Compliance Aware, Agentic Pitch Book Generation For Bankers.

An all-in-one content generation tool providing agentic workflows with RAG search capabilities.

A Queen Mary Final Year Dissertation Interim Report By:
Name: Andrei Rizea
ID: 220300153

Supervised By:
Manesha Peiris

Programme of Study:
Digital & Technology Solutions (Software Engineering)

Module:
IOT635W Project Interim

Contents

Contents.... 1
Problem Statement.... 2
Project Aims.... 2
Organisational.... 2
Educational.... 2
Objectives.... 4
Risks & Mitigation.... 4
Literature Review.... 5
Project Milestones.... 5
Technologies.... 6
References.... 7

Problem Statement

Pitch books (PBs), used in sales presentations to win client business, are key to Investment Banking (IB) yet junior bankers spend nearly 50% of their time creating them (Introhive, 2021). PBs are highly repetitive in structure, data, and phrasing, with little differences apart from routine contextual updates. This makes for an ideal candidate for AI automations. However, this inefficiency persists across the industry, raising the question of why this process has not been streamlined yet.

This introduces three problems:

Reduced productivity
A lot of time is spent manually creating PBs, refreshing data, and iterating on style. This moves the attention away from analytical and client-focused work, reducing productivity.

Scattered precedence material (PM)
Since each banker creates their own PBs, PM (historic deal data and PBs used for reference) is scattered and individualised resulting in no centralised sources, introducing data retrieval delays.

Tough learning-curves
As PBs depend on bank specific templates, data-sourcing processes and narrative expectations, analysts often struggle to produce slides correctly on their first attempts. This results in repeated cycles potentially delaying deadlines.

Project Aims

Organisational

The project aims to deliver a platform that automates PB generation while providing a centralised source of PM. With Multi-Media-Generation (MMG), Retrieval-Augmented-Generation (RAG) and agentic workflows, PBs can be generated, data sourcing streamlined hence, significantly reducing the manual work required.

Educational

We'll explore how AI methods, specifically RAG, agentic workflows, and MMG can be integrated to address real inefficiencies. MMG presents unique challenges as it must combine narratives, data, visuals, and slide design into coherent outputs. Agentic workflows further raise concerns such as how to integrate into existing systems taking into account permissions and responsibilities as it's difficult to distinguish between a human and an agentic user.

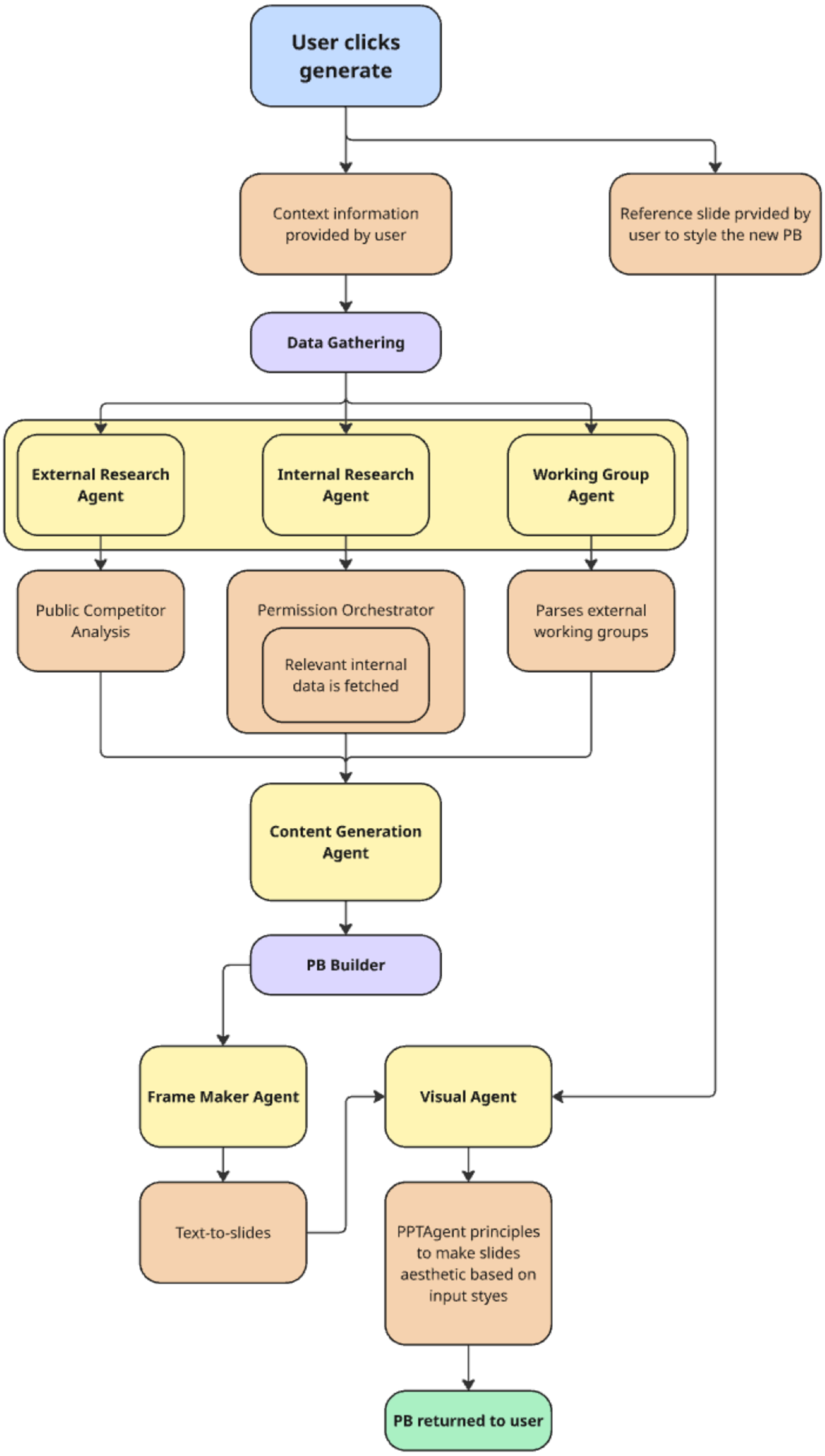
User clicks generate
Context information provided by user
Reference slide prvided by user to style the new PB
Data Gathering
External Research Agent
Internal Research Agent
Working Group Agent
Public Competitor Analysis
Permission Orchestrator
Relevant internal data is fetched
Parses external working groups
Content Generation Agent
PB Builder
Frame Maker Agent
Visual Agent
Text-to-slides
PPTAgent principles to make slides aesthetic based on input styes
PB returned to user

Fig.1 - Abstracted diagram of the proposed flow of the application

Objectives

PB Generation
The application should automatically generate PBs in varying formats, requiring only minimal user input. For example, deal context (company, date range, sector etc). This is to ensure the model is accurately grounded.

Agentic workflows
During generation, the solution must be able to research company and financial data from public or vendor data sources automatically. It should search the web and other sources such as vendor data.

Precedence material
Storage of previously generated PBs is important, providing a single source for PM, simplifying data retrieval. A RAG search should then query across all this historical data.

Risks & Mitigation

Risk	Likelihood	Consequence	Mitigation
AI risks introducing biases and factual inaccuracies. This could lead to reputational and ethical implications.	High	High	RAG systems can be used to narrow down the context, grounding it to truthful information. Human-in-the-loop validation should be introduced to ensure no inaccuracies remain (IBM, 2024). Furthermore disclaimers should be added to remove all liability of misinformation.
Bankers have different data access meaning if incorrect data is accessed, compliance issues arise.	Medium	High	Authentication flows must be built. In cases of access glitches, compliance should be made aware immediately
PBs are created amidst the release of business news with banks in a race to secure lead roles. If the system is down, it would affect the business significantly.	Medium	High	Cloud deployments will ensure new instances can be created upon failures. Comprehensive tests then ensure the functioning of the application.

Literature Review

Anderson, (2025) states that "Business communication runs on presentations". The IB industry is an old-school business relying on direct communication. Studies show that analysts spend ~50% of their time creating powerpoints. Through data automations, such as the research agent proposed, we can cut that down by 30% (Introhive, 2021).

AI slide generation became popular in 2014 focusing on text-to-slide. Recently, there are more models that generate visually appealing slides which, Anderson (2025) states, produces radical time savings to bankers.

PPTAgent specialises in 2 aspects. It follows reference designs and continuously evaluates each slide (Zeng, et al., 2025). PPTAgent excels on content, design and narrative, addressing a gap as past models focused solely on text quality. Indico Labs, (2023) states the time investment required for an initial PB drafting could then be reduced by up to 80% with such automations.

Bandyopadhyay, et al (2024) states that naive AI approaches used to improve PB efficiency, such as basic text-to-slides, lack the narrative required for persuasion. The authors suggest placing Large-Language-Models (LLM) and Visual-Language-Models (VLM) into one structured workflow. Gemini-3-Pro will be used for this application, a VLM with superior multi-modal understanding (Dextralabs, 2025). The model should then be grounded with RAG to reduce factual inaccuracies (Lewis. et al. 2020).

In terms of evaluating the application, traditional accuracy scores are unsuitable as they cannot understand the narrative, creative quality, nor any tangible business improvements. Hence, productivity-based measures will be used. Sauermann (2023) states that effective productivity metrics must be "objective, as opposed to subjective thoughts" and that they must be quantifiable. So, we will evaluate based on time savings, fewer re-works, and total weekly PBs per analyst which is more appropriate, reflecting real productivity gains.

Project Milestones

An agile Software-Development-Lifecycle (SDLC) methodology is being used. This is because it emphasises rapid iteration and close user feedback-loops. Prototypes will be shown to power users for feedback after each phase and relevant literature will be reviewed continuously which will help guide the project.

1. Scope & Literature review (Met)
   - Defining the problem statement, aim and conducting a preliminary literature review.
2. Foundation of application (Weeks 8-10, in-progress)
   - Initial setup allowing users to call mock APIs, used to improve UX.
3. Core AI (Weeks 10-12)
   - Basic text-to-slide functionality using GPT5.1 providing the MVP.
4. VLM (Weeks 13-15)
   - Designs mimicking users reference pitchbooks using PPTAgent principles.
5. Agents (Weeks 16-18)

    - Research agent workflows for data sourcing.
6. RAG (Weeks 19-20)
    - RAG search against PM.
7. Testing & Refinement (Weeks 21-22)
    - Unit, integration, and acceptance testing with compliance seal of approval.
8. Report finalisation (Week 23-24)

Technologies

Frontend
React Typescript will be used, deployed via Vercel rather than AWS due to its one-click simple deployments. Typescript provides code durability using strict typing while React allows for component driven development.

Database
Supabase is a PostgreSQL database. It provides file storage, authentication and Atomicity-Consistency-Isolation-Durability (ACID) compliant transactions ensuring reliability. It also supports complex queries and has extensive documentation helping integration and debugging.

Backend
Python is used over Java because it's the standard language for AI development. It supports libraries such as Hugging Face and TensorFlow, allowing for integration with different models.

API
FastAPI provides asynchronous I/O handling allowing for multiple simultaneous requests ensuring stability, chosen over GraphQL which is a heavyweight service requiring schemas and resolvers, deployed via Render due to its simplicity.

AI
GPT5.1 (LLM) and Gemini-3-pro (VLM) will be used as they rank above all other companies for the most benchmarks (LCouncil, 2025; Vellum-AI, 2025). OpenAI Agent builder will be used allowing workflows to be built and triggered via their Python library. Extensive documentation from all, also simplifies integrations.

References

Anderson, (2025) ‘The Evolution of AI Presentation Generators for Business Slides (and Why Autoppt Stands Out)’. Available at: https://autoppt.com/blog/ai-presentation-generators-evolution-why-autoppt (Accessed: 11 October 2025).

Bandyopadhyay. et al. (2024) ‘Enhancing Presentation Slide Generation by LLMs with a Multi-Staged End-to-End Approach’, pp. 222–229. Available at: https://research.adobe.com/publication/enhancing-presentation-slide-generation-by-llms-with-a-multi-staged-end-to-end-approach (Accessed: 18 November 2025).

Dextralabs. (2025) ‘Top 10 Vision Language Models in 2025’, Dextralabs. Available at: https://dextralabs.com/blog/top-10-vision-language-models/ (Accessed: 18 November 2025).

Introhive. (2021). Cracking the myth of long investment banker hours. Introhive. Available at: https://www.introhive.com/blog/cracking-the-myth-of-long-investment-banker-hours (Accessed: 18 November 2025).

Indico Labs. (2023). The benefits of automating PowerPoint reports. Available at: https://www.indicolabs.io/blog/benefits-of-automating-powerpoint-reports (Accessed: 17 November 2025).

IBM. (2024) ‘What is human in the loop (HITL)?’. Available at: https://www.ibm.com/think/topics/human-in-the-loop (Accessed: 18 November 2025).

Lewis, P. et al. (2020) ‘Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks’. Available at: https://arxiv.org/abs/2005.11401 (Accessed: 18 November 2025).

LCouncil. 2025. *AI Model Benchmarks Nov 2025. Comparing GPT-5, Claude 4.5, Gemini 3, Grok 4*. Available at: https://lmcouncil.ai/benchmarks (Accessed: 17 November 2025).

Sauermann, J. (2023), Performance measures and worker productivity, IZA World of Labor. Available at: https://wol.iza.org/articles/performance-measures-and-worker-productivity (Accessed: 18 November 2025).

Vellum AI. 2025. *LLM Leaderboard*. Available at: https://www.vellum.ai/llm-leaderboard (Accessed: 17 November 2025).

Zheng, H., Guan, X., Kong, H., Zheng, J., Zhou, W., Lin, H., Lu, Y., He, B., Han, X. and Sun, L. (2025) ‘PPTAgent: Generating and Evaluating Presentations Beyond Text-to-Slides’. Available at: https://doi.org/10.48550/arXiv.2501.03936 (Accessed: 2 October 2025).