Product Requirements Document (PRD)

Product: Pharma-Scout: Agentic Innovation Accelerator
Challenge: Techathon 6.0 – Challenge I: Pharmaceuticals
Version: 2.0 (Hackathon Prototype – Competitive Edition)
Date: 2025-09-30

---

1. Goal and Vision

□ Overall Goal

To develop a breakthrough **Agentic AI solution** that drastically reduces the **early-stage product evaluation lifecycle** in pharmaceuticals (2–3 months → under 3 minutes per query). The system leverages multimodal AI agents, knowledge-grounded reasoning, and a dedicated Agentic orchestration server (MCP) to identify viable **molecule repurposing** and **value-added pipeline opportunities**.

✨ Vision

Pharma-Scout transforms pharmaceutical innovation from **manual, fragmented, and reactive research** to an **autonomous, hypothesis-driven agent ecosystem**, enabling:

- **Faster go/no-go** decision making
- **Higher pipeline quality** through deeper data triangulation
- **Interactive insights** for both scientists and executives

---

2. Target Users

User Role	Need/Goal	Key Interaction
Pharma Planner/Strategist	Uncover unmet needs, new products, whitespace opportunities quickly	Input strategic queries & review summary reports
R&D Scientist	Deep analysis of molecule landscape, including MoA, patents, trials	Converses with Master Agent; reviews detailed references
Executive / Portfolio	High-level business impact and ROI overview	Reviews a **5-slider PPT** auto-generated

Mgr		story

---

3. Scope and Key Deliverables

□ In-Scope Features

- **Master Agent Orchestration** (using LangGraph/CrewAI)
- **Hypothesis Generation** (Fine-tuned lightweight SLM)
- **Synthetic Database Integrations:** IQVIA, USPTO, ClinicalTrials.gov (mock APIs)
- **Internal RAG Knowledge Retrieval** from synthetic PDFs
- **Structured Reporting:** exportable reports (PDFs, tables, charts)
- **Executive Deliverable:** Auto-generated **5-slide PPT innovation storyline**

□ Out-of-Scope (Prototype Only)

- Live subscription integrations (will rely on mocks)
- Enterprise user authentication & scaling beyond hackathon demo

---

4. Functional Requirements

FR 4.1: Master Agent (Hypothesis Orchestrator)

ID	Requirement
FR 4.1.1	Accept natural language strategic queries from a React-based UI
FR 4.1.2	Decompose the query into modular subtasks (Market, Patent, Clinical Trials)
FR 4.1.3	Trigger Hypothesis Agent for 3–5 novel molecule repurposing ideas
FR 4.1.4	Delegate subtasks to Worker Agents (mock APIs + RAG retrieval)

FR 4.1.5	Synthesize all Worker outputs into a structured and referenced summary

FR 4.2: Worker Agents & Report Generators

ID	Agent	Responsibility	Output
FR 4.2.1	IQVIA Insights Agent	Simulate querying market size, trends, therapeutics competition	Charts + Tables
FR 4.2.2	Internal Knowledge Agent	Query Vector DB (Chroma/Pinecone mock) for internal documents	Summaries (RAG grounded)
FR 4.2.3	Hypothesis Agent (SLM)	Generate novel repurposing hypotheses	Hypotheses List
FR 4.2.4	Web Intel Agent	Retrieve guidelines, journals, patient forums (simulated web search)	Hyperlinked summaries
FR 4.2.5	Patent Agent	Query synthetic USPTO DB for IP/legal freedom-to-operate	Patent Status Overview
FR 4.2.6	Clinical Trials Agent	Query mock ClinicalTrials.gov data for recruiting/active trials	Tabular Evidence
FR 4.2.7	Report Generator	Final aggregation to **PDF & 5-slide PPT**	Downloadable Reports

---

5. Non-Functional Requirements

NFR 5.1: Performance

- End-to-end query response within **≤180 seconds**
- Ability to process **10 simultaneous queries** in 1 demo batch

NFR 5.2: Architecture & Scalability

- **Agent Orchestration Framework:** LangGraph or CrewAI
- **MCP Server (Dedicated Multi-Cloud Platform):**
    - **Vector DB:** Chroma / Pinecone (mock deployment)

- **Fine-Tuned SLM Deployment:** on GPU endpoint
- **Master Agent Runtime:** containerized (Docker + FastAPI service layer)
- **Streaming Outputs:** via WebSockets for real-time UI updates

NFR 5.3: Security & Data

- All sources **mocked / synthetic** per hackathon rules
- Demonstrated separation of **internal vs external RAG retrieval**

---

6. Full Tech Stack

Frontend (End User Interaction Layer)

- **Framework:** React (TypeScript)
- **UI Toolkit:** TailwindCSS, Material UI
- **App Builder Tool:** Lovable (low-code + React integration)
- **Realtime:** WebSockets for streaming agent responses
- **Visualization:** Recharts/D3.js for trends, charts, graphs

Backend (Agent Orchestration + APIs)

- **Framework:** FastAPI in Python
- **Agent Orchestration:** LangGraph / CrewAI
- **Worker Agents:** Modular containerized microservices
- **Report Generator:** Python + ReportLab / pptx Backend

Dedicated MCP Server Setup

- **Containerization:** Docker + Kubernetes (for modular microservices)
- **Vector Store:** Chroma / Pinecone mock instance
- **Fine-Tuned Models (SLM):** Hosted on NVIDIA GPU instance
- **Knowledge Retrieval:** RAG via LangChain + internal vector DB
- **Data Sync:** Mock APIs simulating IQVIA, USPTO, ClinicalTrials

7. Demonstration Scenario

Example Query (Planner):

"Which respiratory diseases show low competition but high patient burden in India?"

1. User inputs query via React UI
2. **Hypothesis Agent** generates 3–5 candidate molecules
3. **Master Agent Orchestration** triggers research subtasks
4. **RAG Retrieval** pulls internal strategy PDFs for molecule X
5. Worker agents fetch data: Market → Patent → Clinical
6. Synthesized report generated (with citations)
7. Final PDF + Auto-PPT (5-slide narrative) delivered in-app

---

8. Success Metrics

Metric	Target
Orchestration	Trigger all 7 Worker Agents + synthesize correctly
Time-to-Solution	≤180 seconds per complex research query
Data Grounding	100% of claims linked to mock/synthetic references
Deliverables	Professional PDF + 5-slide PPT auto-generated for executive
Innovation Recognition	Judges identify clear differentiation vs simple RAG demos

Expanded Use Cases & Strategic Questions

🧬 A. Molecule Repurposing & Pipeline Expansion (“New Product” Questions)
Agents: Hypothesis Agent (SLM), Clinical Trials Agent

- Orphan Drug Repurposing
  What are the top three Orphan Disease indications that share a similar Mechanism of Action (MoA) to our anti-inflammatory drug, Molecule X?
- Dosage Form Innovation
  Beyond oral administration, what are the technical and IP feasibility flags for developing Molecule X into an inhalable dry powder formulation for respiratory diseases?
- Combination Therapy Screening
  Which second active ingredient (Co-Drug Y), if co-formulated with Molecule X, would maximize efficacy against Disease Z while maintaining a favorable safety profile?
- Shelved Asset Revitalization
  For Molecule A (failed Phase II due to cardiac issues), does recent literature suggest a link to any neurodegenerative disease pathways?
- Pediatric Opportunity Assessment
  What are the unmet pediatric indications where Molecule X is currently used off-label?
  What is the market size potential in the EU for a formal pediatric approval?

📈 B. Market Strategy & Competitive Intelligence (“Go/No-Go” Questions)

Agents: IQVIA Insights Agent, Web Intelligence Agent

- Therapy Area White-space
  Which three therapy areas show high patient burden and pricing potential but have <5 active Phase III competitor trials globally?
- Generic Erosion Analysis
  For Drug Z, what are the five highest-value markets where patent expiry is imminent (≤3 years)? What does the generic competitive landscape look like?
- Target Market Prioritization
  Between Brazil, India, and Mexico, which country has the highest CAGR projection for a generic launch of Molecule Y, considering local regulations & competition?

- Competitor Product Deep Dive
  For competitor Drug P, what are its:
  - Projected sales trends
  - Primary marketing claims (from news/forums)
  - Any active post-market trials
- Market Entry Strategy
  What are the primary reimbursement challenges and KOL sentiment regarding Molecule X in the US market, as inferred from forums and news?

⚖ C. Regulatory & IP Risk Management ("Red Flag" Questions)
Agents: Patent Agent, Internal Knowledge Agent

- Freedom-to-Operate (FTO) Check
  Does a new sustained-release delivery system for Molecule X infringe on any active patents filed by our top two competitors in the US & China?
- Internal Document Compliance
  In our internal MINS quality manuals, what are the storage stability requirements for Molecule X at +10% concentration?
  Do we have the necessary equipment for compliance?
- Patent Extension Strategy
  What method-of-use or formulation patents could be filed immediately for Molecule X to extend lifecycle beyond the primary composition patent expiry?
- Regulatory Risk Assessment
  If pursuing a new dosage form, what are the FDA/EMA bridging study & bioequivalence guidelines that apply?
- Adverse Event Signal Detection
  Have recent papers or news reported new/unusual adverse events for Molecule X that need immediate pharmacovigilance review?

🧫 D. Clinical Trials & Evidence Generation ("Trial Optimization" Questions)
Agents: Clinical Trials Agent, Internal R&D Knowledge Agent

- Patient Population Profiling
  For a Phase II trial in Disease Z, what are the typical inclusion/exclusion criteria?
  Which patient demographics show the highest response rates?
- Trial Site Selection
  Which clinical sites/sponsors have the most experience & enrollment success for trials involving drugs with an MoA similar to Molecule X (past 3 years)?
- Endpoint Validation
  In completed Phase II trials for Disease Z, which secondary endpoints correlated most strongly with primary success endpoints?
- Trial Failure Analysis (Internal)
  From internal R&D docs (RAG), what were the top 3 root causes of failure in the Phase I trial of Molecule B?
  How can those risks be mitigated for Molecule X?
- Biomarker Identification
  What are the most commonly referenced prognostic biomarkers that could stratify patients for a more targeted Phase II trial of Molecule X?