ATS - API documentations

**Gemini API Usage Report**

**1. Token Usage Analysis**

Based on your provided data:

* **Input Tokens:**
  + **Resume:** 1 page ≈ 1,000 tokens
  + **Prompt:** Approximately 250 tokens
    - **Prompt means if do one action like summarize , Generate JD with AI , Percentage match then that action Gemini take approximately 250 tokens**
  + **Total Input:** 1,000 + 250 = 1,250 tokens
* **Output Tokens:**
  + **Model Response:** Approximately 250 tokens
* **Total Tokens per Interaction:** 1,250 (input) + 250 (output) = 1,500 tokens

*Note: Token estimation is based on an average of 1.3 to 1.5 tokens per word.*

**2. Free Tier Usage Details**

As a user of the free tier, you have access to the following models with specific limitations:

* **Gemini 2.0 Flash:**
  + **Input Price:** Free
  + **Output Price:** Free
  + **Context Caching:** Free, up to 1,000,000 tokens of storage per hour
  + **Grounding with Google Search:** Free, up to 500 requests per day
  + **Data Usage:** User data may be utilized to improve Google's products
* **Gemini 1.5 Flash:**
  + **Input Price:** Free
  + **Output Price:** Free
  + **Context Caching:** Free, up to 1,000,000 tokens of storage per hour
  + **Grounding with Google Search:** Not available
  + **Data Usage:** User data may be utilized to improve Google's products

*Note: The free tier is primarily intended for testing and development purposes, with lower rate limits and potential data usage for product improvement.*

**3. Paid Tier Pricing Details**

For production-level applications requiring higher performance and additional features, the paid tier offers enhanced capabilities:

* **Gemini 2.0 Flash:**
  + **Input Price:** $0.10 per 1 million tokens (text/image/video); $0.70 per 1 million tokens (audio)
  + **Output Price:** $0.40 per 1 million tokens
  + **Context Caching:** $0.025 per 1,000,000 tokens (text/image/video); $0.175 per 1,000,000 tokens (audio)
  + **Context Caching Storage:** $1.00 per 1,000,000 tokens per hour
  + **Grounding with Google Search:** 1,500 requests per day free, then $35 per 1,000 requests
  + **Data Usage:** User data is not used to improve Google's products
* **Gemini 1.5 Flash:**
  + **Input Price:** $0.075 per 1 million tokens (prompts ≤ 128k tokens); $0.15 per 1 million tokens (prompts > 128k tokens)
  + **Output Price:** $0.30 per 1 million tokens (prompts ≤ 128k tokens); $0.60 per 1 million tokens (prompts > 128k tokens)
  + **Context Caching:** $0.01875 per 1,000,000 tokens (prompts ≤ 128k tokens); $0.0375 per 1,000,000 tokens (prompts > 128k tokens)
  + **Context Caching Storage:** $1.00 per hour
  + **Grounding with Google Search:** $35 per 1,000 requests (up to 5,000 requests per day)
  + **Data Usage:** User data is not used to improve Google's products

**4. Cost Estimation for Your Usage**

Assuming an interaction involves 1,500 tokens (as calculated above):

* **Gemini 2.0 Flash:**
  + **Input Cost:** 1,250 tokens × ($0.10 / 1,000,000 tokens) = $0.000125
  + **Output Cost:** 250 tokens × ($0.40 / 1,000,000 tokens) = $0.0001
  + **Total Cost per Interaction:** $0.000125 + $0.0001 = $0.000225
* **Gemini 1.5 Flash:**
  + **Input Cost:** 1,250 tokens × ($0.075 / 1,000,000 tokens) = $0.00009375
  + **Output Cost:** 250 tokens × ($0.30 / 1,000,000 tokens) = $0.000075
  + **Total Cost per Interaction:** $0.00009375 + $0.000075 = $0.00016875

*Note: These calculations are based on the assumption that prompts are ≤ 128k tokens. Costs will increase for larger prompts.*

**TOTAL APP USAGE**

APP processes approximately **1,850 tokens** per resume (1-page) when performing the following actions:

* **Resume Input:** 1,000 tokens
* **Generate GD with AI:** 350 tokens
* **Summarize Prompt:** 250 tokens
* **Percentage Match Calculation:** 250 tokens