Assignment
Reimagining a Classic in a New World

Drive Link-
https://drive.google.com/drive/folders/1jJkbyc1JXmgDrIQw6zyn4KkjTFjZBC0J?usp=sharing

Pipeline Flow:

1. **Input Parsing**

- Source PDF is loaded and text is extracted page-by-page.
- Text is chunked into manageable segments to respect model context limits.

2. **Schema Extraction**

- Each chunk is passed through a strict schema-extraction prompt.
- Output is written incrementally to a structured plain-text schema file.

3. **World & Character Mapping**

- The extracted schema is mapped onto a user-defined target world.
- Character roles, motivations, and relationships are preserved.

4. **Plot Transformation**

- Original plot beats are re-expressed within the target world.
- Narrative order, escalation, and resolution style remain unchanged.

5. **Outline Generation**

- Transformed plot beats are organized into a three-act structure.

6. **Long-Form Story Generation**

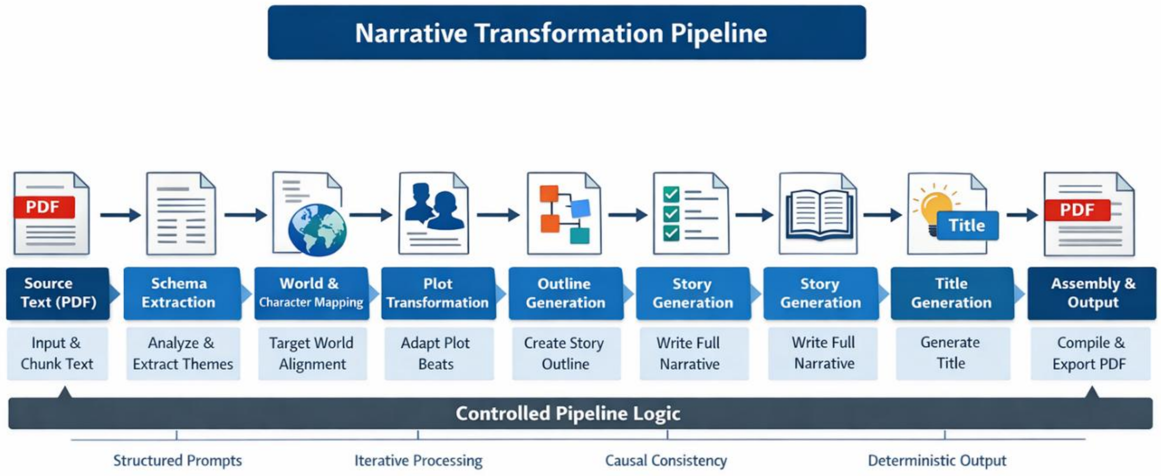
- A full prose narrative is generated from the outline.

7. **Title Generation**

- A concise title is produced based on the completed story.

8. **Assembly & Output**

- Final title and story are assembled and exported as a PDF.



Solution Design

Input Handling

The pipeline begins by reading a source PDF and extracting raw text using a PDF parser. Text is then chunked to ensure that no individual LLM call exceeds context constraints. Chunking also ensures graceful scaling to longer source texts.

Schema-First Extraction

Rather than prompting the model to reinterpret or rewrite the text immediately, the system first extracts a **loss-minimal narrative schema**.
The schema includes:

- Themes
- Tone
- Setting and world type
- Characters and roles

- Ordered plot beats
- Resolution style

This extraction is governed by a rigid prompt that forbids inference or invention

Schema extraction

. Temperature is set to zero to maximize determinism and consistency.

Controlled Transformation

Once extracted, the schema becomes the **single source of narrative truth**. All subsequent steps reference this schema rather than the original text.

World mapping

- **Character mapping** re-anchors characters into a new world while preserving narrative function and emotional intent

. Plot transform

- **Plot transformation** rewrites plot beats into the new setting while enforcing causal consistency and prohibiting structural changes

.

Outline

The transformed plot is then elevated into a structured three-act outline, which serves as a blueprint for long-form prose generation.

Generation

The final story is produced using a novelist-style prompt optimized for:

- Length (3000–4000 words)
- Emotional pacing
- Scene-based expansion
- Internal consistency

Title generation

A short, evocative title is generated last, ensuring that it reflects the completed narrative rather than guiding it prematurely

.

Challenges & Mitigations

- **Schema Extraction Reliability:** Saving in JSON/YAML format were inconsistent, so a strict plain-text (.txt) schema with tightly defined prompts was used to prevent formatting errors and inference.
- **Large File Processing:** Single-pass processing was slow and unstable, resolved by chunking the source text to avoid context limits.
- **Performance Constraints:** Latency was reduced by using a lightweight model for extraction and transformation, reserving heavier generation for final output.

Edge Cases

Source PDFs may contain licensing text, page numbers, author information, or front/back matter. The extraction prompt explicitly restricts output to narrative content only, ignoring non-story elements.

Story will be dependent on the LLM you will chose (Large LLM will produce better, but will take more time).

Future Improvement

Persistent Narrative Memory

- Store schemas and transformations in a database
- Enable iterative revisions and branching narratives (Use of LangGraph)

Interactive UI

- Visual pipeline editor
- Editable schemas and plot beats
- Real-time preview of narrative changes

Evaluation & Validation Layer

- Automated checks for missing plot beats
- Character consistency validation