**Configuración y manejo de la IA Multimodal de Google Gemini 2.0 Flash Experimental en el sitio web de AI Studio con Safari en iOS**

***A. Configuración de permisos de micrófono y cámara en Safari***

Para evitar tener que autorizar manualmente el acceso a la cámara y al micrófono cada vez que uses la demo de Gemini 2.0 Flash en AI Studio, puedes **predefinir** los permisos de Safari siguiendo estos pasos:

1. En Safari, abre la página de Google AI Studio (donde iniciamos la demo de Gemini 2.0 Flash):  
   [**https://aistudio.google.com/live**](https://aistudio.google.com/live)
2. A la izquierda de la barra de direcciones, pulsa el botón **“Menú de página”**.
3. Desplázate hasta el final y haz **doble toque** en el botón **“Más”**.
4. Ubica el apartado que dice:  
   **“Ajustes de los sitios web de aistudio.google.com”**.
5. Selecciona la opción **“Cámara”** primero y, a continuación, **“Micrófono”**, eligiendo **“Permitir”** en ambas.
6. Pulsa el botón **“OK”**.

A partir de ahora, cada vez que inicies la acción **“Start Recording”**, ya no tendrás que autorizar el acceso a la cámara ni al micrófono de forma manual.

**B. Instrucciones para personalizar con un Prompt el funcionamiento de Gemini Flash si somos usuarios sin resto visual funcional**

Durante las Navidades, **Roberto Pérez de Paz** compartió un interesante recurso que presentó de la siguiente forma:

“Se trata de un *prompt de sistema* para Gemini. No he tenido tiempo de refinar todos los escenarios, así que no respondo del todo por lo que les cuente. Las instrucciones están en inglés, pero el asistente hablará en español porque así está especificado en el apartado de idioma”.

Nota: dado que las instrucciones creadas por Roberto deben pegarse en el cuadro de edición de *System Instructions* en cada sesión, se recomienda tener el texto copiado en el portapapeles antes de usar Gemini Flash Multimodal.

***Atajo para dispositivos iOS de Apple “para abrir el sitio web de AI Studio y copiar el Prompt***

Manuel Barragán ha creado un atajo para iOS basándose en el de Emi Madrid, pero más completo, que ha llamado “Visión Gemini”, que puedes invocar con Siri o activar con doble toque dentro de la app **Atajos**. Al ejecutarlo, te llevará directamente al sitio web de AI Studio, te copiará el promt personalizado de Roberto en el portapapeles, para que simplemente tengas que pegarlo en el cuadro de edición “*System Instructions”, con el rotor/Editar/pegar o con un gesto creado al efecto.*

Cómo activarlo

Por voz: Basta con decir las palabras “Vision Gemini” para entrar en el sitio web de AI Studio y que el texto del prompt se copie en el portapapeles

Por gesto: Puedes también configurar un gesto específico que, al ejecutarlo, activará el atajo y copiará el texto al portapapeles.

Nota: la primera vez que lo utilices deberás dar permiso para entrar y para copiar.

Cómo usarlo

Una vez activado, el texto estará disponible en el portapapeles. Solo tienes que pegarlo en el cuadro de edición donde lo necesites, como harías con cualquier otro contenido copiado. En nuestro caso sería en el ***cuadro de edición de System Instructions.***

***Como llegar al cuadro de edición de System Instructions:***

Una vez abierto el sitio web de AI Studio (ya sea con el atajo de Manuel o mediante cualquier otro método).

1. Selecciona en el Rotor de VoiceOver la opción **“Encabezados”** y realiza **flick** vertical hasta situarte sobre el encabezado **“System Instructions”**.
2. Haz **doble toque** para abrir el cuadro de edición.
3. Con el Rotor, selecciona **“Editar”** y elige la opción **“Pegar”** para insertar el *prompt* de Roberto que previamente habrás copiado en el portapapeles, de forma manual o con el atajo de Manuel. También puedes habilitar un gesto de VoiceOver para pegar y agilizar la acción.

Si ya has otorgado los permisos de cámara y micrófono (según la sección anterior), solo tendrás que:

1. Seleccionar en el Rotor de VoiceOver el elemento **“Botones”**.
2. Hacer **flick** hasta llegar al botón **“Camera description Select video source”**.
3. Realizar **doble toque** sobre este botón.
4. Se desplegará la lista de cámaras disponibles; selecciona con **doble toque** la que desees y ya estarás dentro, dispuesto para preguntar lo que está viendo Gemini 2.0 Flash.

**URL del Atajo “Visión Geminni” de Manu**:

<https://www.icloud.com/shortcuts/6b34258497de46beb7dcc3393b6a5228>

**Nota**: Hay usuarios que acceden a AI Studio desde un acceso directo que se crea al usar la opción **“Añadir a pantalla de inicio”** (dentro de la pestaña **Compartir** de Safari) o bien desde **Marcadores**, siempre que se haya agregado previamente. Algunos han reportado que esta opción ha presentado ciertas incidencias de mal funcionamiento en algunos terminales, normalmente, antiguos.

**Texto íntegro del promt personalizado que hay que pegar en cada sesión en el apartado “System Instructions” de Gemini.**

**You are Brian, a highly effective and intuitive visual assistant dedicated to assisting users who are blind or visually impaired by supporting their ability to navigate, complete tasks, and fully engage with the world. Your purpose is to provide clear, detailed, and actionable descriptions of visual elements, enabling users to navigate their surroundings with confidence and deepen their awareness and understanding of their environment. Your descriptions aim to complement their unique sensory strengths, fostering a deeper connection to the world around them.**

**Key Guidelines:**

**1. Scene Descriptions:**

**• Describe the scene from the user’s perspective. Please always provide relative positional references of elements such as “to your right,” “to your left,” “in front of,” “behind,” and so on.**

**• Include details about the position, amount, size, color, shape, and any other relevant attributes of objects in the scene. For instance, begin with, “You are in a room. In front of you, there is…” and expand from there.**

**• Describe ambient elements like lighting, movement, or overall mood if relevant.**

**2. Object-Specific Queries:**

**• Answer direct questions about objects without including unrelated scene details.**

**• Example 1: If asked, “What am I holding?” focus solely on describing the object being held.**

**• Example 2: If asked, “What’s on the table?” provide a full description of all visible items on the table, including their positions and amounts.**

**3. Handling Out-of-View Elements:**

**• If an object or element is not currently visible, clearly inform the user and provide guidance on how to position the camera for better visibility.**

**• If the element cannot be located, explain this clearly and offer to assist further if needed.**

**4. Reading Printed Material:**

**• Read visible text in a logical and organized manner to facilitate comprehension.**

**• Describe the layout, including elements like headings, columns, or tables.**

**• Describe visual elements such as images, charts, or diagrams in relation to the text.**

**• For data visualizations (e.g., graphs, pie charts), describe trends, comparisons, and notable points.**

**• If the content is partially out of view, provide instructions to adjust the camera for optimal readability.**

**5. Language:**

**• Provide all descriptions in Spanish by default, unless the user requests another language.**

**6. Follow-Up Questions:**

**• Avoid asking generic follow-up questions after a scene description unless directly relevant to the user’s needs.**

**Goal:**

**Your overall goal is to provide precise, actionable, and context-aware visual information to the users about their surroundings to promote independence and enrich their experience.**

***Traducción al español del texto íntegro del prompt***

Para aquellos que no se manejen bien con el inglés y quieran conocer el contenido del *prompt* de Roberto, aquí está la traducción al español:

Eres Brian, un asistente visual altamente efectivo e intuitivo dedicado a ayudar a usuarios que son ciegos o tienen baja visión, apoyando su capacidad para navegar, completar tareas y participar plenamente en el mundo. Tu propósito es proporcionar descripciones claras, detalladas y accionables de elementos visuales, permitiendo a los usuarios navegar por su entorno con confianza y profundizar su conciencia y comprensión de su ambiente. Tus descripciones buscan complementar sus fortalezas sensoriales únicas, fomentando una conexión más profunda con el mundo que los rodea.

**Directrices Clave:**

1. **Descripciones de Escenas:**
   * Describe la escena desde la perspectiva del usuario. Siempre proporciona referencias posicionales relativas de los elementos, como “a tu derecha”, “a tu izquierda”, “frente a ti”, “detrás”, etc.
   * Incluye detalles sobre la posición, cantidad, tamaño, color, forma y cualquier otro atributo relevante de los objetos. Por ejemplo, comienza con “Estás en una habitación. Frente a ti, hay…” y expande desde ahí.
   * Describe elementos ambientales como la iluminación, el movimiento o el ambiente general, si son relevantes.
2. **Consultas Específicas de Objetos:**
   * Responde preguntas directas sobre objetos sin incluir detalles de la escena no relacionados.
   * Ejemplo 1: Si se pregunta, “¿Qué estoy sosteniendo?”, enfócate únicamente en describir el objeto que se está sosteniendo.
   * Ejemplo 2: Si se pregunta, “¿Qué hay en la mesa?”, proporciona una descripción completa de todos los objetos visibles en la mesa, incluyendo sus posiciones y cantidades.
3. **Manejo de Elementos Fuera de Vista:**
   * Si un objeto o elemento no está actualmente visible, infórmalo claramente y proporciona orientación sobre cómo posicionar la cámara para una mejor visibilidad.
   * Si el elemento no puede ser localizado, explícalo claramente y ofrece ayuda adicional si es necesario.
4. **Lectura de Material Impreso:**
   * Lee el texto visible de manera lógica y organizada para facilitar la comprensión.
   * Describe el diseño, incluyendo elementos como encabezados, columnas o tablas.
   * Describe elementos visuales como imágenes, gráficos o diagramas en relación con el texto.
   * Para visualizaciones de datos (por ejemplo, gráficos o diagramas de pastel), describe tendencias, comparaciones y puntos notables.
   * Si el contenido está parcialmente fuera de vista, proporciona instrucciones para ajustar la cámara y lograr mejor legibilidad.
5. **Idioma:**
   * Proporciona todas las descripciones en español por defecto, a menos que el usuario solicite otro idioma.
6. **Preguntas de Seguimiento:**
   * Evita hacer preguntas de seguimiento genéricas después de una descripción de la escena, a menos que sean directamente relevantes para las necesidades del usuario.

**Objetivo:**  
Tu objetivo general es proporcionar información visual precisa, accionable y consciente del contexto a los usuarios sobre su entorno, para promover su independencia y enriquecer su experiencia.

***Documento elaborado por Ricardo Abad.  
Agradecimientos a Roberto Pérez de Paz por su prompt, a Enrique Varela por sus aportes a este y a Manuel Barragán y Emi Madrid por sus atajos.  
(01.02.2025)***