SUPERATUS LÍMITES CON GENAI

CREA APPS SIN SABERLO TODO

EDUARDO PIVARAL

DATA ENGINEER / CLOUD ARCHITECT- PYTHIAN



EDUARDO PIVARAL

DATA ENGINEER / CLOUD ARCHITECT - PYTHIAN

20+ years of global IT experience, specializing in large-scale environments and diverse technology stacks, with a focus on Azure, AWS, and Snowflake.

Regular Author for **MSSQLTips.com**Regular Author for **SQLServerCentral.com**MSSQLTips.com Rookie of the year 2018
Local and International speaker





























"La diferencia entre saber y parecer que sabes... a veces son 4 prompts bien armados."

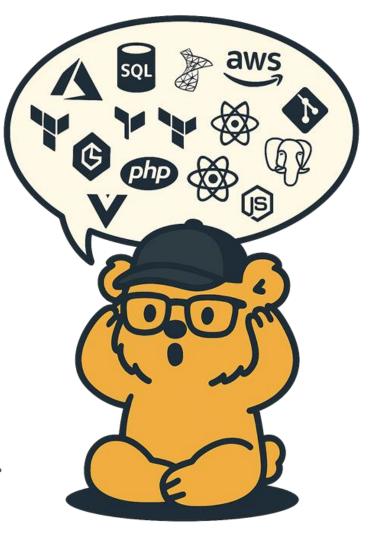


GenBear

INTRODUCCIÓN

- El desarrollo moderno es más complejo que nunca.
- Los desarrolladores deben dominar múltiples tecnologías.

- La IA generativa puede complementar nuestro conocimiento.
- •Objetivo: Crear aplicaciones funcionales y contenido relevante con ayuda de GenAl.



¿QUÉ ES GENAI?



Al que genera contenido original en diversas formas.

Puede crear texto, código, imágenes y más.

Ejemplos:

- Pedirle a ChatGPT que describa una arquitectura de microservicios.
- Generar código para una app de inventarios.
- Generar código SQL para crear una tabla basada en una descripción.
- Crear contenido de marketing para tu producto sin tener un copywriter.
- Generar imágenes para un catalogo de productos.
- Traducir documentos.

¿QUÉ ES UN AGENTE?

Después de recibir instrucciones, actúa por sí mismo.

Capaz de consultar APIs y usar herramientas externas.

Puede tomar decisiones simples para alcanzar un objetivo.

Ejemplo: Copilot, Claude, LangChain... Uno hecho por ti etc.



¿QUÉ NO PUEDE HACER? (TODAVÍA)



Tener iniciativa.

Descubrir cosas nuevas por si mismo.

Inferir intención del usuario con instrucciones vagas.

Comprender contexto de negocio completamente.

No puede tomar decisiones éticas o estratégicas.

RIESGOS Y CONSIDERACIONES

La IA puede generar respuestas incorrectas.

Código vulnerable o revelación de secretos.

Dependencia de la IA puede ser problemática.

Implementar una nueva funcionalidad puede dejar la app totalmente inutilizable.



AHORRA TIEMPO, PERO NO SIEMPRE AHORRA DINERO...



Planes GitHub Copilot

Plan	Precio	Qué incluye
Free	\$0 / mes	- 50 chats o agente- 2,000 completions- Acceso limitado a GPT-4.1, Claude 3.5 Sonnet
Pro	\$10 / mes	- Todo lo de Free +- Chats y completions ilimitados- GPT-4.11, Claude 3.7/4, Gemini 2.5 Pro- 6x más uso de modelos premium- Acceso a revisión de código y coding agent (preview)
Pro+	\$39 / mes	- Todo lo de Pro +- Acceso a todos los modelos (Claude Opus 4, o3, etc.)- 30x más uso de modelos premium

ESTIMACIÓN DE COSTOS PARA GENAI

Modelo	Entrada (1,000 tokens)	Salida (1,000 tokens)	¿Para qué sirve?
GPT-4.1	\$0.01	\$0.03	Tareas complejas, razonamiento, desarrollo serio
GPT-4.1 nano	\$0.0005	\$0.002	Económico para tareas ligeras y chatbots
GPT-4o	\$0.005	\$0.015	Multimodal (texto, imagen, audio) en tiempo real
DALL·E (imagen)	≈ \$0.04 por imagen	N/A	Generación de imágenes
Whisper (audio)	≈ \$0.006 por minuto	N/A	Transcripción precisa y rentable

COMO CALCULAR NUESTROS TOKENS

Tipo de contenido	Fórmula aproximada para tokens	En práctica, aproximadamente
Texto plano	1 token ≈ 4 caracteres (en inglés)1 token ≈ 0.75 palabras	100 palabras ≈ 75 tokens
Documentos (PDF/Word)	1 token ≈ 3-5 palabras (dependiendo del formato, tablas, etc.)	Un documento de 5 páginas ≈ 1,000–2,000 tokens
lmágenes	No se mide en tokens, pero en procesamiento Stable Diffusion: ≈ 1 imagen = 1 call (1–5k tokens equivalente de costo) DALL-E: $\approx $0.02-$0.20$ por imagen	El costo depende más de la generación que del texto del prompt
Audio (voz a texto)	1 minuto de audio = 150−200 tokens1 hora ≈ 10,000 tokens	Modelos como Whisper usan aprox. 1.5–2 tokens por palabra hablada
Video (texto extraído)	1 minuto hablado ≈ 150–200 tokens (si se transcribe)Video completo con análisis visual no está aún masivamente disponible	Transcripción de 10 minutos ≈ 2,000 tokens. Análisis visual puede requerir modelos multimodales
Prompt + respuesta total	Siempre debes considerar que tanto tu input como el output consumen tokens	Un input de 100 tokens + respuesta de 500 tokens = 600 tokens totales

EJEMPLO ESTIMACIÓN COSTOS

- Prompt largo de 500 palabras + respuesta de 1,000 palabras \approx 1,100 tokens \approx \$0.088 (GPT-4.1)
- Transcribir y traducir una entrevista de 30 minutos \approx 6,000 tokens de audio + 2,000 tokens para traducción \approx \$0.036 (Whisper) + \$0.16 (GPT-4.1) \approx \$0.20
- Generar 10 imágenes con DALL·E 3 (1024×1024, estándar) \approx \$0.40
- ■Un agente con múltiples pasos puede usar fácilmente 5,000—10,000 tokens por tarea \approx \$0.40—\$0.80 (GPT-4.1)



¿ES ESTO VIBE CODING?

Respuesta corta: NO!

- Usamos GenAl como copiloto, no como reemplazo del desarrollador.
- Tú sigues dirigiendo, tomando decisiones, entendiendo el código, reescribiéndolo si es necesario.
- Vibe Coding delega todo a la IA...receta para el desastre!
- la lA complementa tu conocimiento, no lo sustituye.
- Aprender fundamentos de software y patrones de diseño sigue siendo necesario.



"Quien no entiende lo básico, ni con lA llegara lejos..."



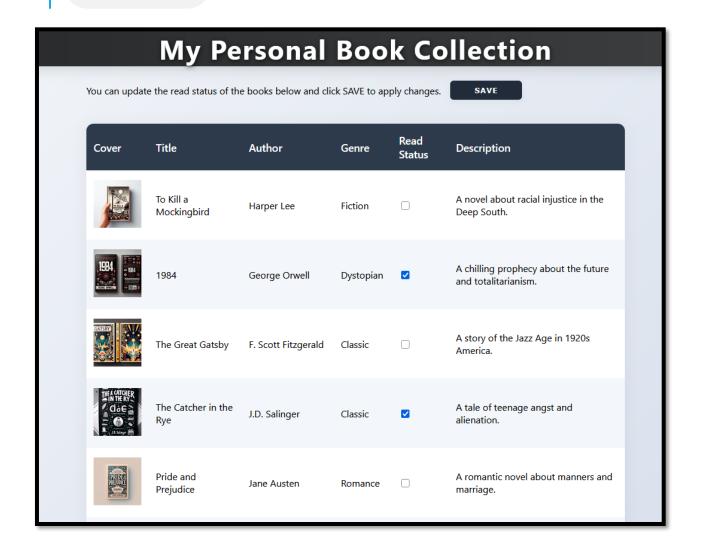
GenBear

DEMOS PUBLICADOS

Ejemplos de que hecho con GenAl Ustedes pueden lograr mucho mas...



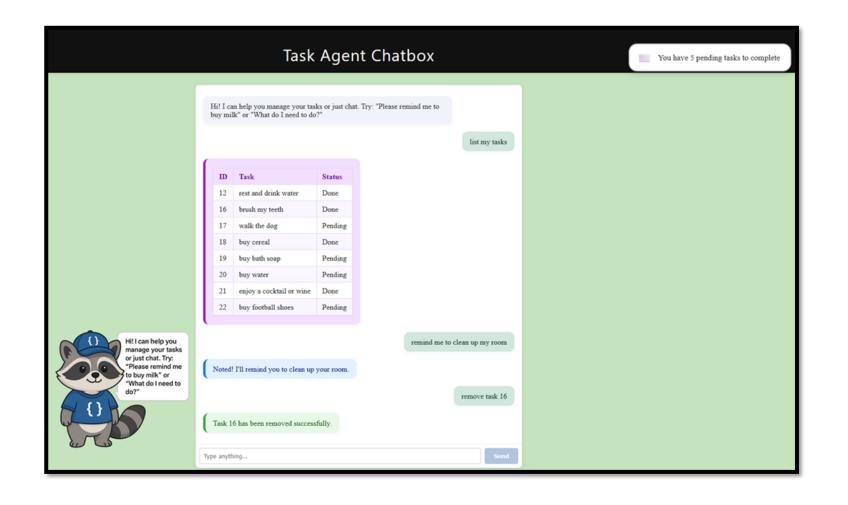
BUILT 100% WITH GENAI: NO-CODE FULL STACK APP





Aplicación web full-stack construida completamente con herramientas de GenAl

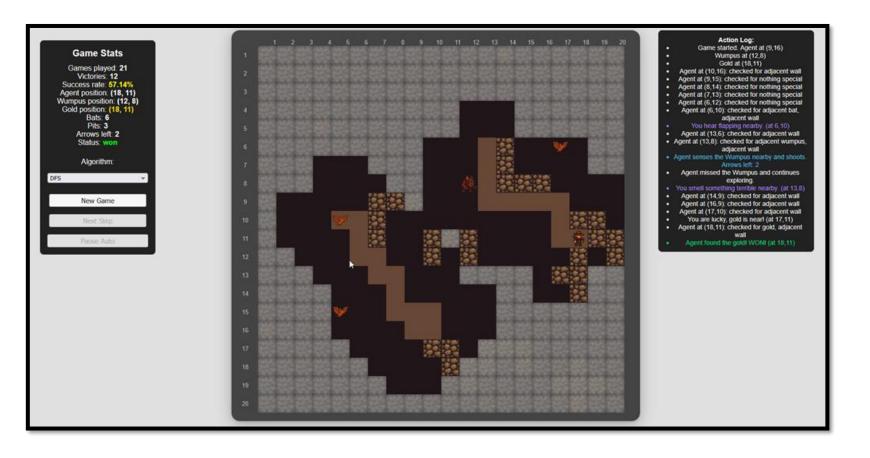
AI AGENT: TASK AGENT CHATBOX





Agente IA que mantiene tareas en una DB Azure SQL

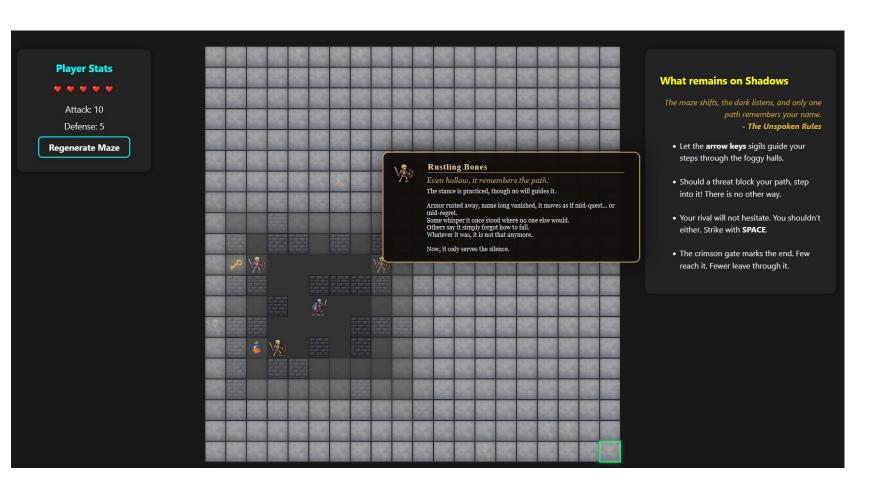
HUNT THE WUMPUS





Mi visión personal del juego, explorado con DFS y A*

WHAT REMAINS IN SHADOW





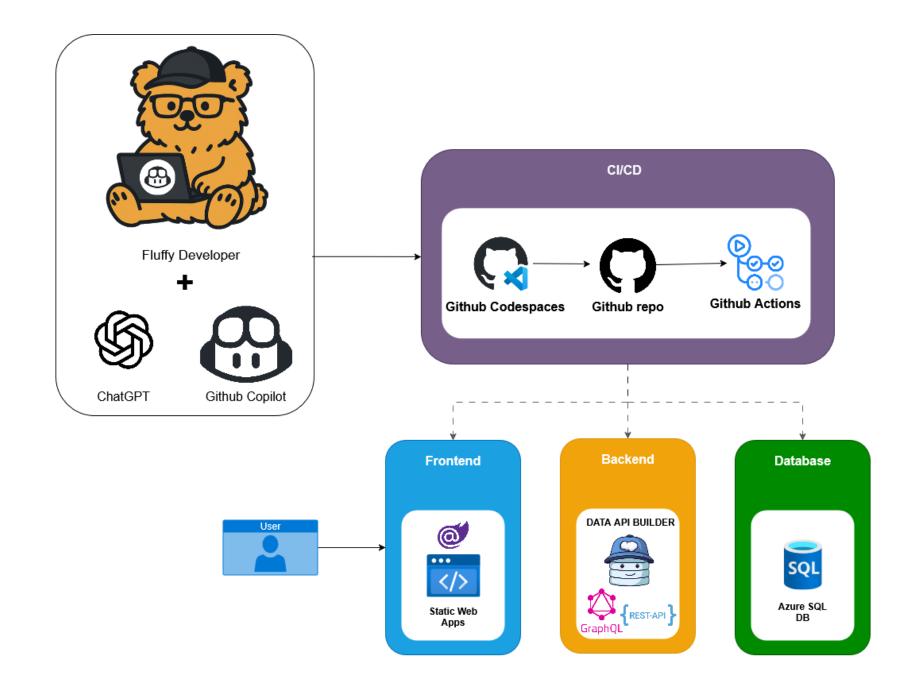
RPG Estilo Soulsborne con mecánicas sencillas, aun en desarrollo.

DEMO

Creemos nuestra app con GenAl



DEMO



MUCHAS GRACIAS

Presentación disponible aquí









