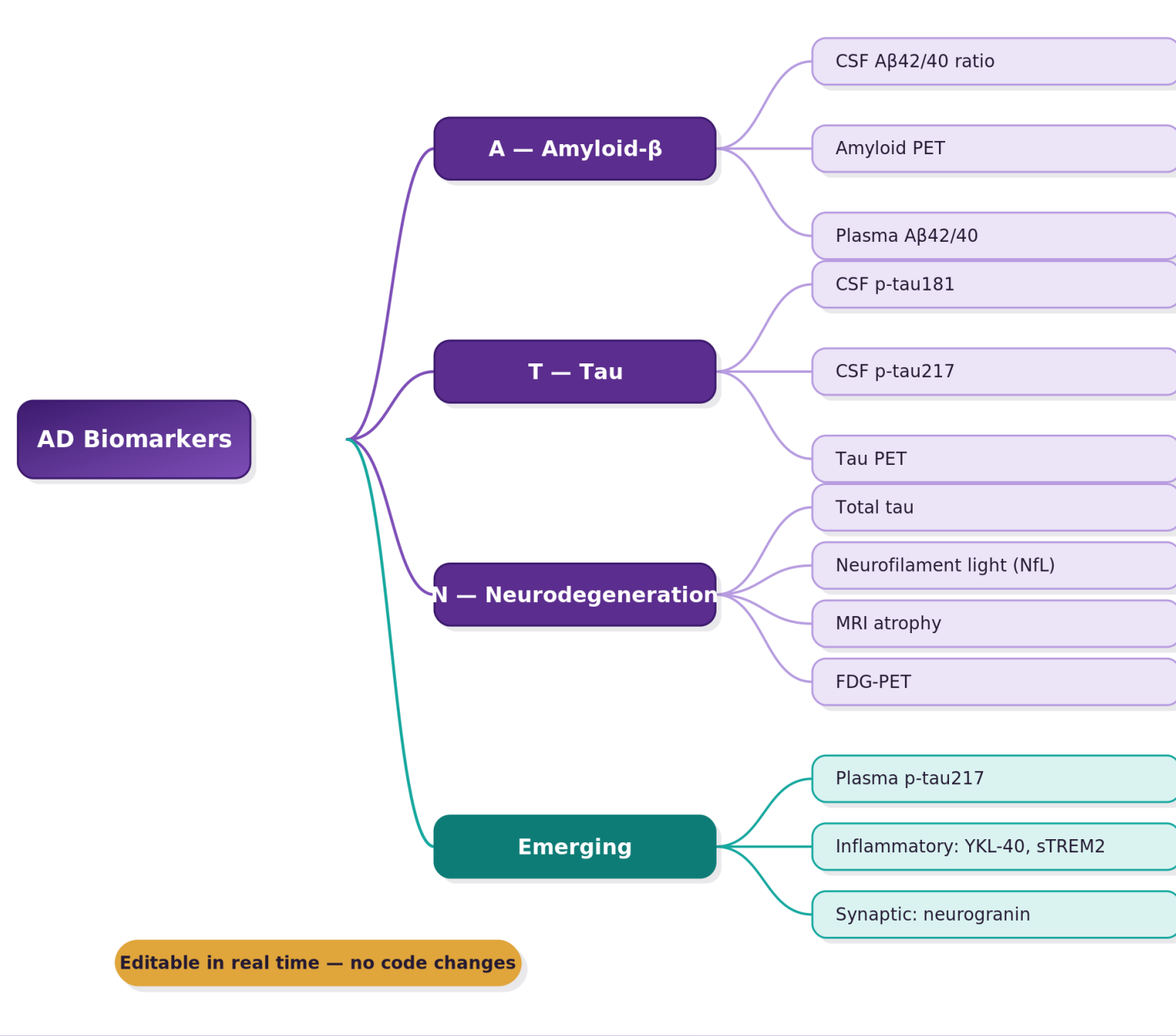
Find a Trial Near You

Patient-facing trial locator with an interactive map and one-click enrollment contacts



Researcher-Editable Biomarker Ontology

AT(N) knowledge tree that guides the string matcher and LLM annotation — curated via Tree Editors



From Free Text to Confidence-Scored Annotations

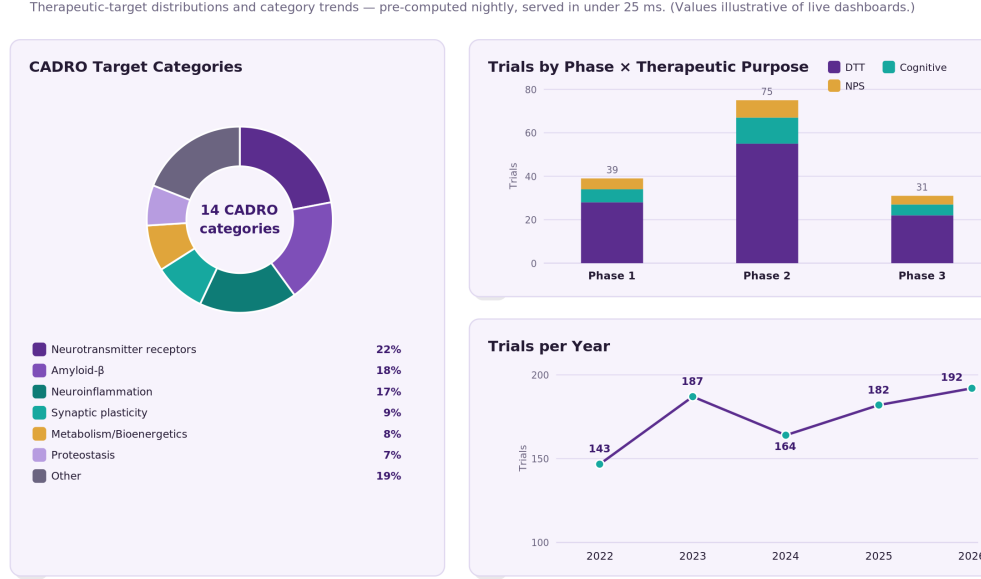
Each trial-text line scored by every method — delivered as color-coded Excel + interactive WebUI

Trial text line	PubMedBERT	BioLinkBERT	Ensemble	String match	LLM biomarker
Plasma p-tau217 ≥ threshold for inclusion	0.98	0.96	✓	p-tau217	p-tau217 (Tau)
Amyloid PET positivity required	0.95	0.93	✓	amyloid PET	Amyloid PET (Aβ)
Age 55-85 years at screening	0.03	0.05	✗	—	—
CSF Aβ42/40 ratio measured at baseline	0.97	0.94	✓	Aβ42/40	Aβ42/40 (Aβ)
MMSE score 20-28 (cognitive measure)	0.12	0.08	✗	—	—

Legend: ■ = biomarker, ■ = not a biomarker

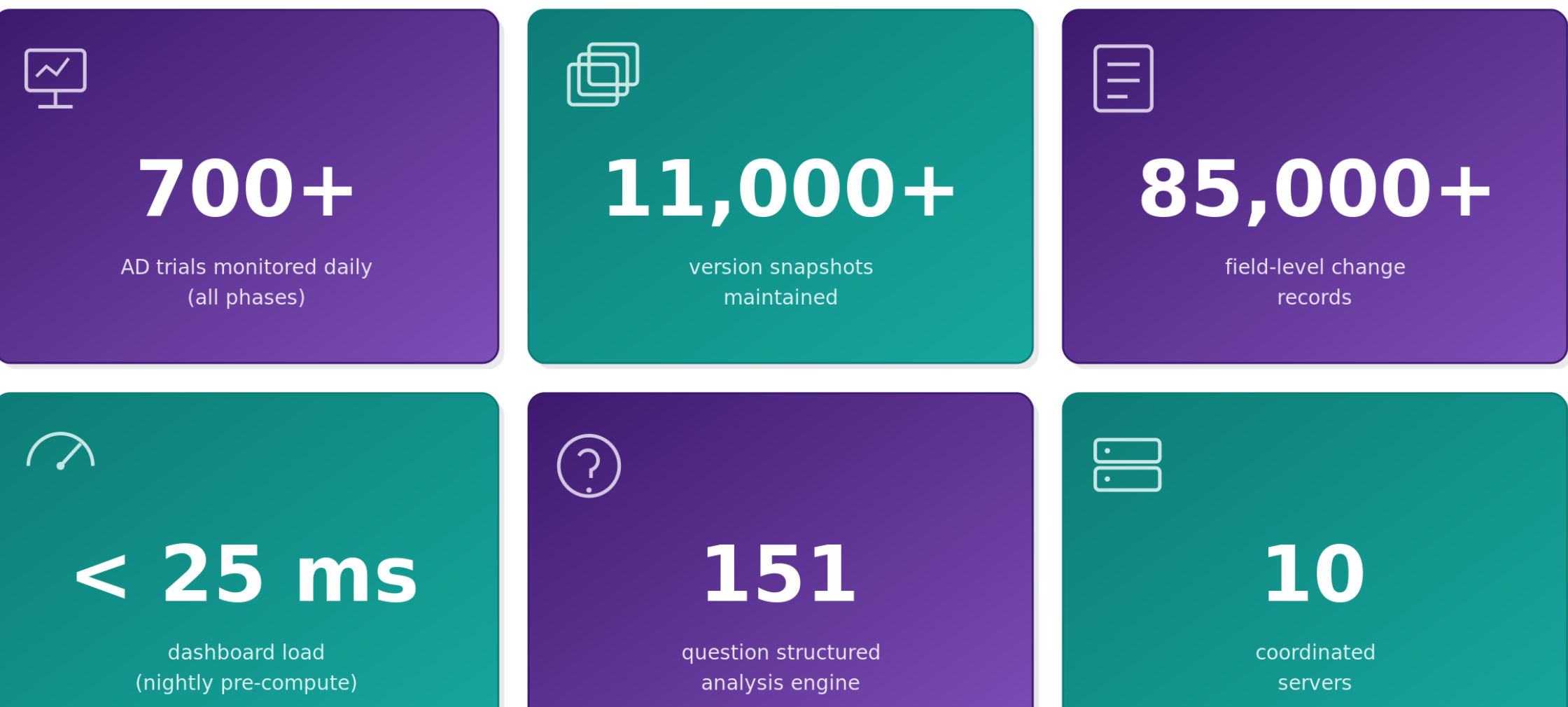
Confidence scores from PubMedBERT / BioLinkBERT classifiers - Ensemble decision (✓ / ✗) - ontology string match - LLM biomarker normalization with AT(N) class.

Live Longitudinal Pipeline Analytics



Built for Scale

A production system replacing months of manual curation with a continuously updated, research-ready resource



Results: The website monitors daily 700+ AD trials across all phases, maintaining 11,000+ version snapshots and 85,000+ field-level change records. Pre-computed nightly analytics serve all dashboards in under 25 milliseconds. The batch annotation pipeline processes trial outcomes and eligibility criteria at scale and generates an interactive webUI and color-coded Excel exports with per-line model confidence scores, string matcher results, and LLM-extracted biomarker annotations. Researchers can access the latest information through a 151-question structured trial analysis engine and AI chatbot. Additionally, modules provide the research team with accurate, real-time, annotated data for generation of clinical trial and biomarker publications, including an eligibility-criteria batch scraper with biomarker annotations, a news aggregation system capturing information on new biomarkers and trials, and an alert system for monitoring ongoing trials. The platform includes educational games promoting public engagement and a patient-facing clinical trial locator with interactive geographic mapping and trial contact information.

Conclusion: We demonstrate that distributed AI systems can automate AD drug development tracking at scale. Its modular architecture supports iterative model improvement, providing a framework for AI-assisted drug development decision support that reduces the time from discovery to actionable outcomes. Furthermore, our platform is built on an architecture that is readily extensible to other neurodegenerative diseases.



Alzheimer's
Drug Discovery
Foundation

CONTACT



Jorge Fonseca
Dept. of Computer Science, UNLV
Jorge.FonsecaCacho@unlv.edu
alz.unlv.edu

Scan to explore the live platform

Funded by the Alzheimer's Drug Discovery Foundation (ADDF). PI: J. L. Cummings.

Refs: Cummings et al., AD pipeline 2021-2026; Leisgang Osse et al., Biomarkers in AD trials 2025.

