Para la investigación

Por favor investiga en los siguientes puntos y realiza un tabla comparativa entre chatgpt copilo gemini:

1) Costos

Explicar la tarifa por tokens de entrada y de salida (precio por 1.000 tokens).

Ejemplo de cálculo por llamada y estimación mensual con N peticiones.

Breve mención de otros costos: embeddings (para RAG), fine-tuning, uso de herramientas.

Fórmula general:

costo = (tokens\_entrada \* precio\_in\_1k / 1000) + (tokens\_salida \* precio\_out\_1k / 1000)

2) Ventana de contexto

Definir qué es: límite total de tokens visibles por el modelo (entrada + salida).

Impacto práctico: truncamiento, degradación de calidad cerca del límite, estrategias (RAG, resúmenes, chunking).

Comparación: ejemplo de qué cabe en 4K vs 128K vs 200K tokens (reglas aproximadas: 700–800 palabras ≈ 1.000 tokens).

3) Diferencias entre modelos

Capacidades relevantes: razonamiento, idiomas, multimodalidad, velocidad, control de formato (JSON), herramientas.

Ventana de contexto y comportamiento cerca del límite.

Tabla comparativa simple.

Diseño de la pagina

Crear una página web informativa sobre la investigación que se realiza en cada uno de los chat generativos (chatgpt copilot Gemini) enfocada en los costos y la diferencia entre los modelos y ventana de contexto, tomando la información que se le suministra

costos:

Explicar la tarifa por tokens de entrada y de salida (precio por 1.000 tokens).

Ejemplo de cálculo por llamada y estimación mensual con N peticiones.

Breve mención de otros costos: embeddings (para RAG), fine-tuning, uso de herramientas.

**Ventana de contexto**

* Definir qué es: límite total de tokens visibles por el modelo (entrada + salida).
* Impacto práctico: truncamiento, degradación de calidad cerca del límite, estrategias (RAG, resúmenes, chunking).
* Comparación: ejemplo de qué cabe en 4K vs 128K vs 200K tokens (reglas aproximadas: 700–800 palabras ≈ 1.000 tokens).