

# Anomaly and Bias Detection Interface

## Pitch Presentation

---

July 26th 2025



AGENDA	Meet the Team	1
	Our Vision	2
	Problem	3
	Impact Hypothesis	4
	Our Solution ( <b>ABD.i</b> )	5
	User Workflow	6
	Addressable Market	7
	Competitive Landscape	8
	Objectives	9
	Investing in ABD.i	10

## Team



# Natasha Kodgi

Co-Founder

Biomedical Engineer with over 3.5 years of specialized industry expertise in immunotherapy, bioinformatics, and clinical trial design. Experience spans CAR-T cell therapy trials, viral vector and vaccine manufacturing at Johnson & Johnson, and data-driven process optimization across biopharma pipelines. Brings a cross-disciplinary skill set at the intersection of therapeutic development and computational modeling, with a commitment to advancing equity in clinical research.



# Kamala Natarajan

Co-Founder

Bioinformatics professional with a strong foundation in biological sciences, computational biology, and product management. She holds a Master of Science in Bioinformatics and Bachelor of Science in Biology from Georgia Tech and has conducted translational research at the intersection of AI and healthcare at institutions like Emory and Duke. Kamala brings academic research depth, having applied machine learning to study biomarkers in Acute Respiratory Distress Syndrome. With experience leading initiatives in cancer bias detection and a commitment to equitable, ethical data use in healthcare, she combines expertise in biomedical science, data analytics, and computational modeling to build impactful, data-driven solutions.



# Nabin Kim

Co-Founder

Computer Scientist with applied expertise in machine learning, biomedical informatics, and AI-powered healthcare tools. Experience spans EEG-based concussion detection, bias analysis in AI-generated medical data, and real-time triage systems integrating EHR and wearable sensors. Has led development across full-stack pipelines, from data modeling and ERP feature extraction to LLM-based decision support. Brings a cross-disciplinary background in computing, neuroscience, and public health, with a focus on equitable and explainable AI solutions in clinical contexts.

## The vision

We aim to become the go-to platform for designing inclusive, bias-aware clinical trials, ensuring that life-saving therapies are developed for everyone, not just a subset of the population. By automating both data preprocessing and demographic bias detection, **we're revolutionizing how trials are designed, analyzed, and trusted.**

---

*"At ABD.i, we're not just building a tool, we're building a standard. Our long-term vision is to integrate equity checkpoints into every stage of clinical research, globally. We want our platform to be the default infrastructure for fair trial design, scalable, intelligent, and indispensable."*

— Kamala Natarajan, Natasha Kodgi, Nabin Kim (Co-Founders)



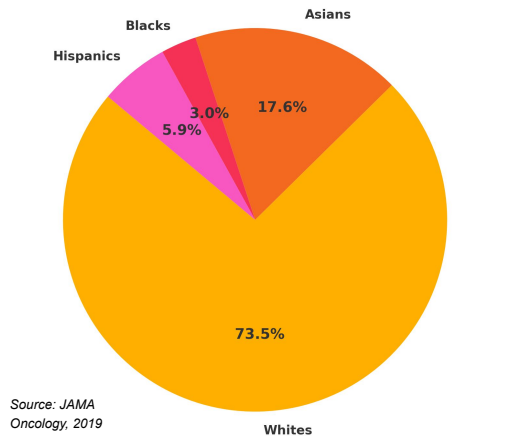
# Problem

## The Critical Problem

Cancer research generates massive datasets, but **hidden biases and data anomalies** are undermining treatment efficacy and perpetuating health disparities:

### Problem 1: Participant Representation

Data Bias Crisis: Cancer Trial Participant Representation



### Problem 2: Data Prep is Time Consuming

**“Data preparation is often 50–70% of project time, sometimes even up to 90%.”**

Source: Pérez et al., J Med Syst, 2015



## Problem

“Although Black women are disproportionately affected by triple-negative breast cancer, they **made up only 5.4 % of participants in recent TNBC clinical trials**—less than half of their representation in the U.S. population (~14 %)”

— *Pharmacy Times Review of FDA-backed TNBC trials*

“Aggregated data hides critical group disparities, obscuring the fact that some Asian American women have among the **lowest breast cancer screening rates** of any racial or ethnic group in the U.S.”

— *Sohn et al., 2021*

- Representation gaps undermine both ethics and efficacy.
- Generalizability compromised
- High risk for aggressive cancers like TNBC





“Of all the forms of inequality, injustice in health is the most shocking and the most inhuman.”

— *Dr. Martin Luther King Jr.*



If clinical trial datasets are automatically screened for data anomalies and demographic bias, then researchers will be better equipped to **identify and mitigate inequities** in trial design and analysis, leading to more **inclusive, accurate, and ethically sound outcomes** that improve health equity across diverse patient populations.





## Our Strength



### Who We Are

Led by **Women in STEM**, we're advancing **representation in both research and leadership**.



### Lived Experience

We've faced **discrimination firsthand**, and built ABD.i to change that reality for future with our representation.



### Scalability

Our solution scales across **geographies, disease areas, and population** settings.



### Commitment

We're committed to **equity-first research design**, starting with **underserved populations and women's health**.

# **ABD.i = Anomaly and Bias Detection Interface**

**Proof of Concept: Python-powered Streamlit Application**

**Two core modules:**

- 1. Anomaly-Aware Preprocessing**
- 2. Bias Detection + Subgroup Visualization**



# ABD.i = Anomaly and Bias Detection Interface

✓ Upload CSVs for auto-analysis

✓ Detect and correct:

- Duplicates (case\_id, project\_id)
- Negative/invalid values (age < 0)
- Outliers (e.g., age > 120)
- Missing values

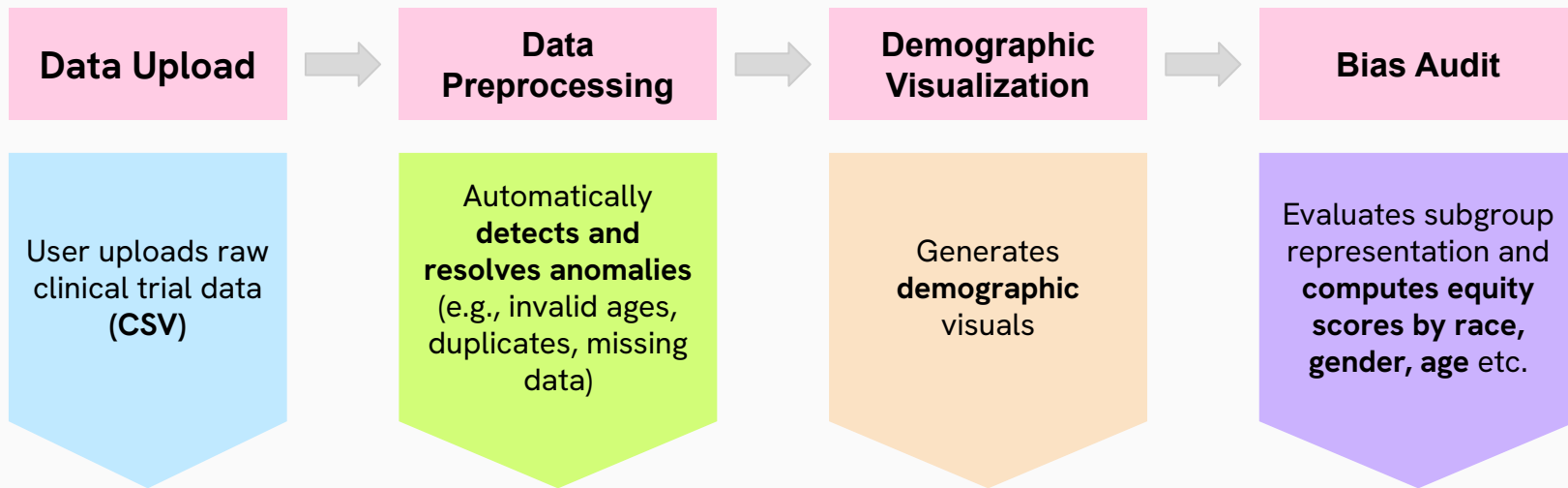
✓ Identify bias across race, gender, age

✓ Demographic bar graph visuals

✓ Distribution Distance Metrics

✓ Fairness-Audit

## ABD.i Workflow: From Upload to Equity Insight



# ABD.i Application Demo



## Bias Audit Dashboard

 Upload CSV



Drag and drop file here

Limit 200MB per file • CSV

Browse files



Please upload a dataset to begin.



**ABD.i**

# User Workflow



## Bias Audit Dashboard

Deploy ⋮



Upload CSV



Drag and drop file here

Limit 200MB per file • CSV

Browse files



TCGA-BRCA - clinical\_dataset\_breast\_cancer.csv 5.8MB



### Preprocessing Legend

Show explanation for each preprocessing option ⌵



### Preprocessing Recommendation Summary

	Data Type (e.g., Diagnoses, Tests)	Total Columns	Columns Suggested for Removal (🔥 Unused or Empty)	Average Missing Data (%)	Columns Suggested for Grouping (Group by Category)	Columns Suggested for Numeric Adjustment (Adjust Numbers)
0	project	1	0	0.0%	1	0
1	cases	9	1	15.7%	5	1
2	demographic	21	10	58.2%	4	3
3	diagnoses	132	101	82.7%	16	2
4	treatments	46	24	81.1%	10	5

✅ File successfully loaded!

☐ 🗒 Show detailed preprocessing logs



### Preprocessing Options

✅ 🔧 Apply Scaling to numeric columns (MinMaxScaler) ⓘ

✅ 🔧 Encode categorical columns ⓘ

✅ 🌱 Handle missing values automatically ⓘ

🔧 Apply Recommended Preprocessing

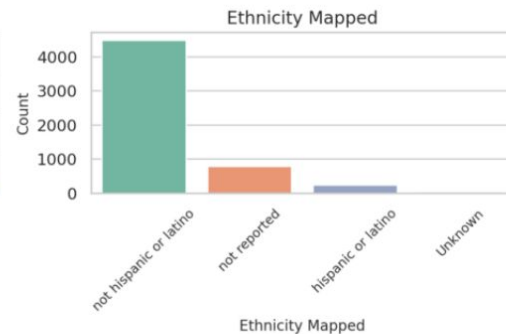
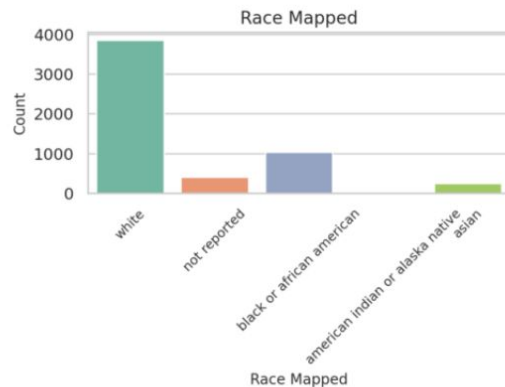
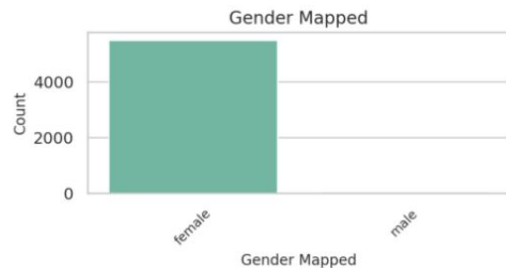


## User Workflow

[Download Processed Data](#)

### Data Preprocessing and Visualization

#### Demographic Distributions



☒ Show visualization



## User Workflow

### Fairness Audit Results

★ Specify benchmark distribution (optional, JSON format: {'GroupA': 0.5, 'GroupB': 0.5})

Enter benchmark distribution as JSON

```
{  
  "6": 0.385,  
  "7": 0.226,  
  "8": 0.143
```

Total Groups

14

Fair Groups

0

↑ 0.0%

Unfair Groups

14

↓ -100.0%

### Distribution Distance Metrics

KL Divergence

0.8773

🚨 Severe deviation

Wasserstein Distance

0.0771

● Mild deviation

Total Variation

0.4993

🚨 Severe deviation





## Addressable Market

### A \$100B+ Opportunity in Clinical Research Tech

- **Global CRO Market Value:**

- **\$65.06B** in 2024 → **\$126.17B** by 2034  
– *Source: Precedence Research, 2024*
- Alternative estimate: **\$55.95B** in 2024 → **\$104.6B** by 2033  
– *Source: BioSpace, 2024*

- **Why It Matters:**

- CROs conduct the majority of clinical trials globally
- Outsourced trials = growing demand for tech-first, automation-focused tools
- **Even 1% of this market = \$1B+ revenue potential**



## Addressable Market

### Pharma R&D Spending Is Exploding

- **\$238B** spent on pharma R&D in 2022  
– *Source: Statista & PR Newswire, 2023*
- **\$83B** in 2019 — nearly 10x 1980s spend (inflation-adjusted)  
– *Source: Congressional Budget Office, 2021*

### Why It Matters:

- Majority of R&D budget flows into clinical trials.
- CROs are the execution arm of this spend
- As complexity rises, the need for smarter, automated tools grows.



# Addressable Market

## Our Tool Targets the CRO and Pharma R&D Sector

- Every trial outsourced to a CRO faces **data harmonization and demographic equity challenges** - we address both.
- As **outsourcing increases** and trial protocols become more complex, **AI tools** will dominate.



# Competitive Landscape

## Acclinate

"Connecting Communities to Care"

## Baseline Trials AI

"Structuring and formatting of clinical trial data"

## Trial Library

"Diverse trial recruitment"

## Tempus Labs

"Predictive healthcare AI"

AI-Powered Platform

✓

✓

✓

✓

Clinical Trials Focus

X

✓

✓

X

Data Anomaly Detection  
and Preprocessing

X

✓

X

✓

Demographic Bias Detection  
& Monitoring

X

X

X

X

Demographic Diversity  
Focus

✓

X

✓

X

## OBJECTIVES

### Marketing

Be the voice of  
inclusive trial design.



Webinars



Whitepapers



Conference  
Talks

Visibility → Credibility → Adoption

### Communication

Speak equity  
Speak clearly

“

A game-changer  
for inclusive  
clinical trials.

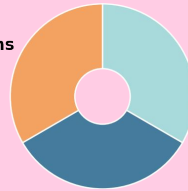
”

“

This tool helped  
me identify and  
address biases  
in our study.

”

Clinicians



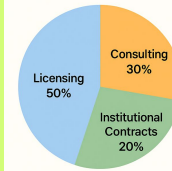
Researchers

Biotech

### Business

Prove, grow, sustain.

Revenue Streams



Pilot Partnerships



## OBJECTIVES

### Objective 1

July-September, 2025

#### Validate Core Functionality

- Finalize MVP with bias/anomaly detection modules.
- Conduct internal testing and demo sessions with oncology researchers.
- Secure interest from MCG, Emory, or similar for pilot discussions.
- Begin Institutional Review Board (IRB) review to ensure ethical compliance.

### Objective 2

October-December, 2025

#### Launch First Pilot & Build Momentum

- Launch small-scale pilot with one institutional partner.
- Gather data on accuracy, usability, and time savings.
- Develop a case study of our progress.
- Apply to at least one accelerator or digital health grant.

### Objective 3

January-March, 2026

#### Expand Based on Pilot Learnings

- Add new features (e.g., subgroup analysis, dashboards, audit trails).
- Reach out to 2-3 new cancer centers or research partners.
- Publish pilot outcomes or present at oncology/health informatics events.
- Enhance backend to support multi-site data and protocols.

### Objective 4

April-June, 2026

#### Scale & Commercialize

- Initiate strategic talks with, CROs, or academic centers.
- Identify 1-2 pharma-sponsored trials to test scalability.
- Begin regulatory prep (HIPAA, FDA, etc.).
- Explore commercialization paths: SaaS, licensing, or research consulting.



APP:

# Why You Should Invest in ABD.i



## Proven Team Advantage

Our team of **engineers and data scientists** spans trials, tech, and health systems, no outsourcing, no cookie-cutter templates.



## Massive Market Access

Target buyers **biotechs, CROs, hospital systems, and public health agencies** represent a multi-billion-dollar global market



## Game-Changing Efficiency

Our **plug-and-play interface** shrinks **6+ months** of manual data work into just weeks. Saving clients time and money.



## Validated and Live Prototype

**Working prototype** has already been tested on real trial datasets, surfacing bias and demographic insights

# Thank you!

---

ABD.i

@abdi-biotech



**venmo**

If you have any inquiries, feel free to contact us at our email  
**abdibiotech@gmail.com**