Gemini AI App – Image Describer

Overview

The **Gemini AI App – Image Describer** is a simple HTML-based web application designed to generate **textual descriptions of uploaded images** using **Google's Gemini AI API**.
It allows users to upload one or more images, specify generation parameters (like temperature, top-p, and top-k), and submit prompts for contextual or guided image description.

This tool is ideal for testing, research, and internal AI workflows that require **vision-language processing**.

Architecture

High-Level Flow

1. **Frontend Interface:**
    - Built using **HTML, CSS, and JavaScript**.
    - Users can adjust generation parameters (max tokens, temperature, top-p, top-k) and specify the Gemini model.
    - Supports **multiple image uploads** and custom prompts.
2. **Backend API Integration:**
    - Sends requests to the **Gemini AI API endpoint**.
    - Each request includes:
        - Uploaded images (base64 or multipart form data)
        - Text prompt
        - Model parameters (temperature, top-p, top-k, max tokens)

- Receives structured AI-generated text describing the image.

3. **Output:**
    - The generated response is displayed directly on the page.
    - No database connection is required; all interactions are handled in-memory or client-side.

System Diagram

```
[User Interface (HTML Page)]
         |
         v
 [JavaScript Frontend Script]
         |
         v
 [Gemini AI API (Model: gemini-1.5-flash)]
         |
         v
 [AI Response]
         |
         v
 [Textual Description Displayed to User]
```

Tech Stack

Component	Technology Used
Frontend	HTML, CSS, JavaScript
AI API	Google Gemini API
Model	`gemini-1.5-flash`
Backend (optional)	Node.js / Python Flask / PHP for API key handling

Hosting	Any static or lightweight web server
Data Storage	None (no database needed)

User Interface Elements

Element	Description
Max Output Tokens	Defines the maximum length of the generated response.
Temperature	Controls the creativity of the response (lower = more focused, higher = more random).
Top P	Nucleus sampling parameter to limit token diversity.
Top K	Restricts sampling to the top-K probable tokens.
Model	Selectable model input (default: `gemini-1.5-flash`).
Upload Image	File input to upload one or more image files.
Prompt Text Area	Allows the user to provide a guiding prompt for the image description.
Send Prompt Button	Triggers the API call to generate the description.

Setup Instructions

1. Prerequisites

- **Google Gemini API Key** from Google AI Studio.
- Basic web server environment (for local or hosted deployment).

2. Configuration

Add your API key in your backend or JavaScript code:

```
const API_KEY = "YOUR_GEMINI_API_KEY";
```

-

Set your preferred default model and parameters:

```
 const model = "gemini-1.5-flash";
const temperature = 0.6;
const top_p = 0.4;
const top_k = 60;
const max_output_tokens = 2048;
```

-

API Request Example

```
const formData = new FormData();
formData.append("file", imageFile); // from file input

const payload = {
  contents: [
    {
      role: "user",
      parts: [
        { text: document.querySelector("#prompt").value },
        { inline_data: { mime_type: imageFile.type, data: imageBase64
} }
      ]
    }
  ],
  generationConfig: {
    temperature: 0.6,
    topP: 0.4,
    topK: 60,
    maxOutputTokens: 2048
  }
};

fetch("https://generativelanguage.googleapis.com/v1beta/models/gemini-
1.5-flash:generateContent?key=" + API_KEY, {
  method: "POST",
  headers: { "Content-Type": "application/json" },
```

```
  body: JSON.stringify(payload)
})
.then(res => res.json())
.then(data => {
  document.querySelector("#output").innerText =
data.candidates[0].content.parts[0].text;
});
```

Output

- The AI model returns a **natural language description** of the uploaded image(s).
- The result is shown on-screen and can be optionally copied or saved by the user.