Internship Report: Deep Venture Partners LLC

Martina Paez Berru

IEOR E4999 Fieldwork — Columbia University

Summer 2025 (June 23 – August 29, 2025)

Executive Summary

During the summer of 2025, I completed a full-time internship as a Data Science and Engineering Intern at Deep Venture Partners LLC (DeepVP), a venture capital firm developing AI-driven platforms for technology scouting and investment workflows. My responsibilities encompassed developing end-to-end automation systems that integrated large language model (LLM) inference pipelines with web scraping, structured data ingestion, and automated presentation generation. I played a central role in integrating and orchestrating the Google Slides API and Google Drive API into this pipeline, including authentication, placeholder management, template logic, and debugging edge cases. I also designed and tested a preview mechanism for users to manually edit LLM-generated slide content before saving.

The goal of this work was to make the process of producing professional investor presentations faster, more intelligent, and fully automated. I developed the logic for handling JSON-based content generated by LLMs, validated the structure against predefined schemas, and ensured accurate population of dynamic content into templated slides. I implemented interactive previews that mimicked the collaborative and editable behavior of Google Slides, ensuring an intuitive handoff from automated generation to human refinement.

This internship was remote and conducted from New York City. It provided an academically rigorous experience that allowed me to apply techniques from data engineering, machine learning, and software architecture to a real-world investment application. I was supervised by the firm's CTO and worked in close collaboration with product managers and AI researchers.

1.   Introduction

The objective of this internship was to engineer scalable, modular systems that utilized AI and data pipelines to streamline internal fund workflows, with a focus on automating early-stage investment prospecting. The experience was designed to reinforce core skills from my graduate studies in business analytics, data science, and machine learning, and to deepen

my exposure to applied innovation in venture capital.

Deep Venture Partners is pioneering a new investment thesis around university innovation as an undercapitalized asset class. Their strategy is built on the idea that scientific breakthroughs from university research represent an untapped source of high-potential innovation that can be translated into commercial ventures. By building proprietary tooling to identify, ingest, and evaluate these research outputs, DeepVP is developing a scalable approach to sourcing and accelerating breakthrough technologies.

For corporate research and development (R&D) departments and limited partners (LPs), it's often difficult to track innovation that matters worldwide (especially as federal research funding declines and scientific projects seek more private-sector support). DeepVP's platform intervenes at this juncture by transforming academic patents and abstracts into structured insight, matching them with business teams, and generating tailored materials that support investment decisions, strategic partnerships, and innovation scouting.

What if we could identify these opportunities early, and translate them into business action? That's what Deep Venture Partners is building toward.

2. Technical Projects and Responsibilities

2.1. Google Slides Automation with LLM Integration

I led the development of a containerized FastAPI-based microservice for Google Slides automation. This microservice ingested structured LLM output (validated JSON) and generated templated slide decks using a dynamic Google Slides template. The workflow was designed to operate asynchronously to support multi-user scalability and non-blocking slide generation.

Key components of the pipeline:

- **Prompt-to-JSON generation**: Used templated prompts to elicit structured, schema-constrained content from multiple LLMs.

- **Structured schema validation**: Validated LLM outputs against JSON schemas for universities, companies, and scientists.

- **Google Slides API integration**: Dynamically inserted LLM-generated content into a fixed template using placeholders (e.g., {{university_name_1}}).

- **Google Drive API orchestration**: Managed publication, access control, and shared links for generated slides.

- **OAuth 2.0 authentication**: Implemented token-based access control for authorized users using DeepVP credentials.

- **Asynchronous slide generation**: Used Celery and Redis to handle background processing and expose preview links via REST API.

- **Interactive preview system**: Implemented a web-based preview mechanism for manual editing of content prior to final download.

- **Containerization and deployment**: Created a Dockerfile and configured the service via `docker-compose`. Kafka integration was outlined in future design tickets but not implemented during this internship.

- **Frontend integration**: Exposed slide links in the frontend using ReactJS and embedded them in an `<iframe>`.

This pipeline reduced deck generation time from hours to minutes, saving dozens of hours per month.

2.2.   Schema Design and Content Pipeline Coordination

In parallel, I supported the upstream and downstream components of Workflow 2:

- **LLM Coordination and JSON Schema Management**: Defined strict JSON schemas and validated outputs for ingestion.

- **Ticket-Based Full-Stack Collaboration**: Contributed to ERDs, API endpoints, and frontend components based on Notion tickets.

- **LLM-to-Frontend Debugging**: Fixed issues in placeholder rendering, authentication, and UI behavior.

- **Product Ownership**: Made architectural decisions about placeholder logic and slide preview design.

3.   Organizational Fit and Impact

My work directly supported DeepVP's core mission: using AI to improve early-stage investment decisions. I enabled the firm to scale up its evaluation of university-based innovations.

I collaborated with seven other Columbia IEOR graduate students and our project manager, Spencer Pao. We met weekly with the founders, Ben Kaplan, Dan Hoffer, and Luc Vincent, to align technical progress with business goals.

I contributed extensively to the web application interface used to generate investor decks, helping reduce manual workload for partners. My focus was on ensuring reliable content formatting, placeholder mapping, and LLM integration.

Through continuous iteration, I influenced product decisions, implemented preview logic, and aligned the interface behavior with the intent of the user. I led architectural planning and execution while receiving regular feedback from the founders, which was highly rewarding.

4.   Educational Relevance

This internship complemented my graduate studies and reinforced both technical and collaborative skills. I gained experience across:

- **Data Engineering**: ETL pipelines for unstructured data.

- **Software Architecture**: Modular API and LLM integration.

- **Machine Learning**: LLM prompt engineering and inference.

- **Frontend Development**: UI features for slide previews.

- **LLM Engineering**: Prompt design and output validation.

- **DevOps Tools**: Docker, GitHub, and multi-service orchestration.

- **Data Modeling**: Relational schema design and ER diagrams.

- **Team Tools**: Notion for planning and documentation.

- **API Integration**: End-to-end Google Slides and Drive workflows.

5.   Conclusion and Reflection

My internship at Deep Venture Partners was a defining experience in my academic and professional development. It allowed me to translate theoretical knowledge into production-grade systems and observe how data science drives real business value. I gained both technical depth and strategic perspective in a high-growth, innovation-oriented environment.

The support and mentorship I received during the internship were instrumental in my learning. The guidance of my supervisor, the collaboration with my peers, and the consistent feedback from the founders created a learning environment that was rigorous and uplifting. The energy of the team made each challenge feel like an opportunity and the feedback I received during weekly meetings with the founders was a source of motivation.

I learned essential tools and workflows that I now feel confident applying to future roles. From Docker and API orchestration to front-end debugging and product iteration, I have gained invaluable skills that will shape my job search and career direction.

Looking ahead, I am excited to continue exploring the technologies and workflows I worked on. While my work was focused on Workflow 2, I would have loved the chance to dive into Workflow 1, which centered on web scraping and ingesting large-scale data from university tech transfer sites. I plan to continue studying and experimenting with the LLMs and agents we tested as a team, especially to understand their strengths, limitations, and application patterns. I also hope to contribute to future iterations of this platform beyond the internship, bringing greater structure and intelligence to how scientific innovation is translated into business value.