Capstone Project on Generative AI for Domain-Specific Business Requirement Document (BRD) Generation

---

Project Title:

Intelligent BRD Generator for Domain-Specific Applications (Pharma and Finance)

---

1. Objective

Develop a **Generative AI-based model** capable of producing complete and contextually rich **Business Requirement Documents (BRDs)** tailored to domain-specific standards, particularly for **Pharmaceutical** and **Financial** sectors. The model should understand the nuances of industry terminology, compliance mandates, and business processes to generate accurate and professional documentation.

---

2. Problem Statement

Organizations in regulated domains like Pharma and Finance spend substantial time drafting BRDs that align with compliance, domain jargon, and stakeholder expectations. These documents often require domain experts and multiple iterations before approval.

The challenge is to build an **AI-powered system** that can: - Understand industry context and regulatory frameworks (e.g., FDA in Pharma, Basel III in Finance). - Generate BRDs automatically from minimal user input (e.g., product description or business problem). - Produce structured, readable, and compliant documents including all key sections such as: - Project Overview - Business Objectives - Functional & Non-Functional Requirements - Key Performance Indicators (KPIs) - Compliance & Risk Sections

This project aims to **reduce the BRD creation time by 70%**, while maintaining **domain accuracy** and **compliance readiness**.

---

3. Project Scenario

Scenario A – Pharma Domain

A pharmaceutical company is planning to launch an internal application to automate adverse event tracking during clinical trials. The company needs a BRD to define system objectives, compliance requirements (21 CFR Part 11), and integration with electronic health record (EHR) systems.

The AI model should be able to: - Understand terms like *adverse event reporting, trial monitoring, patient data anonymization*. - Include mandatory compliance sections like *HIPAA* and *GxP*. - Generate sections outlining objectives, scope, and regulatory traceability.

Scenario B – Finance Domain

A financial institution aims to build a credit risk analytics platform to predict loan default probabilities. The BRD must describe business objectives, model logic (credit risk scoring), KPIs (e.g., default rate, recovery rate), and compliance with *Basel III* and *GDPR*.

The AI model should be able to: - Include relevant financial terminologies like *collateral value, probability of default (PD), loss given default (LGD)*. - Address compliance and audit readiness. - Generate a structured, ready-to-review BRD with KPIs and regulatory elements.

---

4. Technical Architecture

Components:

1. **Frontend Interface:** Streamlit or Flask-based input form for domain, context, and requirements.
2. **Backend Model:**
3. Google Gemini / OpenAI GPT-4 model fine-tuned on domain data.
4. Domain-specific datasets (Pharma: FDA documentation, trial protocols; Finance: BRDs, compliance docs).
5. **Database:** Firestore / BigQuery to store generated BRDs and metadata.
6. **Template Engine:** Jinja2 or custom template layer for generating formatted Word/PDF outputs.
7. **Deployment:** Google Cloud Vertex AI or Cloud Run for hosting the API and frontend.

Architecture Flow:

1. User selects domain (Pharma/Finance).
2. Enters project summary and basic inputs (objectives, stakeholders, etc.).
3. System prompts model to generate BRD draft.
4. Output undergoes validation (domain keywords, compliance tags).
5. Document saved and downloadable in Word format.

---

5. Key Features

- **Domain-Aware Prompting:** Model trained with specific vocabulary and regulatory contexts.
- **Customizable Sections:** Ability to toggle sections (KPIs, Functional Requirements, etc.).
- **Compliance Auto-Tagging:** Detects and inserts compliance clauses.
- **Template Management:** Predefined templates for Pharma, Finance, or custom domains.
- **Document Export:** Output in DOCX or PDF format for stakeholder circulation.

---

6. Evaluation Metrics

Metric	Description	Target
Accuracy	Alignment of BRD content with domain standards	> 85%

Metric	Description	Target
Relevance	Domain-specific terminology correctness	> 90%
Compliance Coverage	Inclusion of necessary regulatory references	> 90%
Time Efficiency	Time saved vs manual drafting	70% reduction
User Satisfaction	Qualitative feedback from SMEs	>= 8/10

---

7. Deliverables

- Domain-adaptive AI model (Finance and Pharma variants)
- Streamlit-based user interface for BRD input/output
- Sample BRDs generated for both domains
- Technical documentation (architecture, model details)
- Word/PDF output templates

---

8. Hands-On Activities

1. **Dataset Preparation:**
2. Collect sample BRDs and compliance frameworks (FDA, Basel III, etc.).
3. Annotate domain-specific terminology.
4. **Model Fine-Tuning:**
5. Use a generative model (Gemini or GPT-4) with domain examples.
6. Test prompts like:
    - “Generate a Pharma BRD for clinical trial management.”
    - “Draft a Finance BRD for credit scoring automation.”
7. **Validation:**
8. Check generated BRDs for structure, terminology, and completeness.
9. **Output Testing:**
10. Export generated BRDs into Word format.
11. Verify document readability and regulatory completeness.

---

9. Supportive Guide for Participants (with Hints)

Step 1: Environment Setup

- Install required libraries: `pip install google-generativeai streamlit python-docx`
- Set up your Gemini API key inside your code for authentication.
- **Hint:** Use environment variables or a config file instead of hardcoding the key for better security.

Step 2: Domain Understanding

- Pharma: Review compliance standards such as FDA 21 CFR Part 11, HIPAA, and GxP.

- Finance: Review Basel III, GDPR, and credit risk calculation methods.
- **Hint:** Create a glossary file with frequently used terms for each domain to help fine-tuning and context preservation.

Step 3: Prompt Engineering

- Start with concise prompts; expand gradually by adding context and output structure.
- Example:

```
prompt = "Generate a detailed Pharma BRD for an adverse event tracking
system compliant with FDA and HIPAA."
```

- **Hint:** Add 2–3 sample BRDs in the input context for better structure learning (few-shot learning).

Step 4: Model Integration

```
import google.generativeai as genai

genai.configure(api_key="YOUR_GEMINI_KEY")
model = genai.GenerativeModel('models/gemini-2.5-pro-preview-03-25')
response = model.generate_content(prompt)
print(response.text)
```

- **Hint:** Use exception handling and logging to capture API rate limits or unexpected errors.

Step 5: Export Output

```
from docx import Document

doc = Document()
doc.add_heading('Business Requirement Document', level=1)
doc.add_paragraph(response.text)
doc.save('Generated_BRD.docx')
```

- **Hint:** Apply proper formatting such as bold headings, numbered lists, and section dividers to make the document presentation-ready.

Step 6: Validation & Evaluation

- Verify compliance-specific keywords (e.g., “Basel III,” “21 CFR Part 11”).
- Compare generated BRDs against authentic templates.
- **Hint:** Develop a BRD checklist template to validate coverage, clarity, and compliance.

Step 7: Optional Advanced Tasks

- Implement version tracking with Firestore or Git.
- Enable feedback integration to iteratively refine future document generations.
- **Hint:** Log user edits to identify recurring gaps in generated outputs.

---

10. Expected Outcome

A fully functional prototype capable of generating **comprehensive, compliant, and domain-specific BRDs** for Pharma and Finance sectors. The generated BRDs will feature: - Customizable sections and structure. - Domain-specific terminology and compliance adherence. - Export-ready Word documents suitable for review and approval workflows.