

Gemini no tiene su origen en OpenAI. Son proyectos desarrollados por empresas competidoras con trayectorias tecnológicas distintas.

Orígenes y Desarrollo

- **Gemini (Google):** Es el resultado de la fusión de dos laboratorios de investigación de Google: **Google Brain** y **DeepMind**. Su arquitectura se basa en innovaciones propias de Google, principalmente la arquitectura **Pathways**.
- **GPT / ChatGPT (OpenAI):** Desarrollado por OpenAI (con fuerte inversión de Microsoft). Utiliza la arquitectura **Transformer**.

Relación Irónica

Aunque son entidades separadas, existe una conexión fundamental: **Google inventó la tecnología que OpenAI usó para el éxito.**

- En 2017, investigadores de Google publicaron el artículo "*Attention Is All You Need*", donde presentaron la arquitectura **Transformer** (la "T" en GPT).
- OpenAI tomó esta arquitectura, la escaló y lanzó ChatGPT, lo que obligó a Google a acelerar su propio desarrollo (pasando de proyectos internos como LaMDA y PaLM hasta llegar a Gemini).

Diferencias Clave

Característica	Gemini (Google)	GPT-4 (OpenAI)
Desarrollador	Google DeepMind	OpenAI
Arquitectura	Pathways / Multimodal nativo	Transformer (basado en diseño de Google)
Integración	Ecosistema Google (Android, Docs, Search)	Microsoft (Azure, Office, Bing)

Pathways es la infraestructura de IA de próxima generación de Google, diseñada para entrenar modelos que sean generalistas, eficientes y multimedios. A diferencia de los modelos tradicionales que se entrenan para una sola tarea (especialistas), Pathways busca imitar la flexibilidad del cerebro humano.

Pilares Tecnológicos

- **Entrenamiento Generalista:** Permite que un solo modelo aprenda miles de tareas diferentes (texto, imágenes, audio) de forma simultánea, transfiriendo conocimiento de una a otra.
- **Activación Dispar (Sparse Activation):** En lugar de activar toda la red neuronal para cada consulta, Pathways solo activa las rutas (vías) relevantes.
 - *Ejemplo:* Si preguntas por el clima, el modelo no activa sus "neuronas" de cálculo matemático complejo, ahorrando energía y tiempo.
- **Eficiencia en Hardware:** Distribuye el procesamiento de forma dinámica a través de miles de chips **TPU** (Tensor Processing Units) de Google, permitiendo modelos de una escala masiva (billones de parámetros) con menor latencia.

Comparativa: Tradicional vs. Pathways

Característica	Modelos Tradicionales	Google Pathways
Especialización	Un modelo por tarea	Un modelo para todo
Cómputo	Activación densa (100% del modelo)	Activación dispar (neuronas seleccionadas)
Aprendizaje	Desde cero para cada dominio	Aprendizaje transferido entre dominios
Escalabilidad	Limitada por la memoria de un clúster	Distribuida en miles de aceleradores

Pathways

Pathways es la infraestructura de IA de próxima generación de Google, diseñada para entrenar modelos que sean generalistas, eficientes y multimodales. A diferencia de los modelos tradicionales que se entrenan para una sola tarea (especialistas), Pathways busca imitar la flexibilidad del cerebro humano.

Pilares Tecnológicos

- **Entrenamiento Generalista:** Permite que un solo modelo aprenda miles de tareas diferentes (texto, imágenes, audio) de forma simultánea, transfiriendo conocimiento de una a otra.
- **Activación Dispar (Sparse Activation):** En lugar de activar toda la red neuronal para cada consulta, Pathways solo activa las rutas (vías) relevantes.
 - *Ejemplo:* Si preguntas por el clima, el modelo no activa sus "neuronas" de cálculo matemático complejo, ahorrando energía y tiempo.
- **Eficiencia en Hardware:** Distribuye el procesamiento de forma dinámica a través de miles de chips **TPU** (Tensor Processing Units) de Google, permitiendo modelos de una escala masiva (billones de parámetros) con menor latencia.

Comparativa: Tradicional vs. Pathways

Característica	Modelos Tradicionales	Google Pathways
Especialización	Un modelo por tarea	Un modelo para todo
Cómputo	Activación densa (100% del modelo)	Activación dispar (neuronas seleccionadas)
Aprendizaje	Desde cero para cada dominio	Aprendizaje transferido entre dominios
Escalabilidad	Limitada por la memoria de un clúster	Distribuida en miles de aceleradores

1. Modelos de Lenguaje (NLP)

- **BERT (2018):** Revolucionó el procesamiento de lenguaje natural al ser bidireccional (entiende el contexto izquierdo y derecho de una palabra). Es la base actual del Buscador de Google.
- **T5 (Text-to-Text Transfer Transformer):** Unifica todas las tareas de NLP (traducción, resumen, clasificación) en un formato de texto a texto.

- **LaMDA (2021):** Modelo especializado en diálogo fluido y natural. Fue el precursor de Bard.
- **PaLM / PaLM 2 (2022-2023):** "Pathways Language Model". Escalado masivo de parámetros (540 mil millones) usando la arquitectura **Pathways**. Es el motor que impulsó las primeras versiones de IA generativa de Google.

2. Modelos Multimodales y Especializados

- **Gemini (1.0, 1.5 Pro/Flash):** El modelo más actual. Nativo multimodal (entiende texto, imagen, video y audio desde su creación, no por módulos añadidos).
- **Med-PaLM 2 / Med-Gemini:** Versiones ajustadas específicamente para el dominio médico, capaces de aprobar exámenes de licenciatura médica y analizar radiografías.
- **Sec-PaLM:** Especializado en ciberseguridad para detección de malware y análisis de vulnerabilidades.

3. Modelos de DeepMind (Investigación Científica)

- **AlphaFold:** Modelo que predice la estructura 3D de las proteínas. Resolvió un problema biológico de 50 años.
- **AlphaGo / AlphaZero:** Famosos por derrotar a campeones mundiales en el juego de Go y ajedrez mediante aprendizaje por refuerzo.
- **AlphaCode:** Especializado en programación competitiva y generación de código complejo.

4. Generación de Medios

- **Imagen 2 / 3:** Modelos de generación de imágenes a partir de texto (competencia de DALL-E).
- **VideoPoet / Veo:** Modelos avanzados para la creación de video de alta definición a partir de prompts.
- **MusicLM:** Generación de música y audio de alta fidelidad.

Resumen de Evolución

Era	Modelo Clave	Enfoque
Pre-2017	RNN / CNN	Especialistas aislados
2017-2021	BERT / T5	Comprensión de lenguaje (Transformers)
2022-2023	PaLM / LaMDA	Razonamiento y escala (Pathways)
2024+	Gemini	Multimodalidad nativa y ventanas de contexto masivas

La "diáspora" de talento de Google ha sido incluso más masiva que la de OpenAI, debido a que Google (específicamente Google Brain y DeepMind) fue la cuna de la arquitectura Transformer.

Los "cerebros" de Google se diversificaron en las siguientes empresas líderes:

1. Los Autores de "Attention Is All You Need"

Los 8 autores del paper que inventó el **Transformer** (la base de todo ChatGPT y Gemini) fundaron sus propias compañías:

- **Character.ai:** Noam Shazeer (recientemente re-absorbido por Google).
- **Cohere:** Aidan Gomez (foco en IA para empresas).
- **Essential AI:** Ashish Vaswani y Niki Parmar.
- **Inworld AI:** Illia Polosukhin (también fundó el protocolo **NEAR**).
- **Sakura AI:** Llion Jones.

2. Competidores Directos (LLMs Genéricos)

- **Anthropic:** Aunque se asocia con la "mafia de OpenAI", muchos ingenieros clave de Google Brain se unieron para desarrollar **Claude**.
- **Mistral AI:** Arthur Mensch (ex-DeepMind) fundó en Francia esta empresa, que es el referente europeo en modelos abiertos.

3. IA Aplicada y Especializada

Empresa	Fundador / Talento de Google	Especialidad
Perplexity AI	Denis Yarats (ex-Google/Facebook)	Buscador conversacional
Glean	Arvind Jain (ex-Google VP)	Búsqueda inteligente para empresas
Adept AI	David Luan (ex-Google Brain)	Agentes que operan software
Mistral	Arthur Mensch (DeepMind)	Modelos eficientes y abiertos