Technical Documentation – Custom Content Generator Tool

# 1. Tool Architecture

The system is composed of the following components:

|  |  |
| --- | --- |
| Tool | Used for |
| Figma UI | Front-end design prototype. |
| Input Handler | Captures user inputs such as subject, topic, and format. |
| Prompt Generator | Constructs prompts based on user inputs. |
| Gemini API | Processes the prompt and returns the generated content. |
| Response Formatter | Formats and filters the AI output. |
| Exporter | Allows saving outputs to text or PDF. |
| Performance Tracker | Logs generation time and token usage. |
| Error Handler | Manages API failures and invalid inputs. |

# 2. API Selection Rationale

|  |  |
| --- | --- |
| Criteria | Gemini API |
| Cost | Free tier available via Google Cloud. |
| Performance | Quick response time and accurate outputs for educational prompts. |
| Features | Supports structured prompting and filters inappropriate content. |
| Ease of Use | REST-based API with simple integration and documentation. |
|  |  |

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

# 3. Prompt Engineering Methodology

The prompt engineering strategy includes:  
- Role-based prompting (e.g., 'Act as a science teacher')  
- Structured instructions and format-based guidance  
- Few-shot prompting when needed  
- Iterative refinement to improve clarity, tone, and relevance  
- Use of placeholders for customization parameters (e.g., [topic], [subject])

# 4. Performance Optimization Techniques

|  |  |
| --- | --- |
| Technique | Description |
| Token minimization | Reduced unnecessary words in prompts. |
| Content filtering | Removed disclaimers and off-topic content. |
| Smart defaults | Auto-filled commonly used inputs to reduce form time. |
| Output trimming | Removed lengthy explanations or repeated data. |
|  |  |

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

# 5. Error Handling

The tool includes basic error handling for:  
- API failures: Displays user-friendly error message and logs the issue.  
- Invalid inputs: Validates that required fields are completed before submission.  
- Output errors: Allows users to regenerate the response.

# 6. Rate Limits & Usage Costs

The Gemini API used in this tool has the following limitations and costs:  
- Free tier allows up to 60 requests per minute.  
- Maximum tokens per request: ~8192 tokens (combined input and output).  
- Premium usage may incur costs depending on the Google Cloud billing tier.  
- Developers are encouraged to monitor token usage to avoid exceeding quotas.