InnovateGPT: LLMs to Measure Innovation
An Application to the Life Sciences Industry

Company Advisor: Vincenzo Palermo   Faculty Advisor: Jose J Pacheco

Amjed Belkhiri   Reid Taylor

Problem Statement and Objective

Our aim is to develop a comprehensive framework using Large Language Models (LLMs) to quantify innovation through patents and help Accenture clients define effective innovation and research investment strategies and understand their competitive landscape.

Innovation is a vital determinant of productivity, competitive advantage and financial success. However, companies and Accenture clients face challenges in measuring and harnessing innovation. So, **what is innovation?**

Patents are the best-known proxy for innovation. However, they are lengthy, technical and difficult to understand. Moreover, there is no universal quantitative link between innovation and patents. So, **how do we measure Innovation?**

Innovation not only influences financial outcomes but also boosts productivity and competitive advantage. But **how do we value Innovation's impact on corporate success?**

Data and Scope

Patents Data
Claims, Status, Owner,...

Corporate Financial Data
Sales, R&D Spends,...

Clinical Trials Data
Drug, Company, Status,...

Scope: Life and Science | 2015-2022 | Worldwide

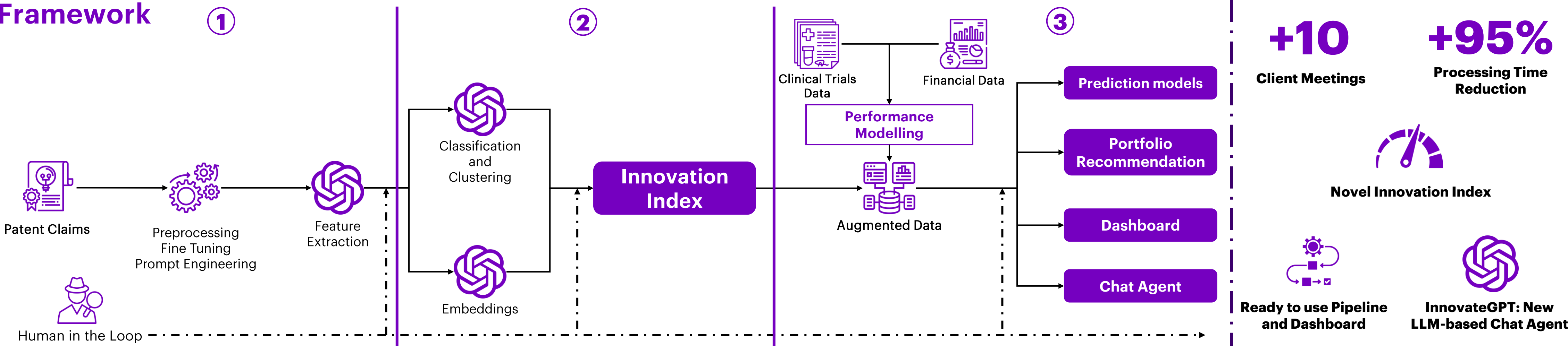**45** Companies   **200K** Patents   **4M** Patent Claims

Methodology

Framework

+10 Client Meetings   **+95%** Processing Time Reduction

Novel Innovation Index

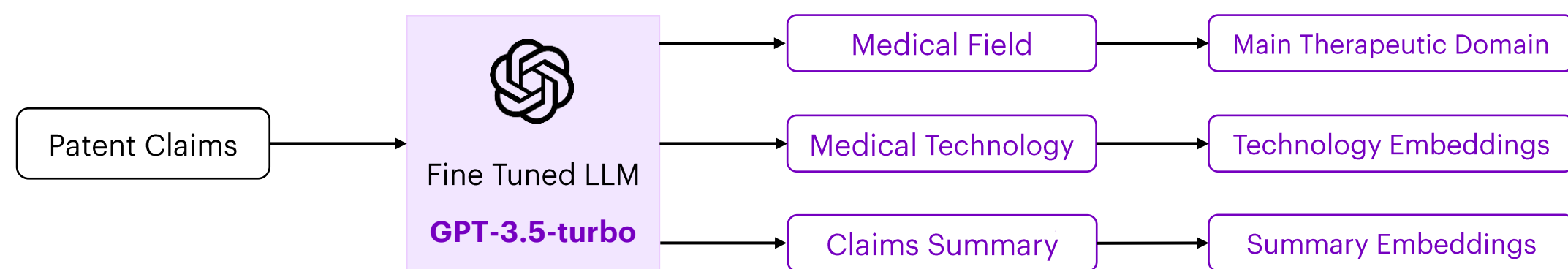Ready to use Pipeline and Dashboard   InnovateGPT: New LLM-based Chat Agent

1. Understanding Innovation

Feature Extraction

Patent claims are lengthy and complex. We use Large Language Models to extract key features.

Patent Claims → Fine Tuned LLM GPT-3.5-turbo →
- Medical Field → Main Therapeutic Domain
- Medical Technology → Technology Embeddings
- Claims Summary → Summary Embeddings

The two main components of a patent: technology and purpose (legal protection).
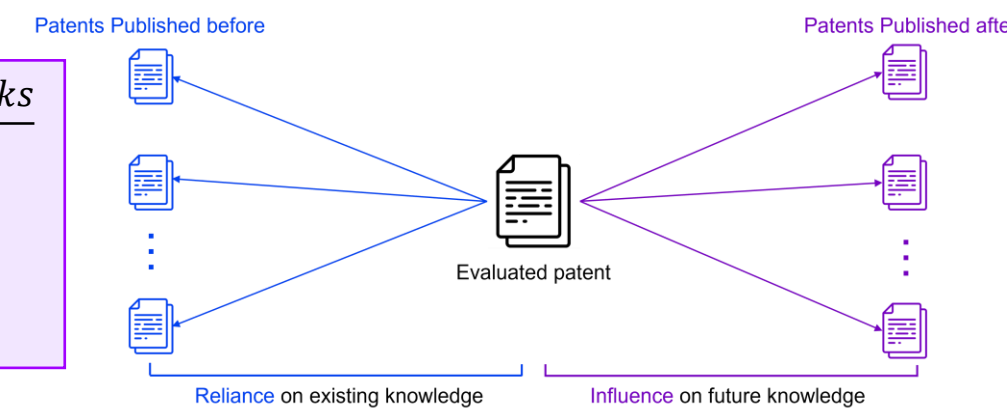
Using fine tuning and prompt engineering, we adapt the LLM to extract key features while respecting the scientific and specific nature of our task.

Medical Fields Grouping

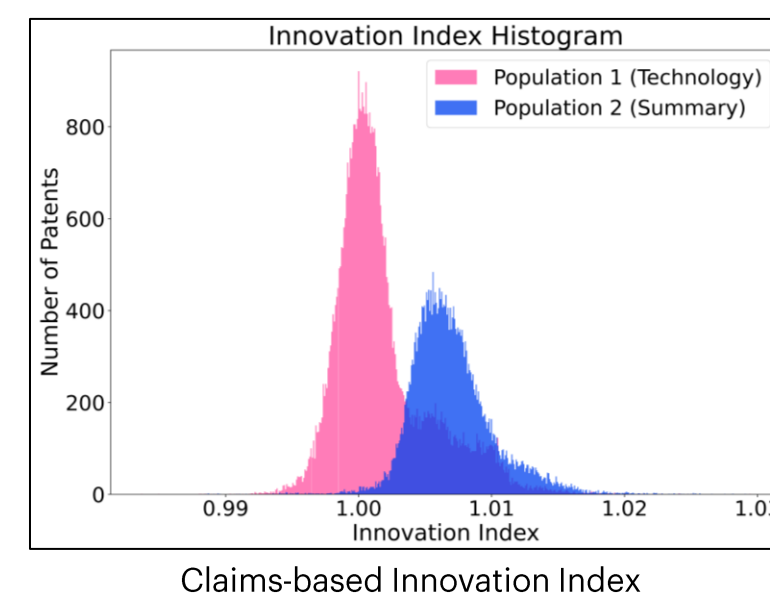319 extracted medical Field → Zero-shot Few-shots Clustering → 14 Main Therapeutic Domains

[Oncology, Chemotherapy, cancer, hematology, hemostasis, thrombosis...] → Oncology and Hematology

+90% Accuracy

Feature extraction and fields grouping examined and validated by industry experts
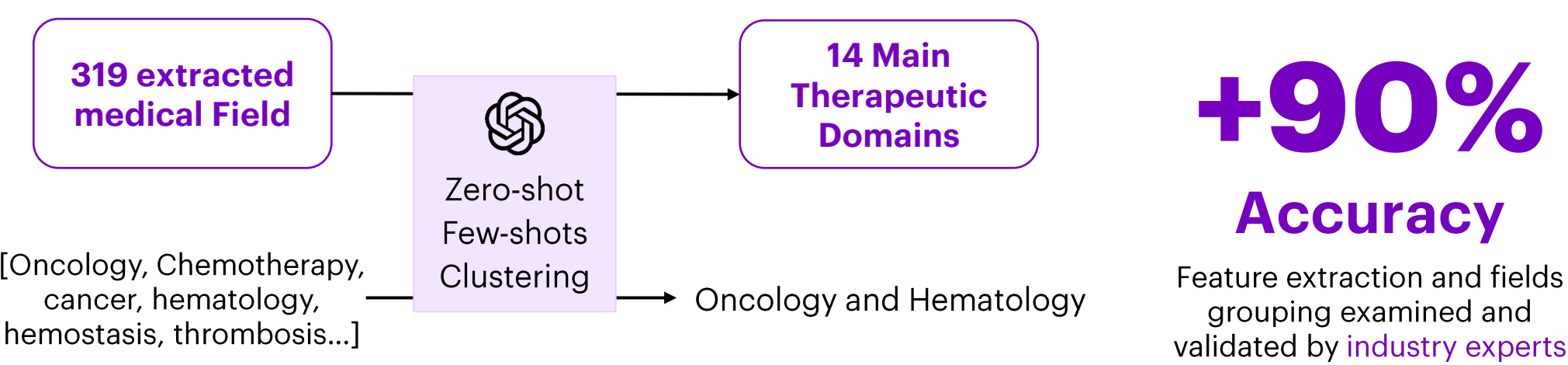
2. Measuring Innovation

$$Innovation\ Index = \frac{How\ influential\ is\ the\ patent\ on\ future\ works}{How\ reliant\ is\ the\ patent\ on\ past\ works}$$

$$= \frac{Average\ Similarity\ to\ all\ future\ works}{Average\ Similarity\ to\ all\ past\ works}$$

Patents Published before ... Evaluated patent ... Patents Published after

Reliance on existing knowledge   Influence on future knowledge

Cosine Similarity is computed using Claims, Technology and Summary embeddings.

Innovation Index Histogram — Population 1 (Technology), Population 2 (Summary) — Claims-based Innovation Index

Gaussian Mixture Models and Weighted Average

Weighted Innovation Index Histogram — Final Innovation Index

Claims-based Innovation Index has a **Bimodal** distribution.

The two populations correspond to the two main features: technology and legal protection (summary).
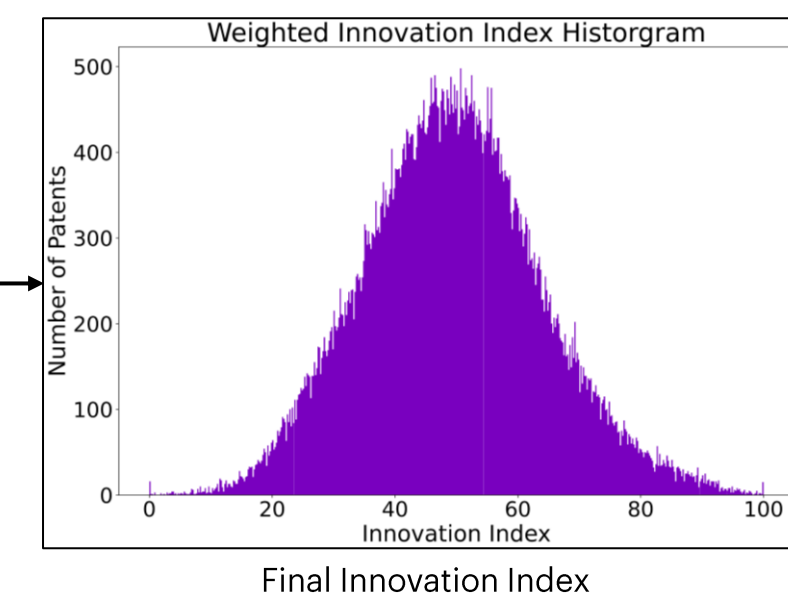
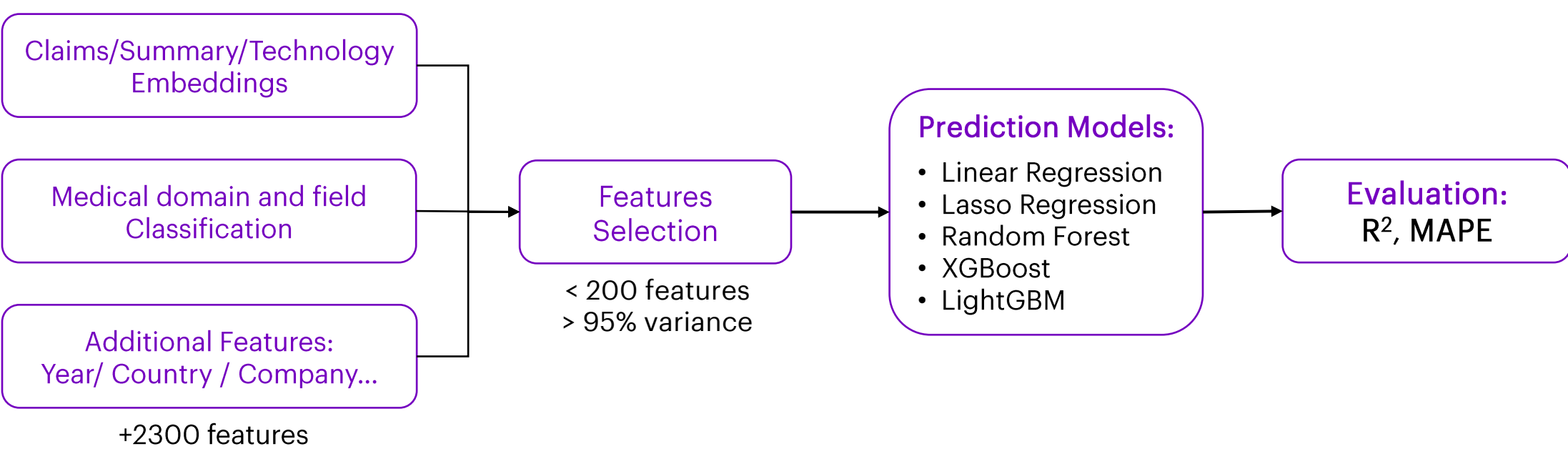The final Innovation Index is a **weighted composed index** that encodes both technology and legal protection.

It is a **relative** index computed on the patent level within the same medical domain and then aggregated to higher levels.

3.1. Harnessing Innovation: Prediction Models

Opportunity: Using the new extracted features and prediction models, we can assess the expected Innovation Index of new patents, projects or assets faster and with high accuracy.

- Claims/Summary/Technology Embeddings
- Medical domain and field Classification
- Additional Features: Year/ Country / Company...

+2300 features

→ Features Selection (< 200 features, > 95% variance) → Prediction Models:
- Linear Regression
- Lasso Regression
- Random Forest
- XGBoost
- LightGBM

→ Evaluation: R², MAPE

Average Out-Of-Sample MAPE
8%

Average Out-Of-Sample R²
0.84

Advantage:
Faster Innovation Index assessment. 30 times faster then running the full pipeline.

Applications:
Projects selection, licensing, M&A valuation, verbiage and legal protection drafting...

3.2. Harnessing Innovation: Portfolio Recommendation

Opportunity: Using the Innovation Index, we can guide Accenture clients' innovation strategies to optimally select the medical domains and the patents or research projects to invest to maximize impact and financial performance.

Step 1 Simplified Formulation

Decisions:
The percentage of budget to invest in each therapeutic domain:
$$X_i, for\ i \in \{Therapeutic\ domains\}$$

Objective:
Maximize the company's overall chances of success in clinical trials and reaching the market as a function of its spends and Innovation Index.
$$Max \sum_i X_i \times Importance_i \times Success\ Factor_i^* \times Innovation_i$$
*Success Factor_i models the relationship between innovation, spends and market success.

Constraints:
- Budget Constraint:
$$\sum_i X_i \leq 1$$
- Diversification Constraints:
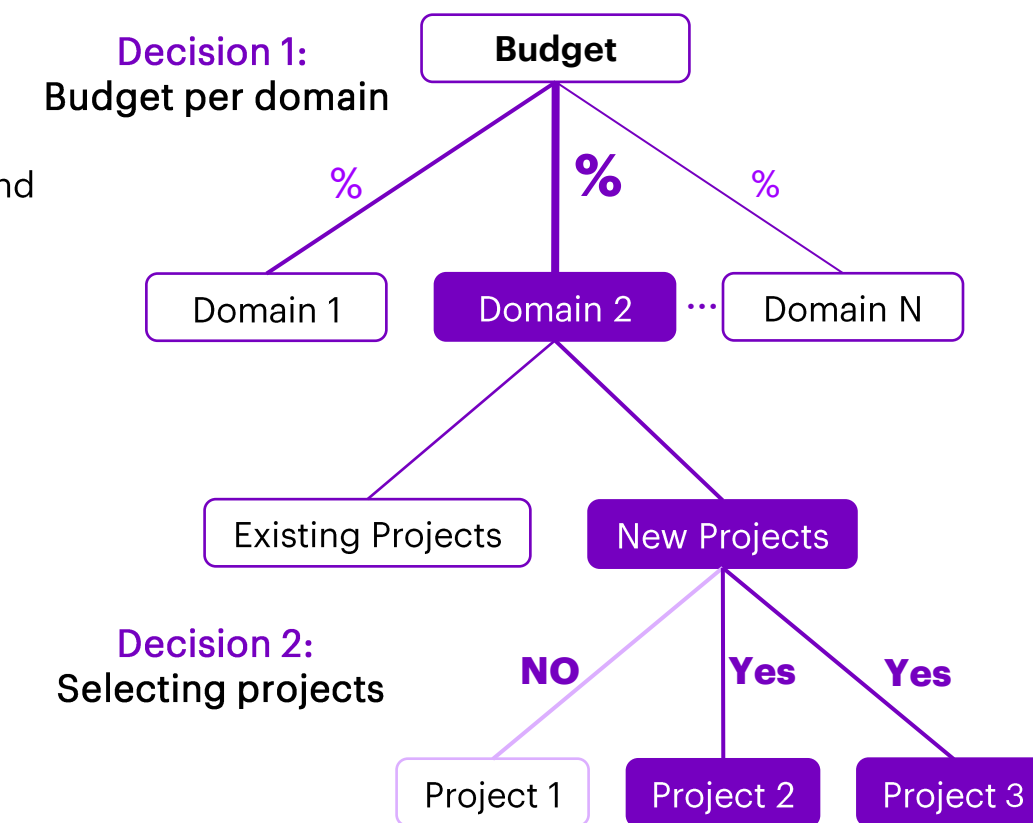Example: Minimum number of domains to invest in
$$\sum_i Z_i \geq n$$
- Innovation Index Constraint:
$$\sum_i Innovation_i \times Importance_i \geq Innovation\ Index + Desired\ Improvement$$
- Budget Expansion Constraints:
$$Historical\ Budget_i - spread_i \leq X_i \leq Historical\ Budget_i + spread_i$$

Two-Step Portfolio Optimization

Decision 1: Budget per domain — Budget → Domain 1, Domain 2, ... Domain N

Decision 2: Selecting projects — Existing Projects, New Projects → Project 1 (NO), Project 2 (Yes), Project 3 (Yes)

Results and Impact

Potential revenue generation: 10+ client meetings to share methodology and identify pharma client-specific actions in the following areas:

- Augment the strategic decisions on innovation strategy including which patents are pursued, their potential future value and how they are positioned in the competitive landscape
- Guide R&D budget allocation decisions
- Support asset valuation for external innovation (e.g., licensing and M&A) by integrating the Innovation Index in asset evaluation
- Improve productivity of patent reviews by +95% (Assumed 2 hours per patent to read, understand and classify for 200K Patents. Processing time reduced from 400K hours to less than 1000 hours), translated into a 199 FTE reduction.

Internal Accenture asset for continuous insights:
- 2-4 global thought leadership reports to strengthen Accenture's value proposition in Life Sciences and beyond
- Expand to other industries to sense future innovation areas and proactively plan offerings and advise clients

Scan for a demo of the fully deployed Dashboard and InnovateGPT, our LLM based Chat Agent.