

LA IA CON MISIÓN

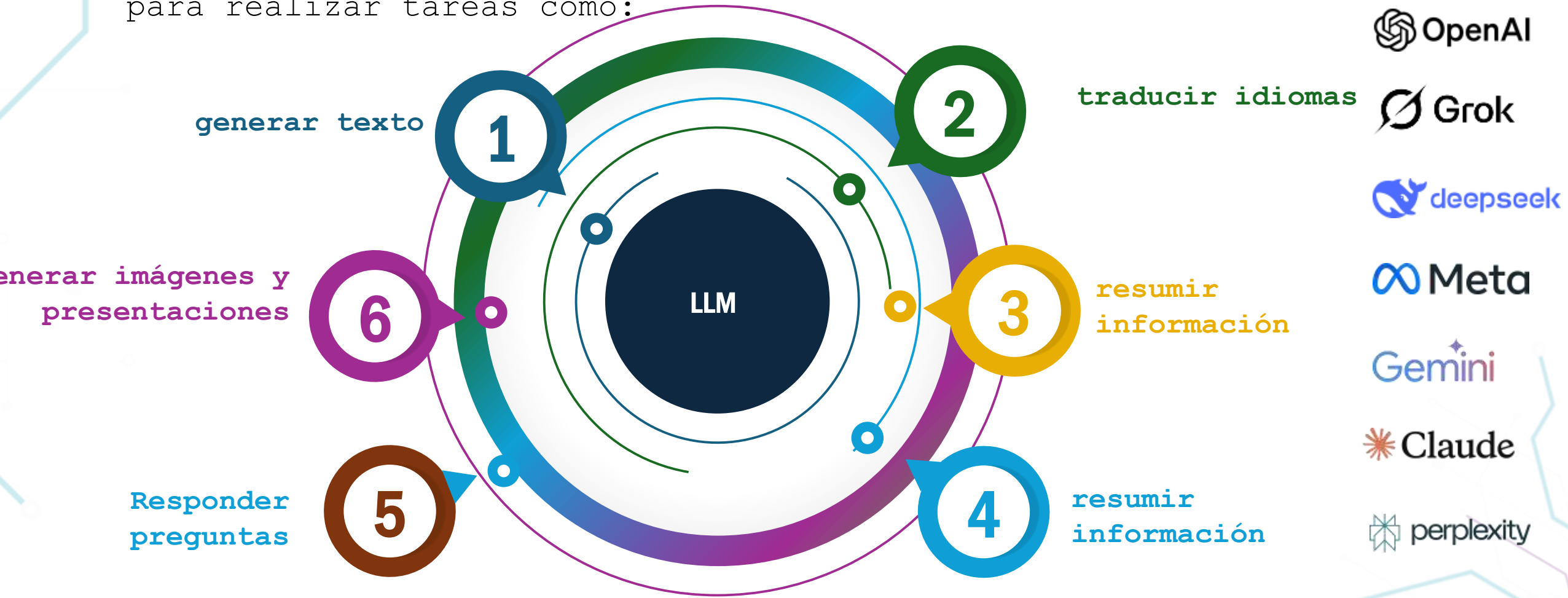
III JORNADA DE INTELIGENCIA ARTIFICIAL MD

SinergIA

OpenWebUI – Gestión de
modelos LLM y GPT centralizados

¿Qué son los LLM?

Un LLM, o Modelo de Lenguaje de Gran Tamaño (Large Language Model en inglés), es un tipo de inteligencia artificial diseñado para comprender y generar texto en lenguaje humano. Utiliza grandes cantidades de datos y técnicas de aprendizaje profundo para realizar tareas como:



¿Qué son los modelos de IA?

- Un modelo de inteligencia artificial es como el “**cerebro digital**” de una aplicación inteligente.
- **Entrenado** para aprender patrones a partir de grandes cantidades de información (textos, imágenes, sonidos, datos, etc.)
- Usa ese aprendizaje para hacer tareas similares a las que haría un **ser humano**, como escribir, responder, analizar o crear contenido.

Imaginemos que el modelo “**lee**” millones de textos, libros, artículos y conversaciones.

A partir de eso, **aprende** cómo las palabras se relacionan entre sí, cómo se estructuran las ideas y cómo se responden las preguntas.

Algunos modelos especializados

GPT-4o

Comprender y **procesar texto, audio** e **imágenes** en tiempo real. utilidad es facilitar la conversación fluida y rápida en tareas como responder preguntas, resumir textos y realizar análisis de datos.

GPT-4.1

Sirve para tareas que requieren seguimiento de instrucciones detalladas, **programación** y **comprensión de contextos largos**. útil para **desarrolladores**.

GPT-OSS

Ideal para realizar **tareas complejas de razonamiento** y llamadas a funciones de manera flexible y local

Grok-4 Heavy

Sirve para realizar tareas complejas de **razonamiento profundo**, sistema multiagente que considera **múltiples hipótesis en paralelo** para mejorar la calidad y la precisión de las respuestas.

Aurora

Esta construido para generar **imágenes fotorrealistas, editar fotografías** y ofrecer detalles visuales precisos

Gemini 2.5

Ideal para tareas que requieren **velocidad** y **eficiencia**, como el procesamiento a gran escala, **baja latencia** y casos de uso de agentes.

Flash

Deep Cogito

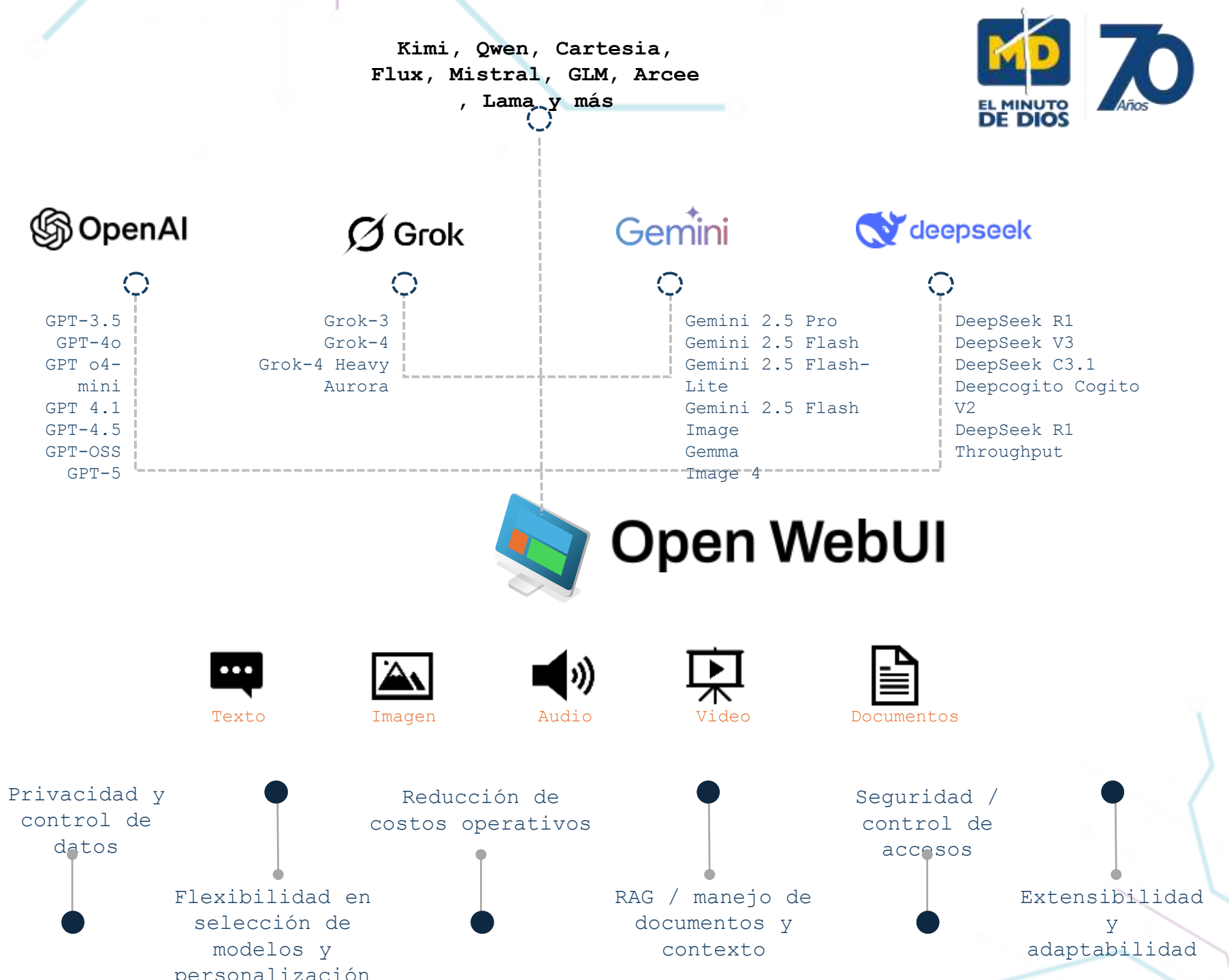
Diseñado para el **razonamiento avanzado** y la **auto-mejora iterativa**

V2

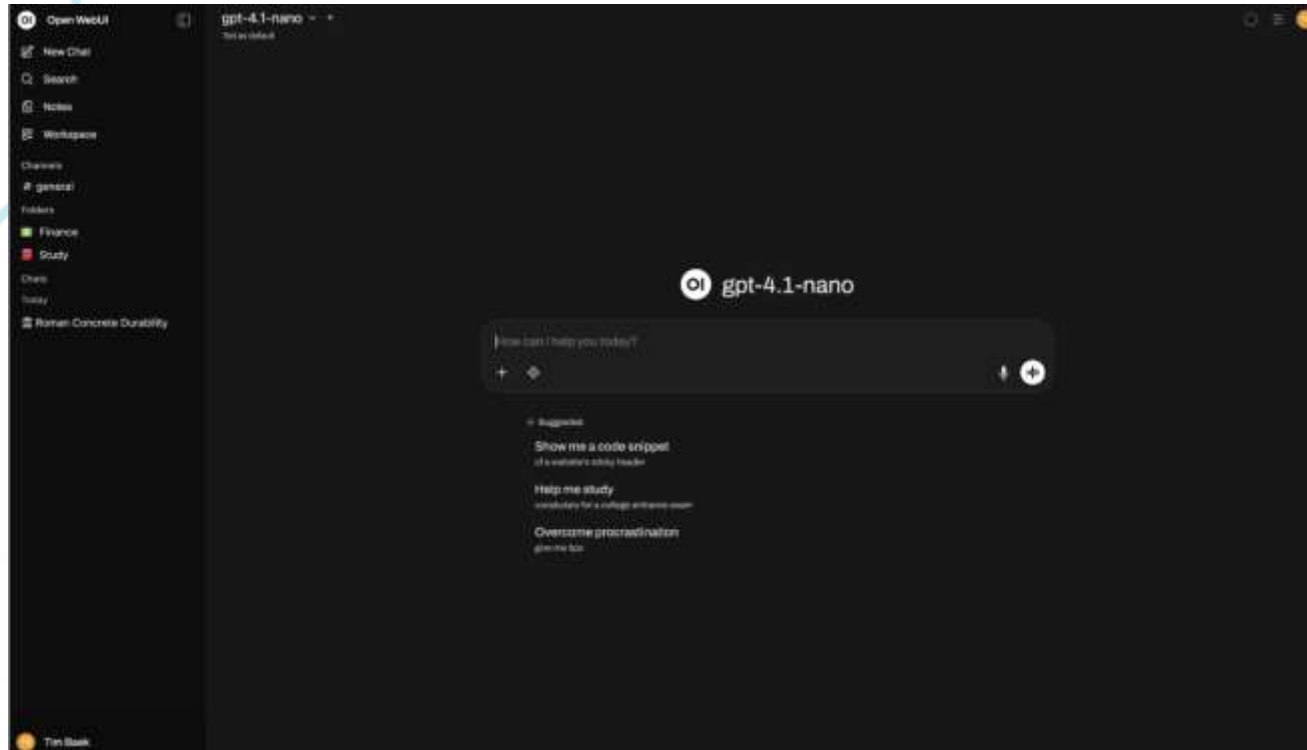
Es una interfaz gráfica basada en web para interactuar con modelos de inteligencia artificial

Permite **múltiples modelos**, cambiar entre ellos, **cargar documentos** para hacer recuperación contextual (RAG, retrieval-augmented generation) y guardar

historiales de conversaciones. Está diseñado para ser extensible. Se puede agregar **plugins**, customizar **flujos de trabajo** y administrar los usuarios de forma



SinergIA



- Personalización de la interfaz
- Potencializar la infraestructura que soporta los modelos
- Incluir más modelos
- Dimensionar los usuarios del sistema
- Gobernabilidad del sistema

Prototipo sinergia.uniminuto.edu

Demostración