
HorizonX
CONSULTING
HorizonX Hackathon: LLM Catalog Platform

Agenda

Introduction

Project Overview

Project Requirements

Deliverable

Hackathon Criteria

The Team

CEO AND FOUNDER

Steve Suarez
HorizonX Consulting Founder and CEO
Board Member
External Advisor for Bain & Company
London, United Kingdom

ENGAGEMENT MANAGER

Jonathan Suarez
Engagement Manager for HorizonX Consulting
Boston University Alumni
London, United Kingdom

Agenda

Introduction

Project Overview

Project Requirements

Deliverable

Hackathon Criteria

Lighthouse: A LLM Catalog Platform

Main problem

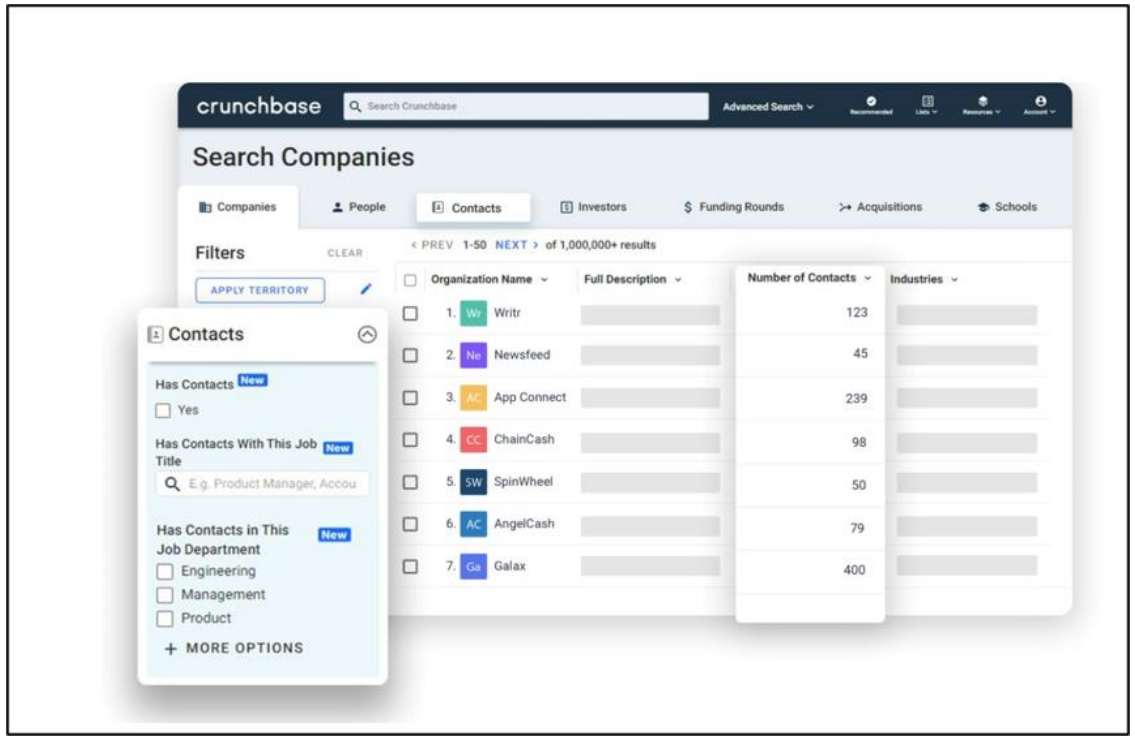
- What LLMs for **regulated industries** (banking, pharma, etc.) are out there?
- Which ones should I look into, which ones should I avoid?

Who encounters these problems

- Market Researchers in businesses
- AI teams
- Governance teams (risk, tech, legal, & compliance)

What can help them

- Centralized repository of LLM information
- Catalog of 400+ LLMs currently in the market
- Insights on specific details describing LLMs e.g. release dates, created by, # of parameters, training data, lawsuits



Platform Similarity

- Think of Crunchbase - Leader in private company data
- Our Platform will display the biggest LLM database
- Stanford Database for platform: Link

Agenda

Introduction

Project Overview

Project Requirements

Deliverable

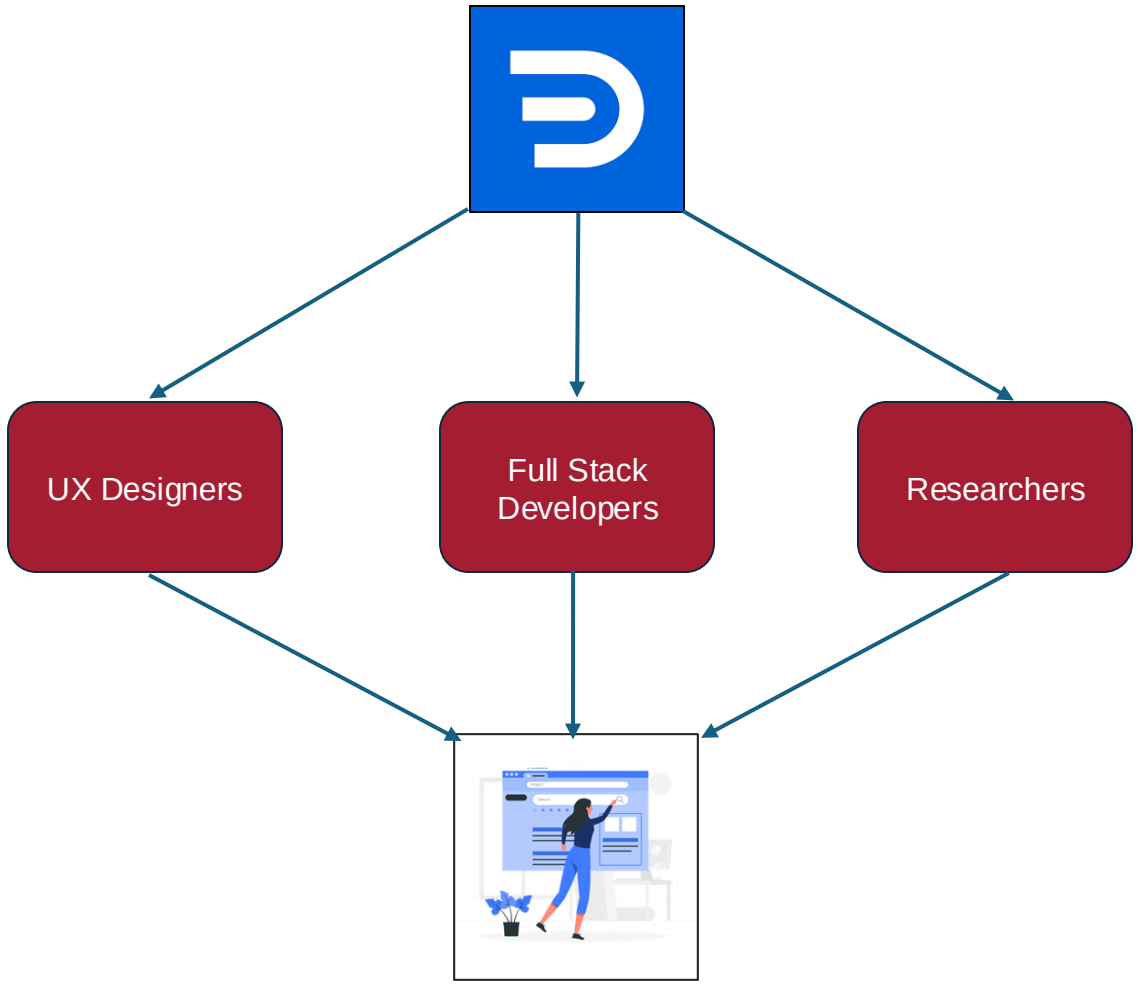
Hackathon Criteria

Resource and Project Requirements

UX Designers
- Design a Large Language Model Catalog with user-friendly visualization.
- Create a baseline report matrix to evaluate LLMs and their value based on accuracy, harm, and framework criteria.

Full Stack Developers
- Develop a centralized repository website for LLM information with insights on specific details such as release dates, creators, parameters, training data, and lawsuits.

Researchers / Data Entries
- Conduct thorough research to compile a catalog of LLMs, including details like release dates, creators, parameters, training data, and legal issues.
- Leverage generative AI for content creation, filling gaps in model lists (as much as possible), and maintaining an up-to-date, comprehensive catalog.



Lighthouse

Agenda

Introduction

Project Overview

Project Requirements

Deliverable

Hackathon Criteria

Deliverable, part 1: LLM Matrix

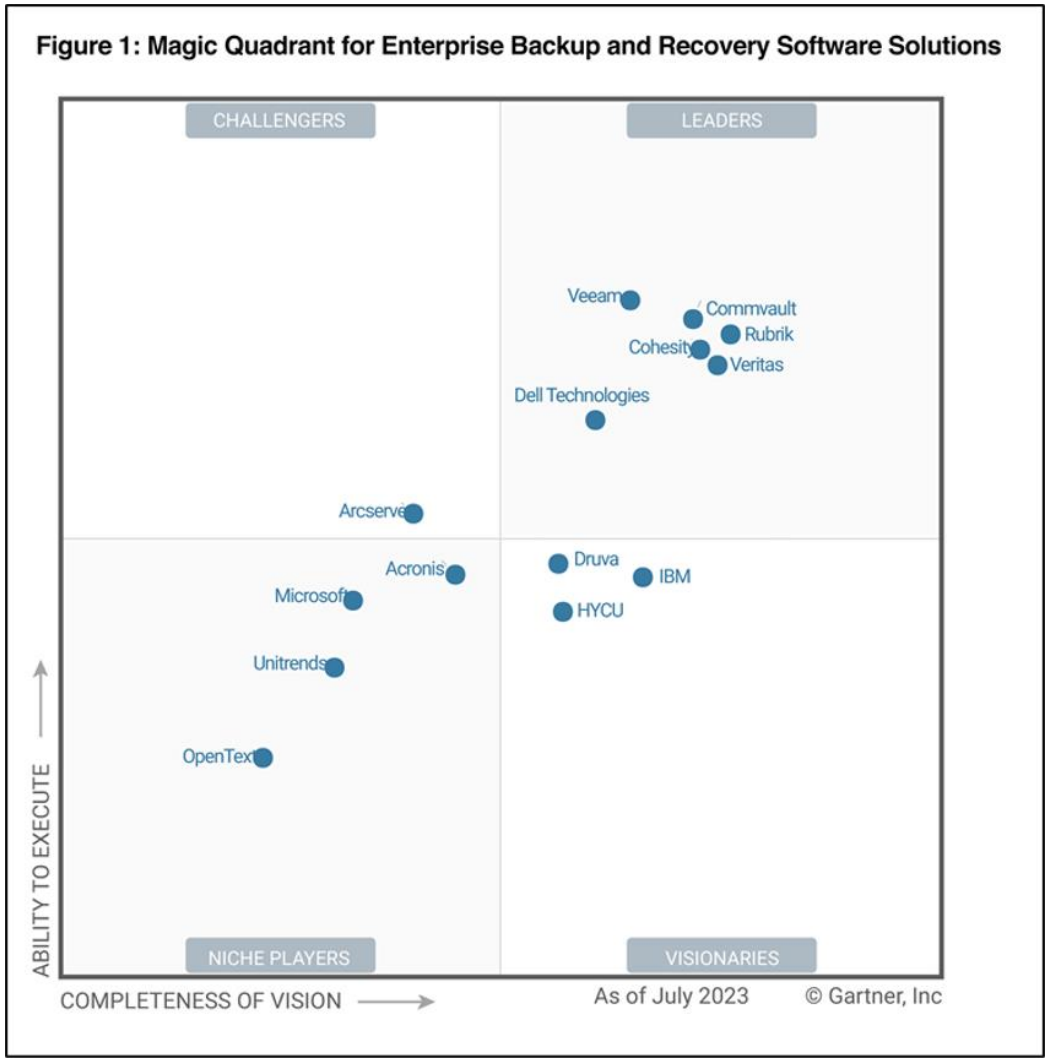
Detailed Requirements

1. Timeline
- **Duration:** 1-1.5 weeks

2. Demo Page
- **Purpose:** The page will showcase a Gartner-like matrix.
- **Content:** The matrix will include several points representing different LLMs. 6-8 LLMs to be plotted (not all of them have to be in the catalog)

3. Total Page(s): 1

4. Matrix Specification
- **Matrix Type:** Similar to Gartner's Magic Quadrant.
- **Points:** Each point represents an LLM.
- **Axes: Business Readiness** and **Perceived Business Value** (see next slide)
- **Creativity:** determine the names of each quadrant, and weigh each important factor of criteria with percentages (see next slide)



Gartner Magic Quadrant: Link

Criteria for Evaluating LLMs

Horizontal Axis: Business Readiness
Quantitative Assessment

Credibility

- Criteria: The reputation and trustworthiness of the LLM, including the organization behind it.
- Evaluation: Assess the LLM based on its track record, endorsements from industry experts, and recognition from reputable institutions.

Harmfulness

- Criteria: The potential for the LLM to produce harmful, dishonest or biased outputs.
- Evaluation: Examine the safeguards in place to prevent harmful outputs, including bias detection and mitigation measures.

Accuracy

- Criteria: The precision and correctness (helpfulness) of the LLM's responses.
- Evaluation: Measure the LLM’s performance on standardized benchmarks and its ability to produce accurate, relevant information consistently.

Benchmark Performance

- Criteria: Performance on industry-standard benchmarks.
- Evaluation: Compare the LLM's scores on benchmarks like GLUE, SuperGLUE, SQuAD, etc., to other models. (room for adaptation how we benchmark!)

Vertical Axis: Perceived Business Value
Qualitative Assessment

Capabilities

- Criteria: The range of functions and features offered by the LLM.
- Evaluation: Consider the LLM’s ability to perform various tasks, such as natural language understanding, generation, summarization, translation, etc.

Success Stories

- Criteria: Documented cases where the LLM has successfully been applied in business scenarios.
- Evaluation: Review case studies, user testimonials, and documented implementations showing tangible business benefits.

Popularity

- Criteria: The widespread adoption and usage of the LLM in the industry.
- Evaluation: Assess the LLM’s market penetration, the number of active users, and the breadth of applications using the LLM.

Example of Evaluation: McDonald's Partnership with IBM

Business Readiness:

Credibility: IBM's AI solutions, backed by a reputable and long-established organization, are widely recognized and endorsed by industry experts, though recent issues highlight areas for improvement.

Harmfulness: While IBM implements safeguards against harmful outputs, the McDonald's drive-thru errors emphasize the need for better error prevention mechanisms in real-world applications.

Accuracy: The AI system's inaccuracies in the McDonald's trial suggest a need for enhanced precision in interpreting complex, real-time interactions.

Benchmark Performance: IBM's AI models perform well on industry-standard benchmarks, although the McDonald's trial suggests the need for benchmarks that better simulate practical, real-world applications.

Perceived Business Value:

Capabilities: IBM's AI, including the Watson suite, offers extensive capabilities across various tasks, but the drive-thru trial indicates limitations in specific contexts like voice recognition in noisy environments.

Success Stories: IBM has numerous success stories across industries, though the McDonald's drive-thru failure contrasts with these successes and highlights variability in performance.

Popularity: Despite the drive-thru trial issues, IBM's AI technology remains popular and widely adopted across multiple industries

Links: Article 1, Article 2

Deliverable, part 2: LLM Catalog

Detailed Requirements

1. Timeline
- **Duration:** 1-1.5 weeks

2. Demo Page
- **Main Catalog Page:** A single page listing the two LLMs.
- **Individual LLM Pages:** Each LLM will have a dedicated page with detailed information.

3. Content and Layout

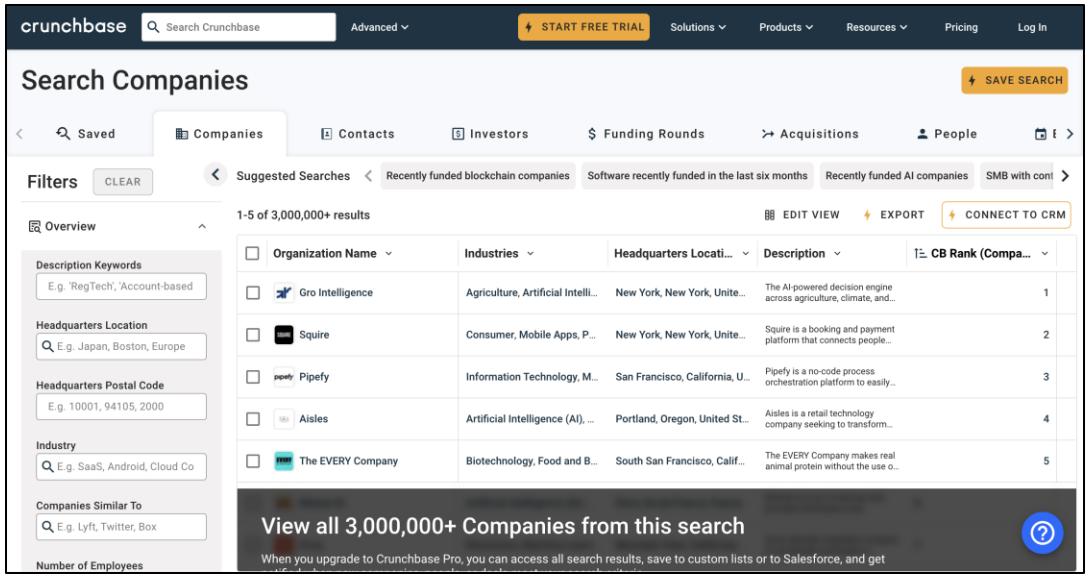
Main Catalog Page:
- Lists two LLMs.
- Each LLM entry will link to its detailed page.

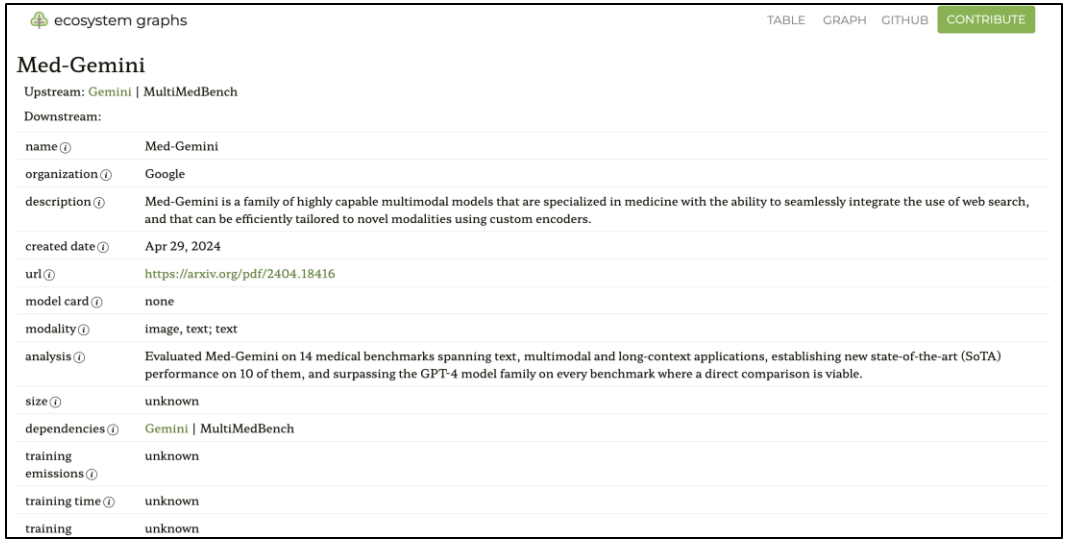
LLM Detail Pages:
- Important details about each LLM (similar to Stanford database)

Creativity:
- Determine which features might be most useful to our audience

4. Total Pages: 3 (1 Catalog Page + 2 LLM Detail Pages)

5. Design Style
- **Similar to Crunchbase:** Clean and professional layout, easy to navigate (See next slide)





Crunchbase Design Characteristics

Clean and Professional Layout

- Minimalist design with consistent typography.
- Ample white space for readability.

Easy Navigation

- Intuitive menu and prominent search bar.

Detailed Content Structure

- Comprehensive profiles with organized sections.

Interactive Elements

- Clickable entries and subtle hover effects.

Content Examples

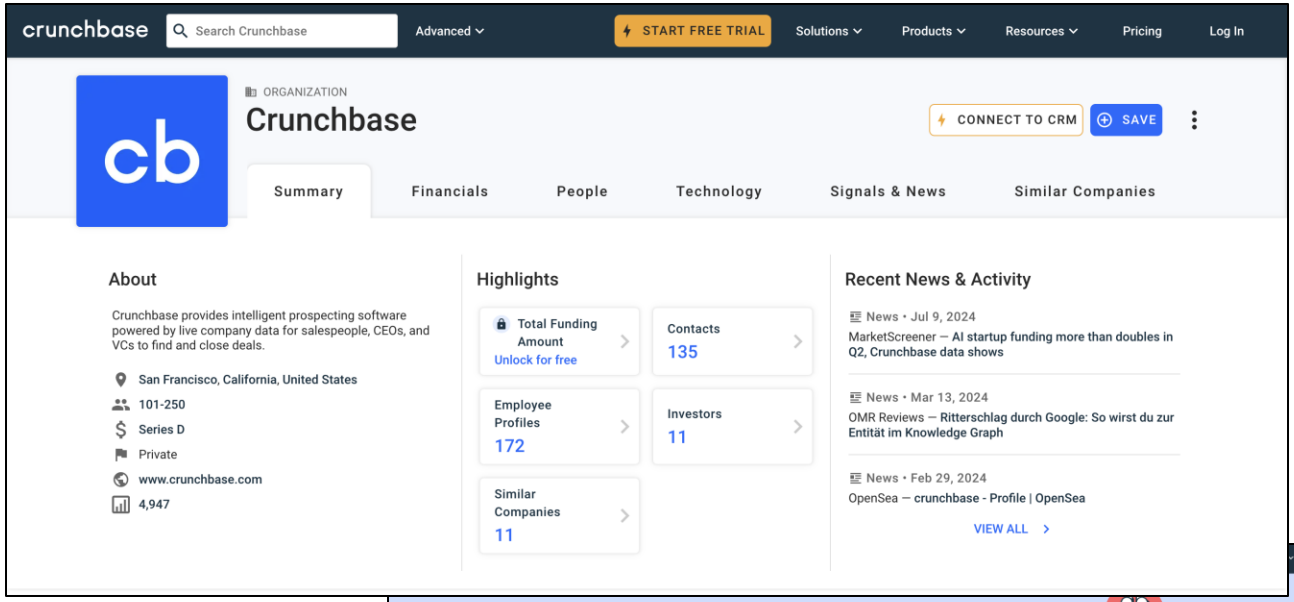
- Overview, Key Features, Technical Specs, Use Cases.

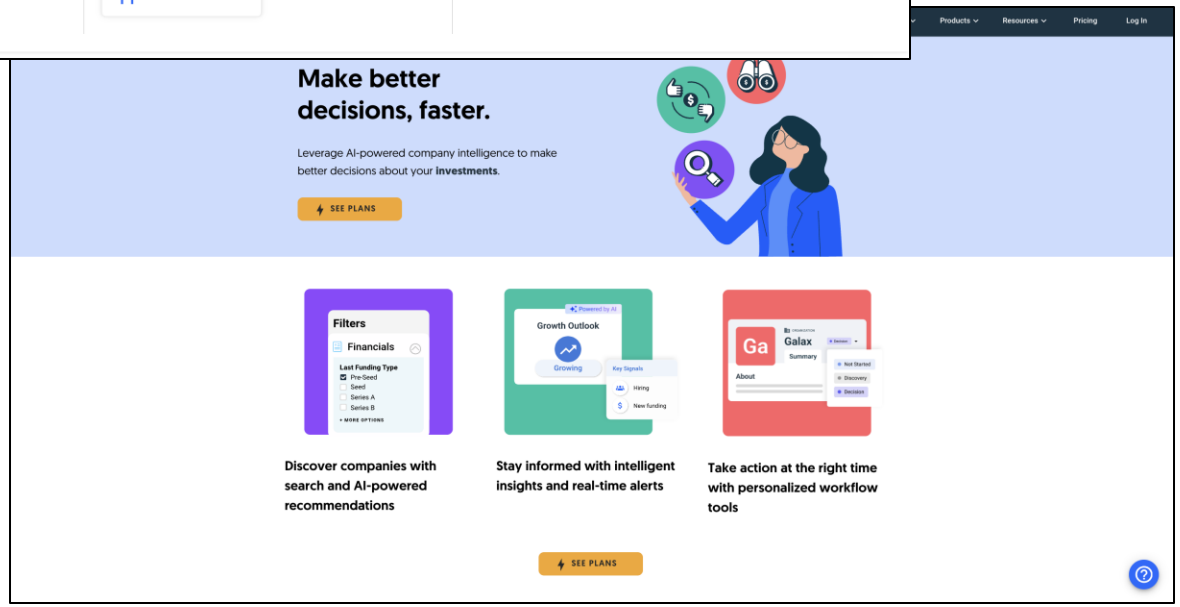
Visual Consistency

- Consistent color scheme, high-quality icons/images.
- Responsive design for all devices.

User Engagement

- User feedback options





Agenda

Introduction

Project Overview

Project Requirements

Deliverable

Hackathon Criteria

HorizonX Hackathon Criteria

Engineering (30%)

Evaluation of the technical aspects and functionality of the project.

Criteria:

- **Code Quality:** Clean, readable, and well-organized code.
- **Functionality:** The project works as intended and meets its goals.
- **Innovation:** Use of new or creative technology/approach.

Content (30%)

Assessment of the information presented within the project.

Criteria:

- **Relevance:** How well the content supports the project's goals.
- **Clarity:** Clear and understandable presentation of information.

Business Development (bonus 10%)

Assessment of the project's potential business impact and feasibility.

Criteria:

- **Market Potential:** The project's potential to succeed in the market.
- **Business Plan:** Feasibility and thoughtfulness of the business strategy.

Visual (15%)

Evaluation of the design and user interface of the project.

Criteria:

- **Design Aesthetics:** Overall visual appeal and creativity.
- **User Interface (UI):** Ease of use and navigation.

Stakeholder Engagement (25%)

Evaluation on how well the participants communicate engage with stakeholders in the project.

Criteria:

- **Clarity of Presentation:** Clear and structured presentation.
- **Engagement:** Ability to engage with stakeholders and maintain communication


HorizonX
CONSULTING
Any Questions?