  **UNIVERSIDAD BANCARIA DE MÉXICO**

*“Constancia Unidad y trabajo”*

**INGENIERÍA EN SISTEMAS COMPUTACIONALES**

RECONOCIMIENTO DE VALIDEZ OFICIAL DE ESTUDIOS DE LA SECRETARÍA DE EDUCACIÓN PÚBLICA No. 2022241 DE FECHA 13 DE SEPTIEMBRE DE 2002.

NOMBRE DE LA MATERIA

Inteligencia Artificial 2

NOMBRE DEL PROFESOR(A)

Gabriel Temoatzin Tlatuani González

CUATRIMESTRE

11o

TÍTULO DEL TRABAJO O INVESTIGACIÓN

Proyecto Final

NOMBRE DE ALUMNO(S)

Amezcua Sagrero Sergio Daniel

Martínez Cruz Diego Alberto



Documentación de los modelos y herramientas que se usaron para la generación de imágenes, con un enfoque particular en las razones por las que no funcionaron en mi entorno local.

**1. Stable Diffusion**:

* Descripción:
  + Modelo de difusión latente de texto a imagen de código abierto.
  + Utiliza una red neuronal convolucional (CNN) para generar imágenes a partir de descripciones textuales.
  + Ofrece un alto grado de control sobre el proceso de generación a través de indicaciones y parámetros.
* Uso:
  + Principalmente a través de la biblioteca diffusers Hugging Face.
  + Existen también interfaces web y API que facilitan su uso.
* Requisitos:
  + Hardware:
    - GPU NVIDIA con al menos 8 GB de VRAM (recomendado).
    - CPU potente con al menos 16 GB de RAM.
    - Almacenamiento rápido (SSD).
  + Software:
    - Python 3.7+.
    - Bibliotecas: torch, diffusers, transformers, accelerate.
* ¿Por qué no funcionó?
  + Mi laptop tiene 4 GB de RAM, lo cual es muy por debajo de los requisitos mínimos.
  + La CPU integrada no es suficiente para procesar los cálculos complejos del modelo.
  + La falta de una GPU dedicada impide la aceleración del procesamiento.

**2. DALL-E 2/3 (DALL-E mini):**

* Descripción:
  + Modelos de generación de imágenes desarrollados por OpenAI.
  + Utilizan una combinación de modelos de difusión y transformadores para generar imágenes a partir de texto.
  + Dall-e 3, es la última versión y la que actualmente integra Bing Image Creator.
* Uso:
  + A través de la API de OpenAI (DALL-E 2/3).
  + Bing Image Creator, es una forma gratuita de uso.
* Requisitos:
  + Cuenta de OpenAI con acceso a la API (DALL-E 2/3).
  + Cuenta de Microsoft para Bing Image Creator.
  + Conexión a internet estable.
* Consideraciones:
  + Servicio propietario, controlado por OpenAI.
  + Bing Image Creator, es gratuito, pero con limitaciones de uso.
  + Más fácil de usar que Stable Diffusion, pero menos controlable.
* Por qué no funcionó (en el contexto de mi entorno local):
  + No se intentó ejecutar localmente, ya que es un servicio basado en la nube.

**3. SDXL Turbo:**

* Descripción:
  + Versión optimizada de SDXL para generación de imágenes en tiempo real.
  + Reduce significativamente los pasos de inferencia, lo que acelera la generación.
* Uso:
  + A través de la biblioteca diffusers Hugging Face.
* Requisitos:
  + GPU NVIDIA de gama alta con al menos 12 GB de VRAM (recomendado).
  + CPU potente y mucha RAM.
* ¿Por qué no funcionó?
  + A pesar de ser más ligero que SDXL, aún requiere un hardware potente.
  + Las limitaciones de RAM y CPU de mi laptop impidieron su funcionamiento.

**4**[**. Small Stable Diffusion**](https://www.reddit.com/r/StableDiffusion/comments/10exzcx/small_stable_diffusion_similar_image_generation/)**:**

* Descripción:
  + Modelo de difusión extremadamente ligero diseñado para dispositivos con recursos limitados.
  + Sacrifica calidad de imagen por eficiencia.
* Uso:
  + A través de la biblioteca diffusers.
* Requisitos:
  + CPU y RAM mínimos.
* ¿Por qué no funcionó?
  + Problemas para descargar el modelo desde Hugging Face Hub (errores de autenticación).

**5. Distilled Diffusion Models:**

* Descripción:
  + Versiones optimizadas de modelos de difusión grandes, creadas mediante "destilación".
  + Ofrecen un buen equilibrio entre calidad de imagen y eficiencia.
* Uso:
  + A través de la biblioteca diffusers.
* Requisitos:
  + Menos RAM que los modelos completos, pero aún se recomienda una GPU.
* Por qué no funcionó (en mi entorno local):
  + Aunque son más ligeros, seguirían teniendo problemas con la limitación de Hardware.

**6. Generación de Imágenes Basada en API:**

* Descripción:
  + Servicios que ofrecen API para generar imágenes a partir de texto.
  + Ejemplos: DALL-E 2/3, Midjourney, API de difusión estable.
* Uso:
  + A través de solicitudes HTTP a la API del servicio.
* Requisitos:
  + Cuenta en el servicio y, en muchos casos, pago por uso.
  + Conexión a internet estable.
* Por qué no funcionó (en el contexto de mi entorno local):
  + No se intentó ejecutar localmente, ya que son servicios basados ​​en la nube.

**7. Optimización local:**

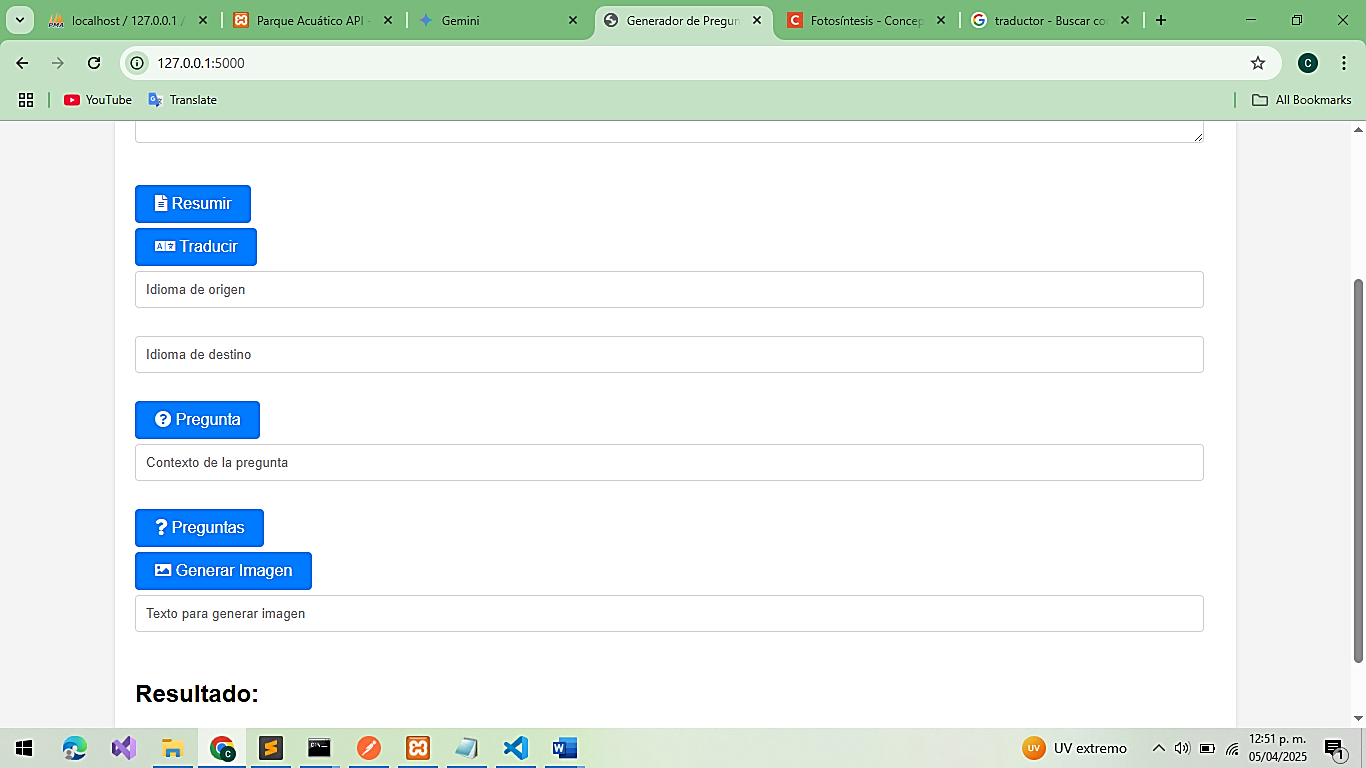
* Generación por partes:
  + Divide la generación de imágenes grandes en mosaicos más pequeños.
  + Reduce el consumo de RAM.
* Cuantización de modelos:
  + Reduzca la precisión de los pesos del modelo para disminuir el consumo de memoria.
* Por qué no funcionó (en tu entorno local):
  + La optimización local es compleja de implementar y requiere conocimientos técnicos avanzados.
  + Aun con la optimización, la limitación de Hardware, seguirá siendo un problema.

**8. Herramientas de Python:**

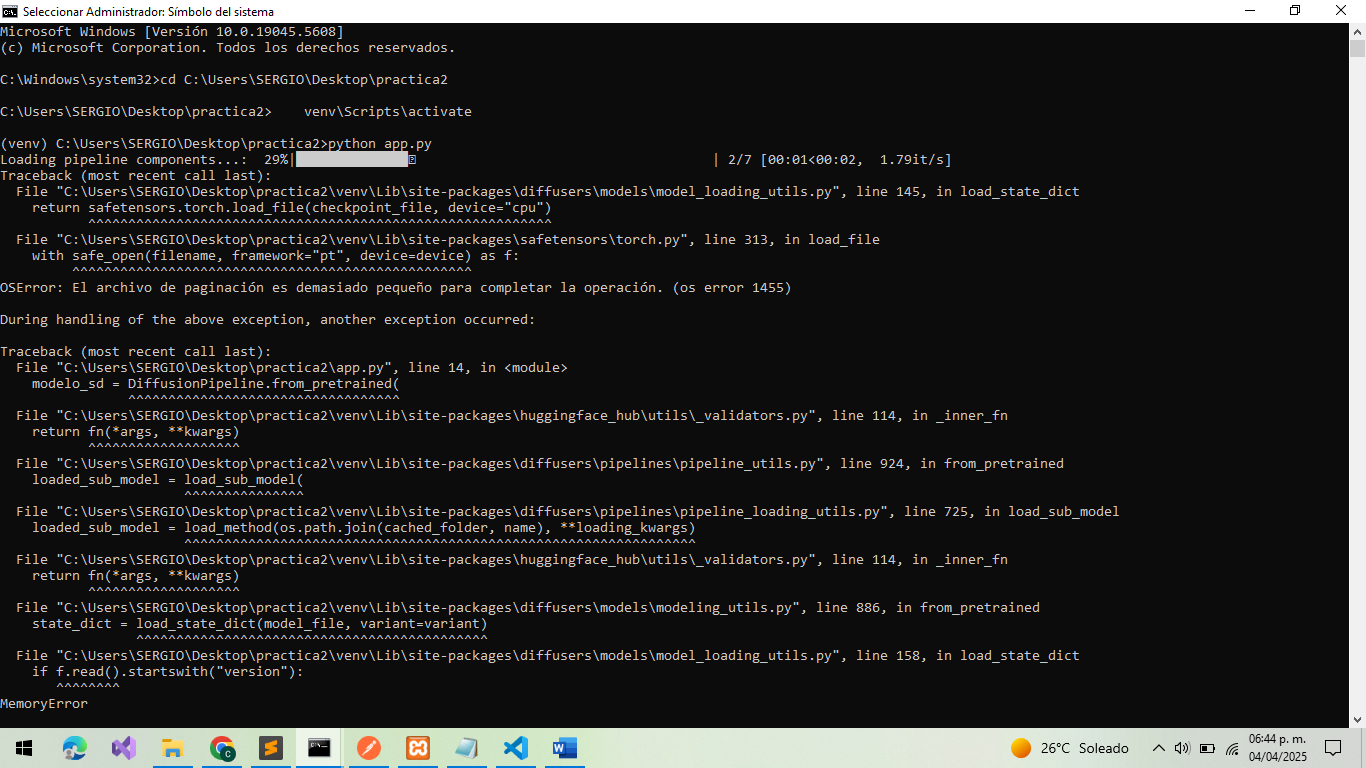
* diffusers:
  + Biblioteca para cargar y utilizar modelos de difusión.
* transformers:
  + Biblioteca para modelos de lenguaje (T5).
  + Procesamiento de texto para generar indicaciones.
* flask-ngrok:
  + Extensión de Flask para exponer servidores locales a través de ngrok.
  + Esencial para ejecutar Flask en Google Colab.
* torch:
  + Biblioteca PyTorch para aprendizaje profundo.
  + Manejo de tensores y operaciones de GPU.
* Por qué no funcionó (en mi entorno local):
  + Estas herramientas no son la causa del problema, sino las limitaciones de hardware de mi laptop.

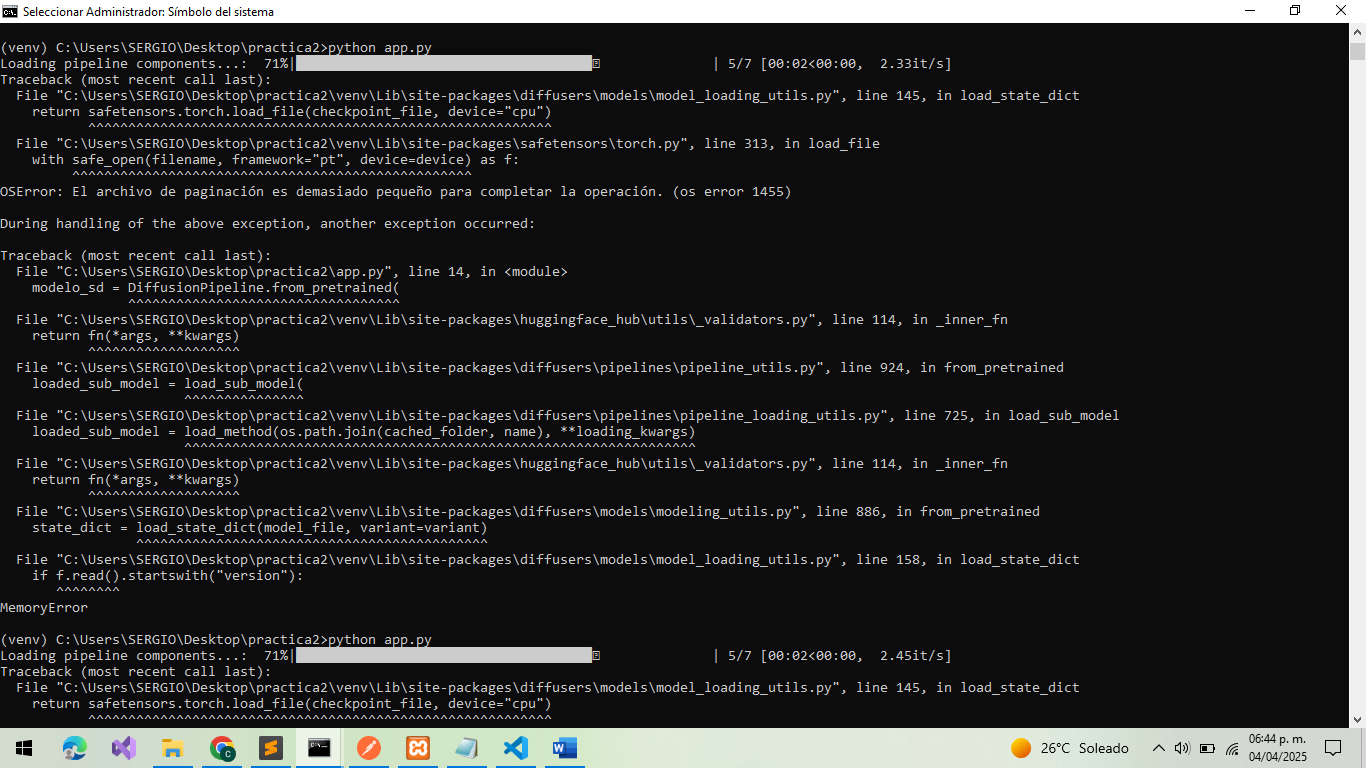
**Evidencias**

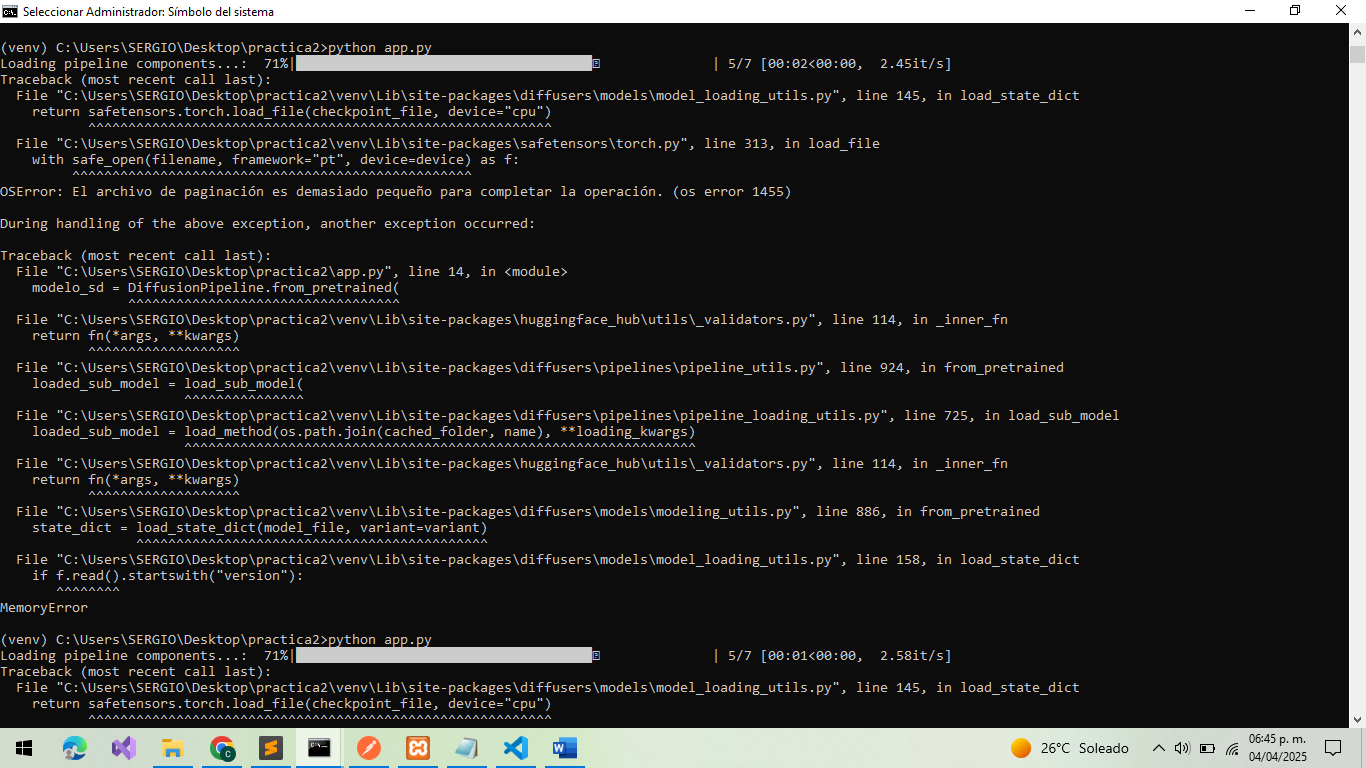
Prueba de ingresar el texto y generar imágenes…



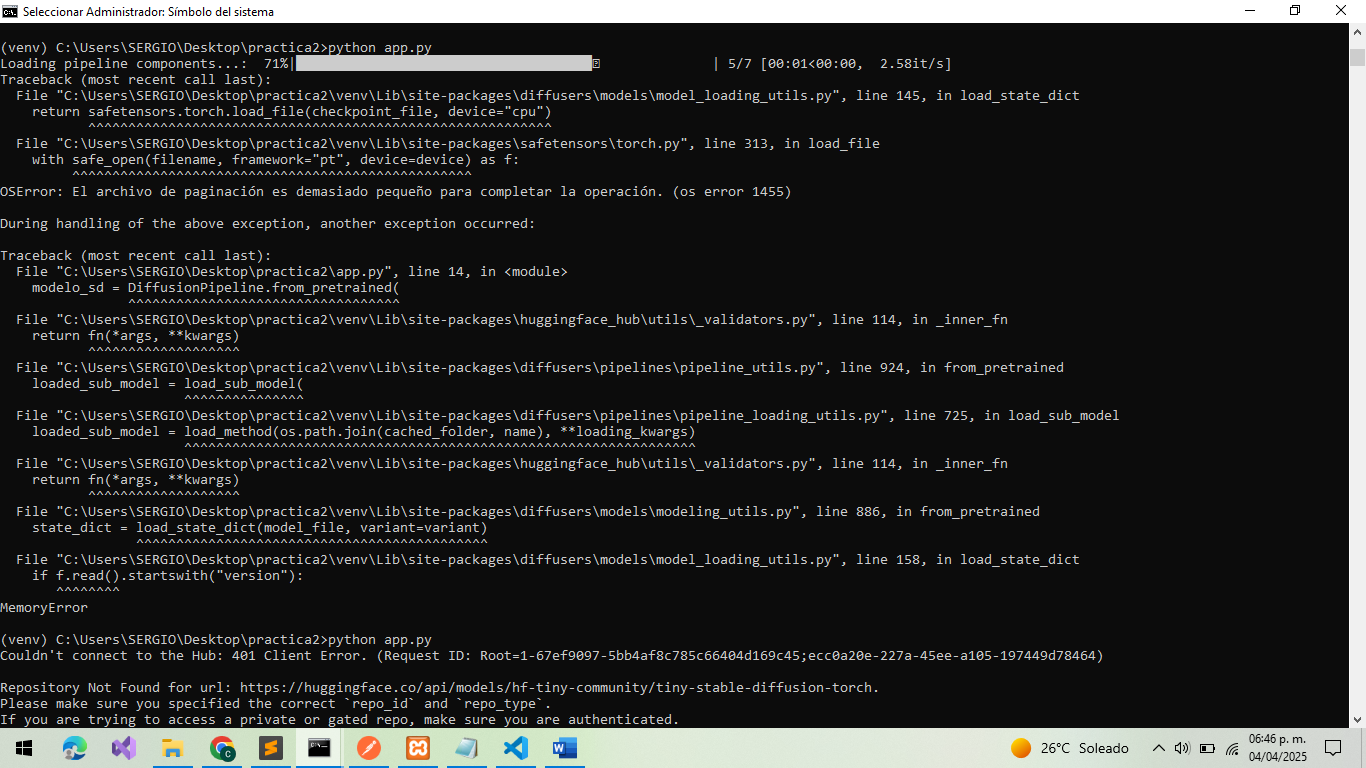
Pruebas de intentos con diferentes modelos:

Difucceess

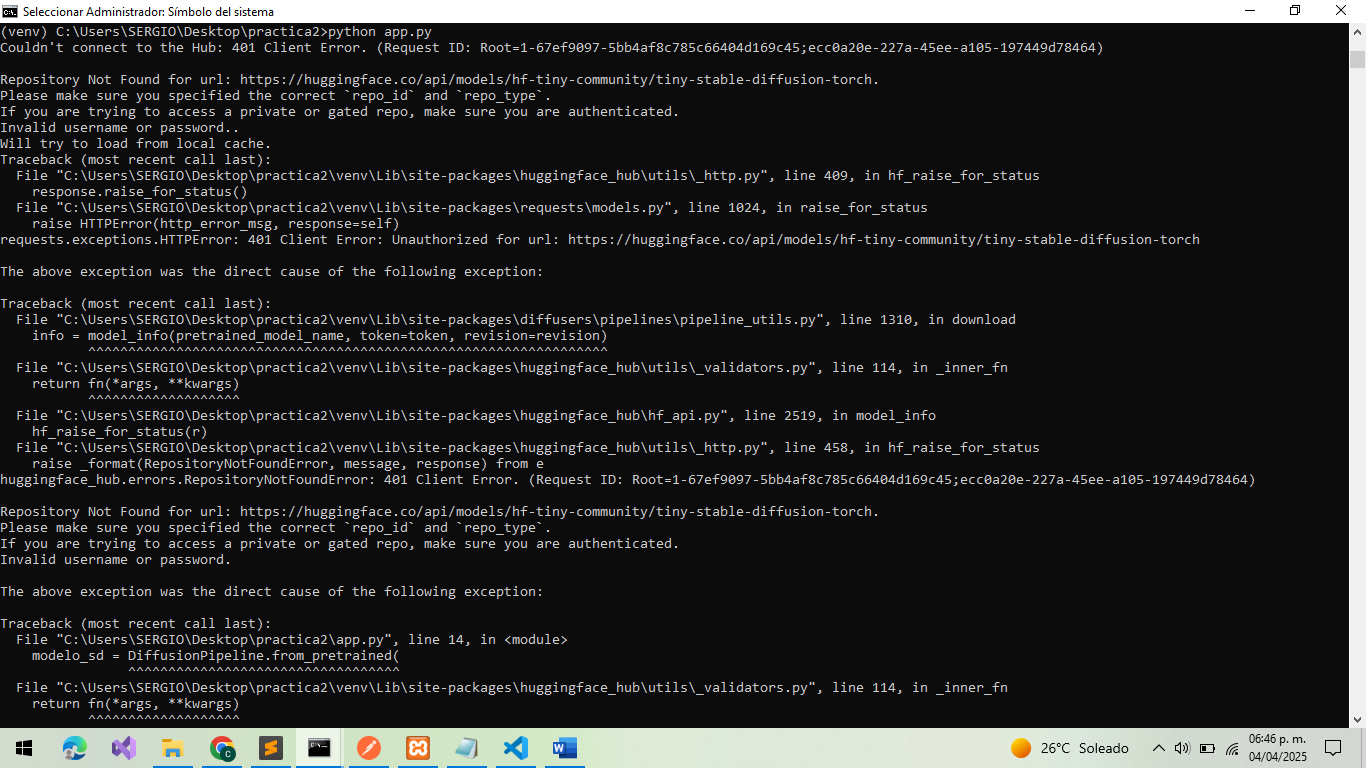




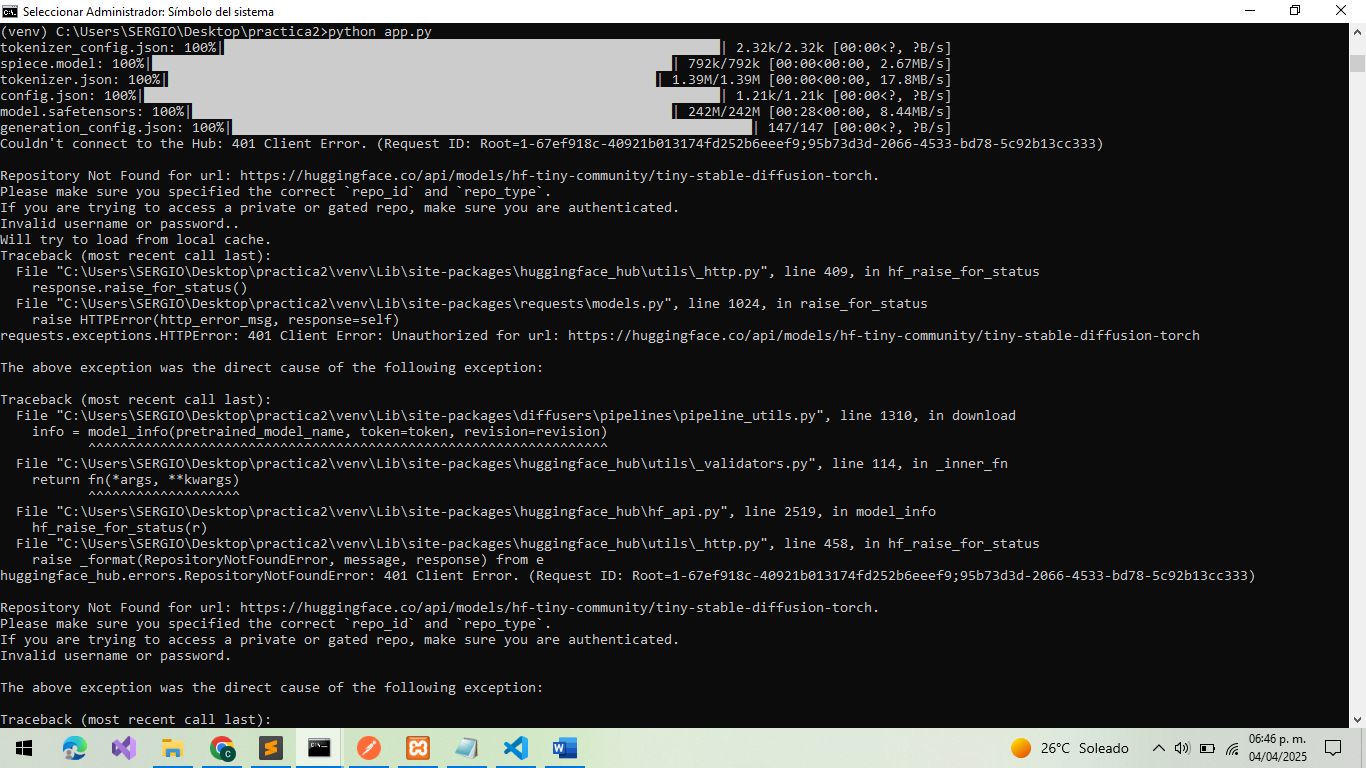
e-mini

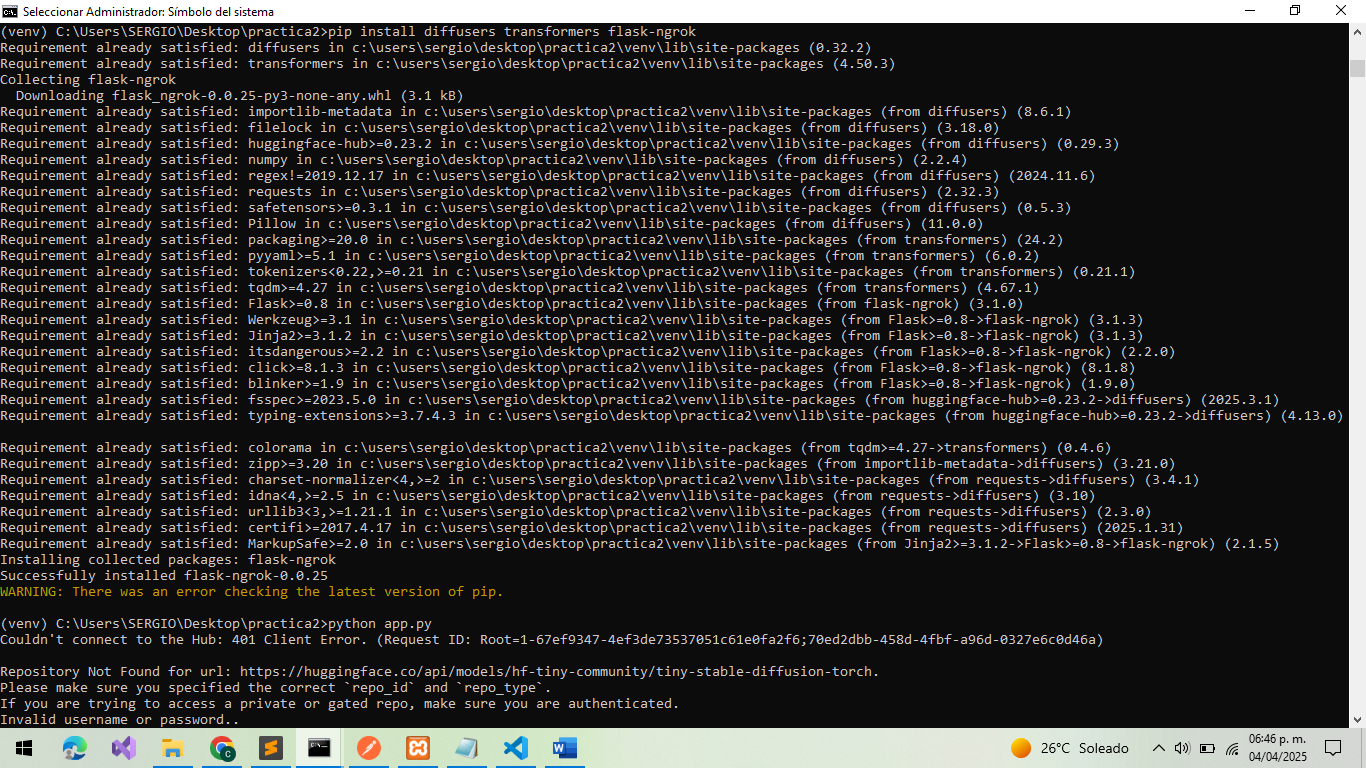


Intento de large

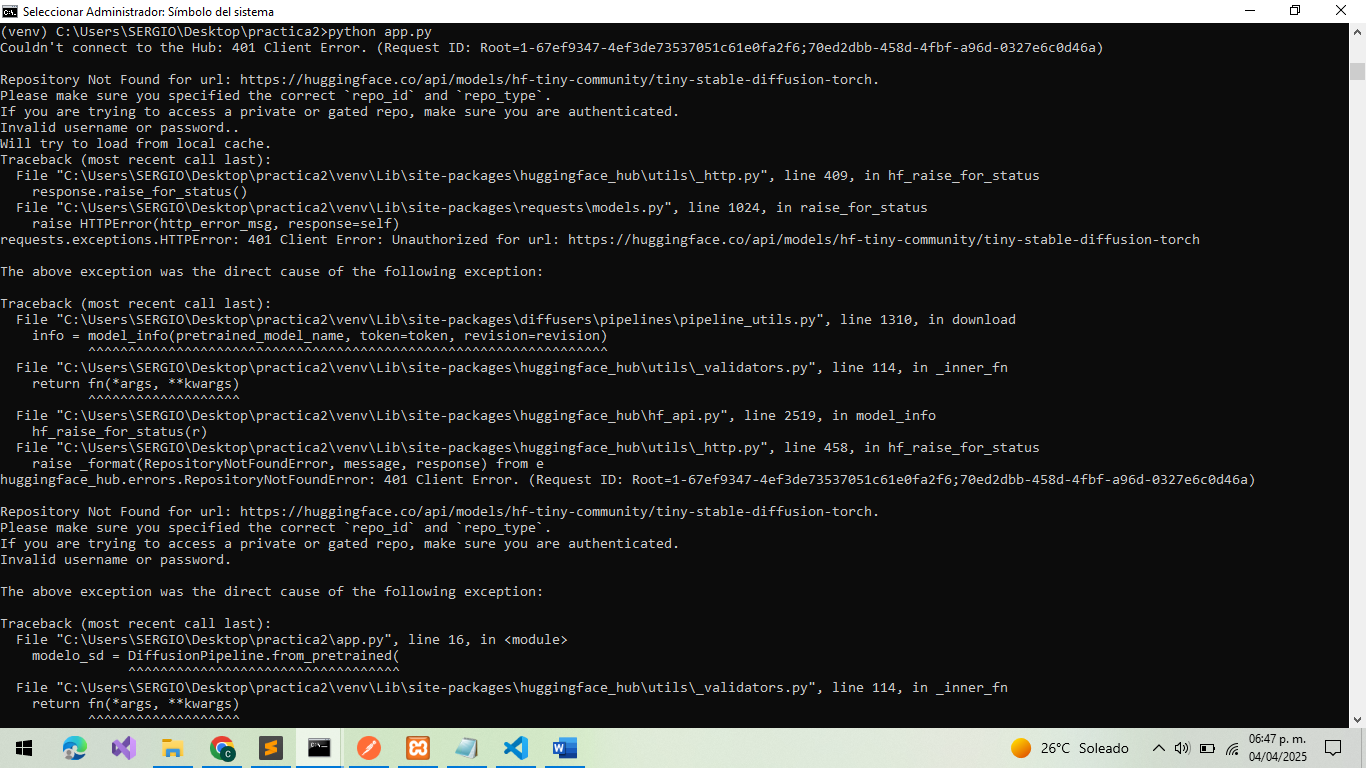


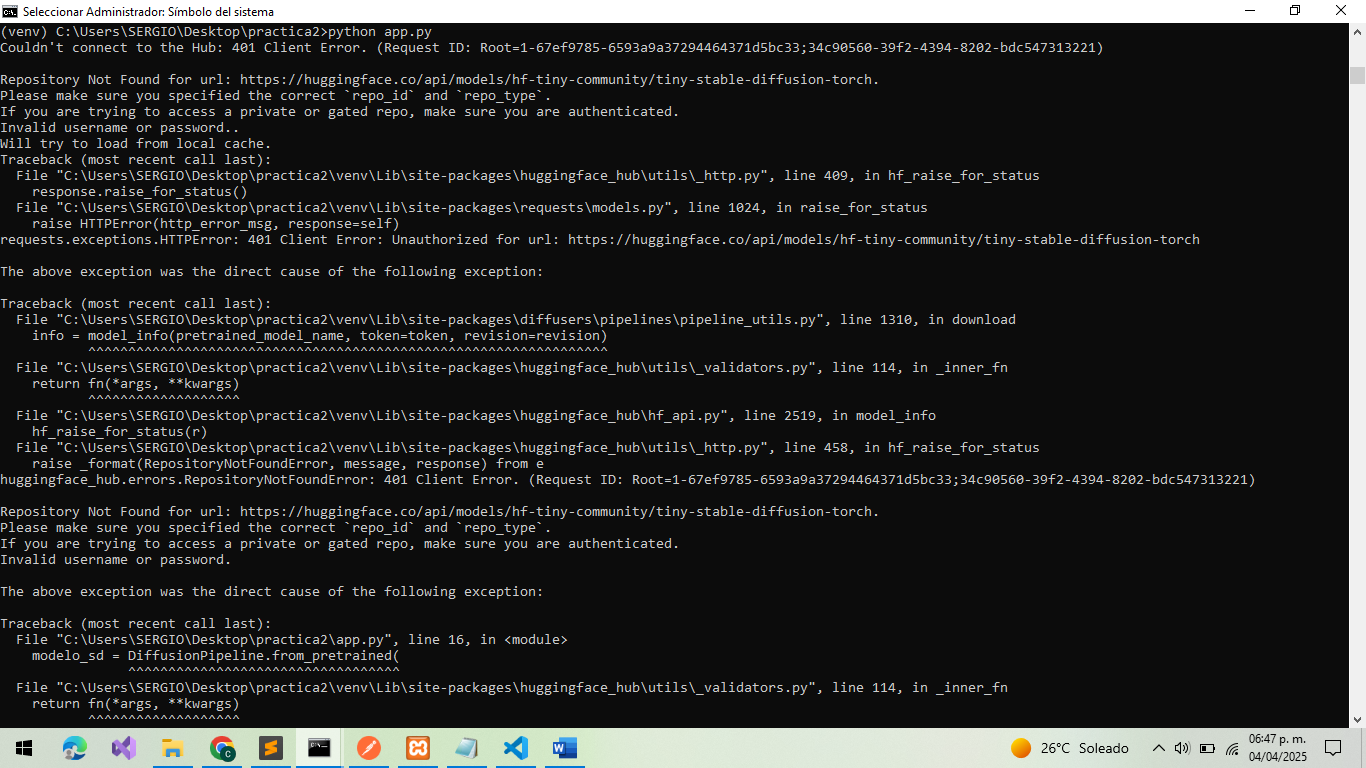
Intento de large





Google





Intento de large

