

Position Title: Volunteer AI Engineering Intern – Agents, LangGraph, FastAPI, and Agentic RAG

Organization: PredictaBio Innovations

Location: Remote

Type: Unpaid Internship (Part-Time, Flexible)

Duration: 2–3 months (with potential for extension)

About PredictaBio

PredictaBio Innovations is an AI-first biotechnology startup building modular systems that combine large language models, agent-based orchestration, and scientific data pipelines to accelerate research workflows in bioprocess development and life sciences. Our platform is focused on transforming unstructured biomedical information into structured, retrieval-ready knowledge that can support downstream reasoning, recommendation, and decision-making.

As we expand our engineering efforts, we are seeking volunteer AI Engineering Interns who are excited to work on real-world agentic AI systems, scalable backend services, and retrieval pipelines for scientific and technical documents.

Position Overview

We are looking for highly motivated Volunteer AI Engineering Interns to help build and improve our next-generation AI infrastructure. In this role, you will work on agentic workflows, document processing pipelines, FastAPI-based services, and modular microservices that support RAG and agent-based decision systems.

This position is ideal for students or early-career engineers who want hands-on experience developing production-oriented AI systems using modern frameworks such as LangGraph, MCP-compatible tools, FastAPI, Docker, PostgreSQL, and retrieval pipelines for scientific literature.

Key Responsibilities

- Build and refine agent-based workflows for document understanding, information extraction, and reasoning tasks.
- Develop backend services using FastAPI for AI pipeline orchestration and internal tool integration.
- Design and implement Agentic RAG pipelines over scientific PDFs, structured metadata, and extracted document chunks.
- Work with LangGraph or similar orchestration frameworks to create multi-step, tool-using AI agents.
- Build modular microservices for ingestion, extraction, retrieval, ranking, and response generation.
- Contribute to document processing workflows involving PDFs, XML, JSON, and biomedical text sources.
- Integrate AI services with PostgreSQL and retrieval layers for storing structured outputs and metadata.
- Containerize services using Docker and support reproducible local development workflows.
- Collaborate on internal tools and MCP-compatible interfaces for connecting models, tools, and data sources.
- Participate in code reviews, system design discussions, and weekly team check-ins.

Desired Qualifications

- Strong Python programming skills.
- Experience building backend APIs using FastAPI or similar frameworks.
- Familiarity with LangChain, LangGraph, or agent-based LLM orchestration frameworks.

- Understanding of RAG architectures, chunking, retrieval, and structured extraction workflows.
- Experience with Docker and modular service development.
- Comfortable working with PostgreSQL and structured/unstructured data.
- Exposure to document processing workflows involving PDFs, XML, or JSON.
- Interest in LLM systems, tool-using agents, and applied AI engineering.
- Strong debugging, problem-solving, and self-learning ability.
- Prior work samples, GitHub projects, or prototypes are a plus.

Application

If you are interested in this position, please send your resume along with a brief motivational paragraph outlining your interest and relevant experience to **predictabiogeneral@gmail.com**.