

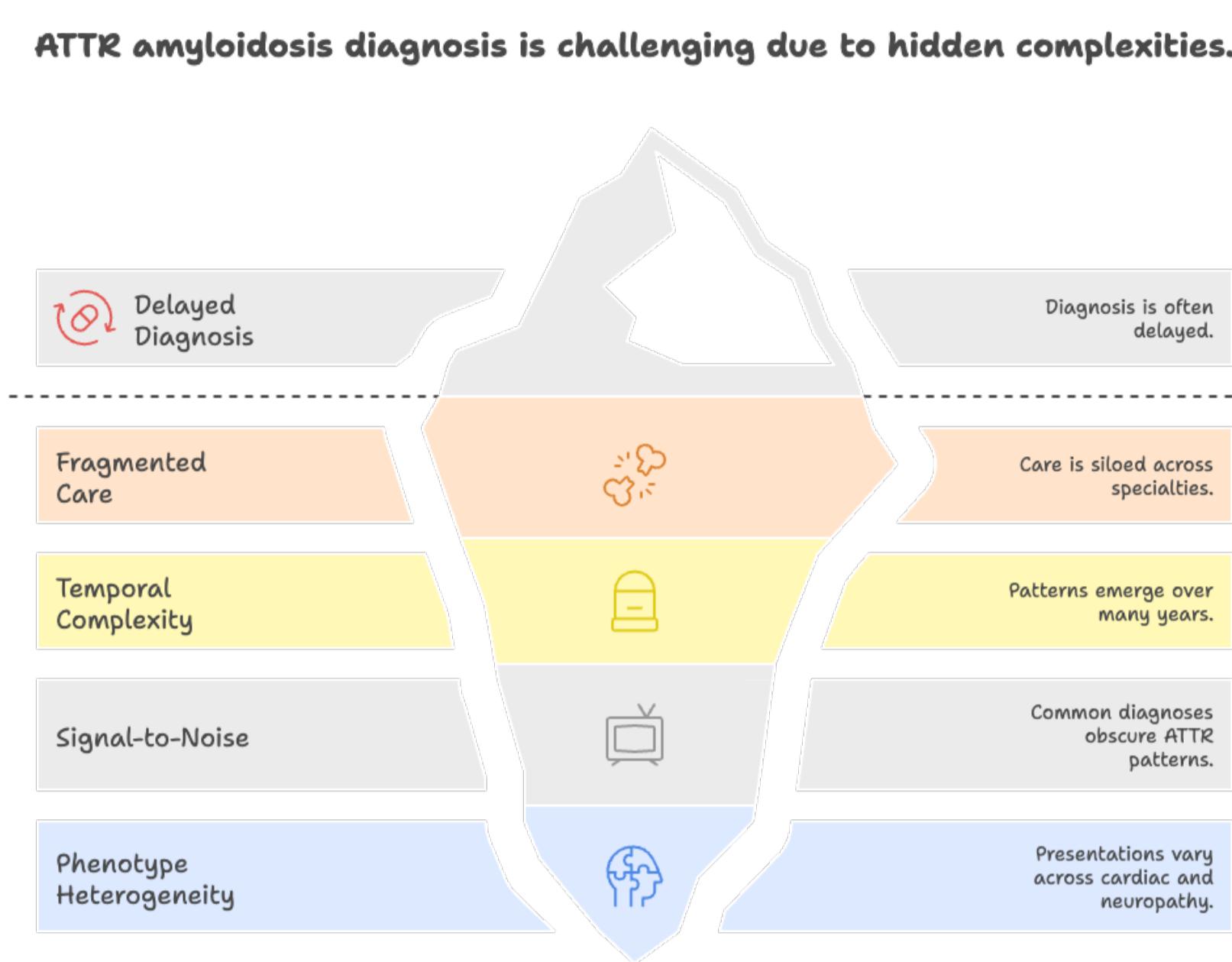
Leveraging Large Language Models and EMR Data to Identify Undiagnosed Rare Diseases: A Hybrid AI Approach

RWD117

Sandy D. Balkin, Ph.D.

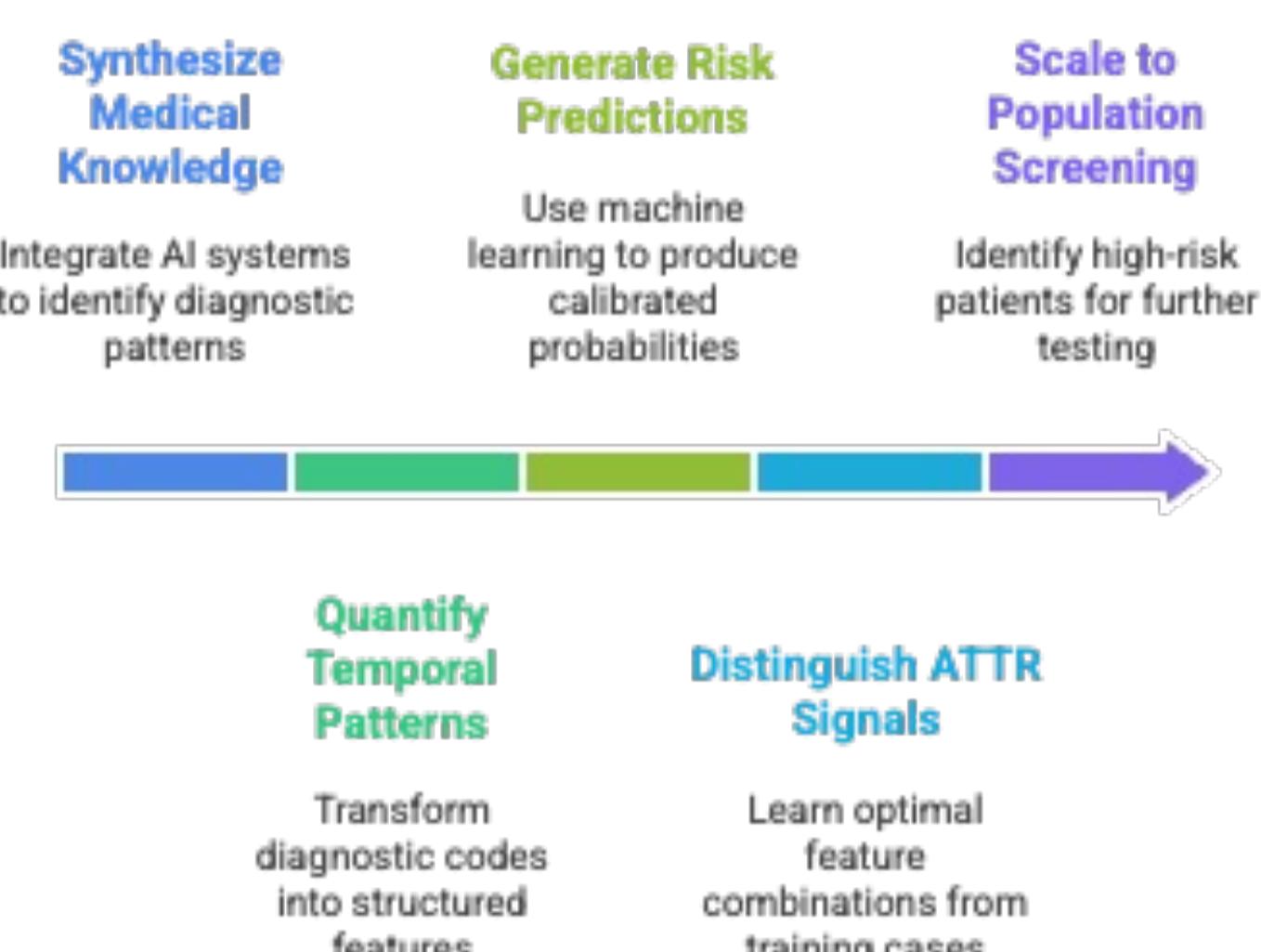
SVP Strategy & Analytics Royalty Pharma, New York, NY USA

INTRODUCTION



OBJECTIVE

EMR Screening Algorithm Development

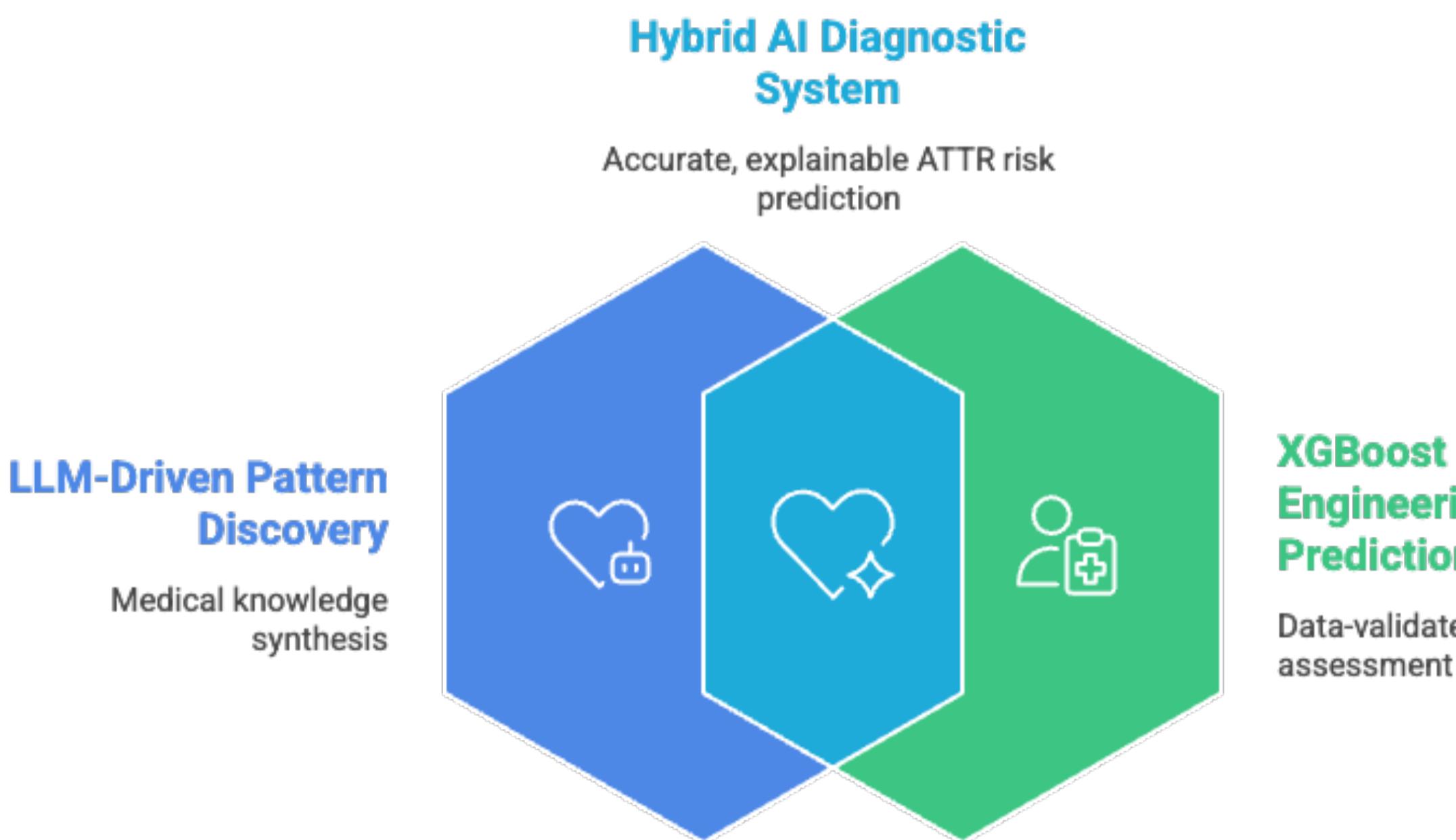


- Synthesize Medical Knowledge**
 - Integrate AI systems to identify diagnostic patterns
- Generate Risk Predictions**
 - Use machine learning to produce calibrated probabilities
- Scale to Population Screening**
 - Identify high-risk patients for further testing

- Quantify Temporal Patterns**
 - Transform diagnostic codes into structured features
- Distinguish ATTR Signals**
 - Learn optimal feature combinations from training cases

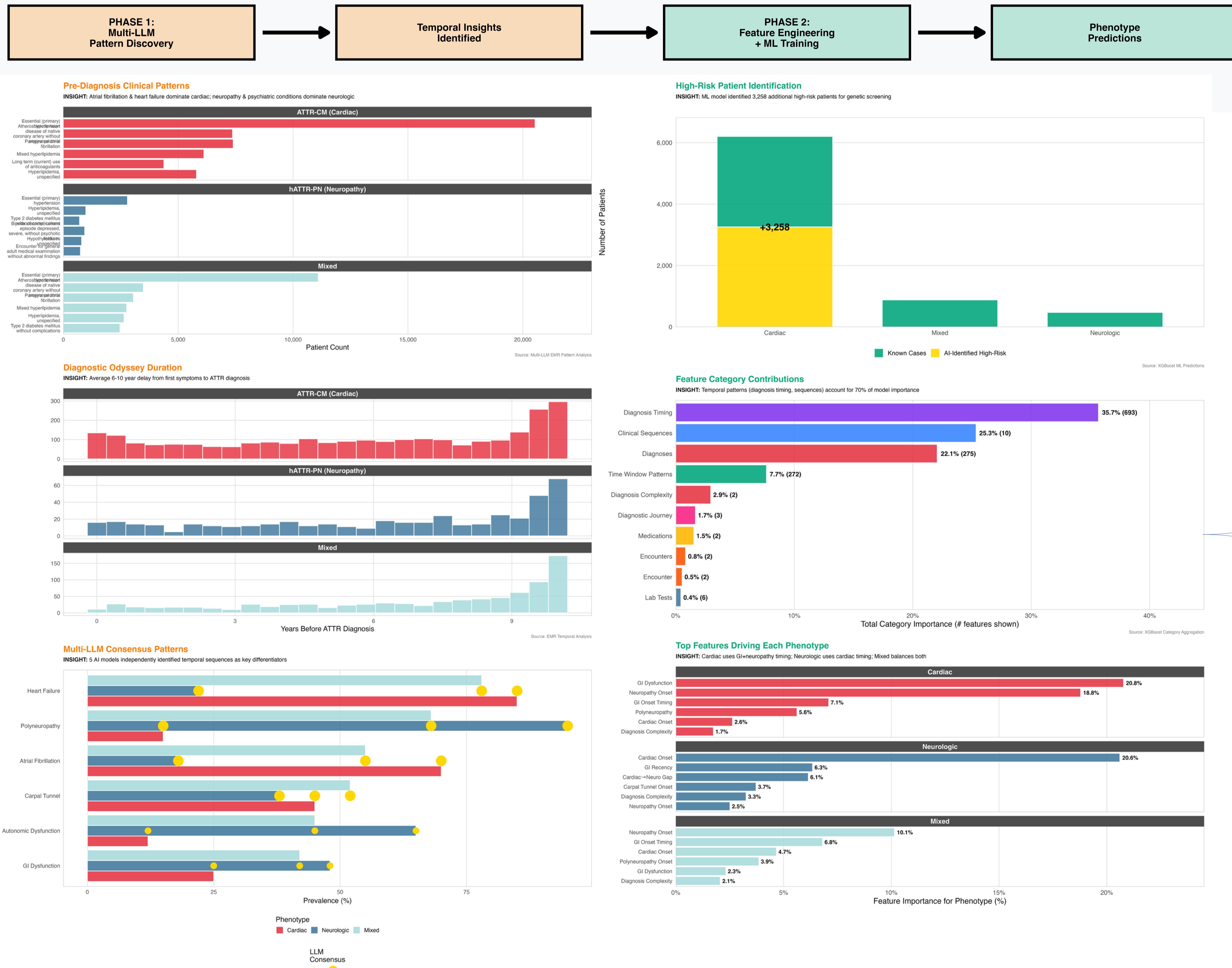
METHOD

Synergy of LLMs and Machine Learning in ATTR Diagnosis



RESULTS

HYBRID AI METHODOLOGY: Phase 1 Pattern Discovery → Phase 2 Machine Learning



CONCLUSIONS

- Novel Hybrid AI Architecture:** First multi-LLM pattern discovery pipeline informing ML feature engineering for ATTR phenotyping, achieving 86.73% accuracy across 7,562 patients with 1,288 comprehensive features
- Temporal Features Dominate:** 979 temporal features account for 70% of model importance, proving that symptom timing and clinical sequences are more predictive than symptom presence alone
- Phenotype-Specific Feature Signatures:** Cardiac predictions rely on neurologic timing, neurologic predictions rely on cardiac timing - revealing how the model differentiates phenotypes through temporal patterns
- Actionable Clinical Impact:** Identified 3,258 high-risk undiagnosed patients (43% increase over known cases), enabling proactive screening and potentially reducing the 6-10 year diagnostic odyssey

DATA DESCRIPTION

HealthVerity Source 42:

- Company overview** - NextGen Healthcare, Inc. is an American software and services company headquartered in Atlanta, Georgia that develops and sells electronic health record (EHR) software and practice management systems to the healthcare industry
- Provider reach** - More than 124,000 providers across the United States use solutions provided by NextGen Healthcare, including large physician groups
- Practice settings** - NextGen EHR is designed specifically for ambulatory practices, serving small practices (1-10 physicians) and larger organizations (10+ physicians)
- Specialty coverage** - NextGen EMR software accommodates a wide range of specialties, including but not limited to Cardiology, Mental and Behavioural Health, Neurology, Gastroenterology, Orthopedics and Urology
- Scale** - contains clinical records for nearly 60 million unique patients. With 500,000-750,000 patients added on a monthly basis, contains both ambulatory and specialty practices medical record volume.