## ## Project Overview: iZACH

**iZACH** is an advanced, Python-based **AI Desktop Assistant** inspired by the concept of JARVIS (Iron Man). Unlike standard voice assistants, iZACH is designed to be a **multimodal agent** that can see the user's screen, understand context, control system hardware, and execute complex automations with high speed and sarcastic wit.

### ### 1. Project Scope & Vision

The vision for iZACH is to bridge the gap between "chatbots" and "operating systems."

* **The Idea:** A seamless, hands-free interface where the user can command the computer as if they were talking to a human collaborator.
* **Scope:** \* **Visual Intelligence:** Real-time screen analysis to "read" what the user is doing.
  + **Hardware Control:** Direct interaction with mouse, keyboard, and system volume.
  + **Entertainment Automation:** Deep integration with Spotify and YouTube for background tasks.
  + **Proactive Assistance:** Moving beyond simple "replies" to actual "actions" (clicking, typing, searching).

### ### 2. Core Architecture (The Tech Stack)

We have optimized the stack for **low latency** and **high precision**, especially handling Windows-specific display scaling.

| **Component** | **Technology Used** |
| --- | --- |
| **Main AI Brain** | **Google Gemini 2.5 Flash** (High-speed multimodal) |
| **API Client** | google-genai (The latest official Google SDK) |
| **Vision/Eyes** | PIL (Pillow) & ImageGrab for DPI-aware screenshots |
| **Hardware Control** | pyautogui for instant coordinate-based clicking |
| **Voice Recognition** | SpeechRecognition (Google Web API) |
| **Voice Synthesis (TTS)** | pyttsx3 (Offline SAPI5 for zero-latency speech) |
| **System Interaction** | ctypes (for DPI awareness), psutil (health monitoring) |
| **UI Framework** | customtkinter (Modern Dark-themed GUI) |

### ### 3. Progress Milestones (What is Working)

We have successfully overcome several critical technical hurdles:

* **Multimodal Integration:** Switched from Groq/Llama to Gemini 2.5 Flash to ensure stable vision and chat support.
* **DPI & Scaling Correction:** Implemented a **DPI-aware coordinate system** that works perfectly on laptop screens with **125% zoom**.
* **Normalized Precision Mapping:** iZACH uses a 0–1000 coordinate grid to map the AI's spatial reasoning to actual screen pixels.
* **Instant Interaction:** Replaced slow "sliding" mouse movements with **instant hardware-level clicks** for a "live" feel.
* **Contextual Memory:** Enabled chat sessions that remember previous commands within a session.
* **Media Automation:** Built modules to search and play specific songs on Spotify and YouTube, including "background mode."

### ### 4. Future Roadmap (The "iZACH Evolution")

* **Double-Click & Drag Support:** Enhancing vision modules to handle folder navigation and file moving.
* **Voice Switching:** Integrating a more natural, human-like voice (possibly via ElevenLabs or Gemini's native TTS).
* **Gesture Control:** Using a webcam to trigger commands via hand movements (Iron Man style).
* **Deep File Indexing:** Allowing iZACH to search through local documents and summarize PDFs on the fly.
* **Self-Correction:** A feedback loop where iZACH "checks" if a click was successful by looking at the screen again after an action.

### ### 5. Technical Challenges Solved

**The 404/429 Quota Battle:** We successfully navigated the deprecation of Gemini 1.5 and the quota limits of Gemini 2.0 to land on **Gemini 2.5 Flash**, which provides the best balance of speed and availability for this project.