

Project presentation

Presented by Abdelrahman Khaled

Audio to Image

Problem Statement: There's a need for converting spoken Arabic into meaningful visuals for applications like storytelling, education, or multimedia projects.

Goal: Transform Arabic speech into English text and generate images from this text, providing a complete end-to-end solution.

Use Cases: Storytelling, educational tools, automated video generation



Ideation process

Audio Input

Record audio through the microphone of Gradio

Speech-to-Text

Convert Arabic audio into Arabic text.

Translation

Translate Arabic text into English text.

Text-to-Image

Generate an image based on the English text.

Audio-to-Text (Speech Recognition):

Model: OpenAI Whisper (Large version) for Arabic speech recognition
Frameworks: PyTorch, Gradio (for capturing audio)

Arabic-to-English Translation:

Model: Pretrained Transformers facebook/m2m100_418M

Text-to-Image Generation:

Model: Stable Diffusion / stabilityai/stable-diffusion-3-medium-diffusers diffusion models

Overall Tech Stack:

Python, PyTorch, Hugging Face Transformers, Gradio, and Diffusers library

Step 1: Speech-to-Text

Description:

Using the Whisper model to convert Arabic audio into text.

Workflow:

- Microphone input
- Model inference
- Output in Arabic script

Challenges:

mistakes correction, noisy backgrounds, and varying speech speeds.



Step 2: Arabic to English

Description:

Using translation models to convert the recognized Arabic text into English.

Workflow:

- Input: Arabic text
- Translation Model: Uses transformers
- Output: English text

Challenges:

Maintaining context and nuances between languages.



Step 3 – Text-to-Image

Description:

Generating images from the translated English text.

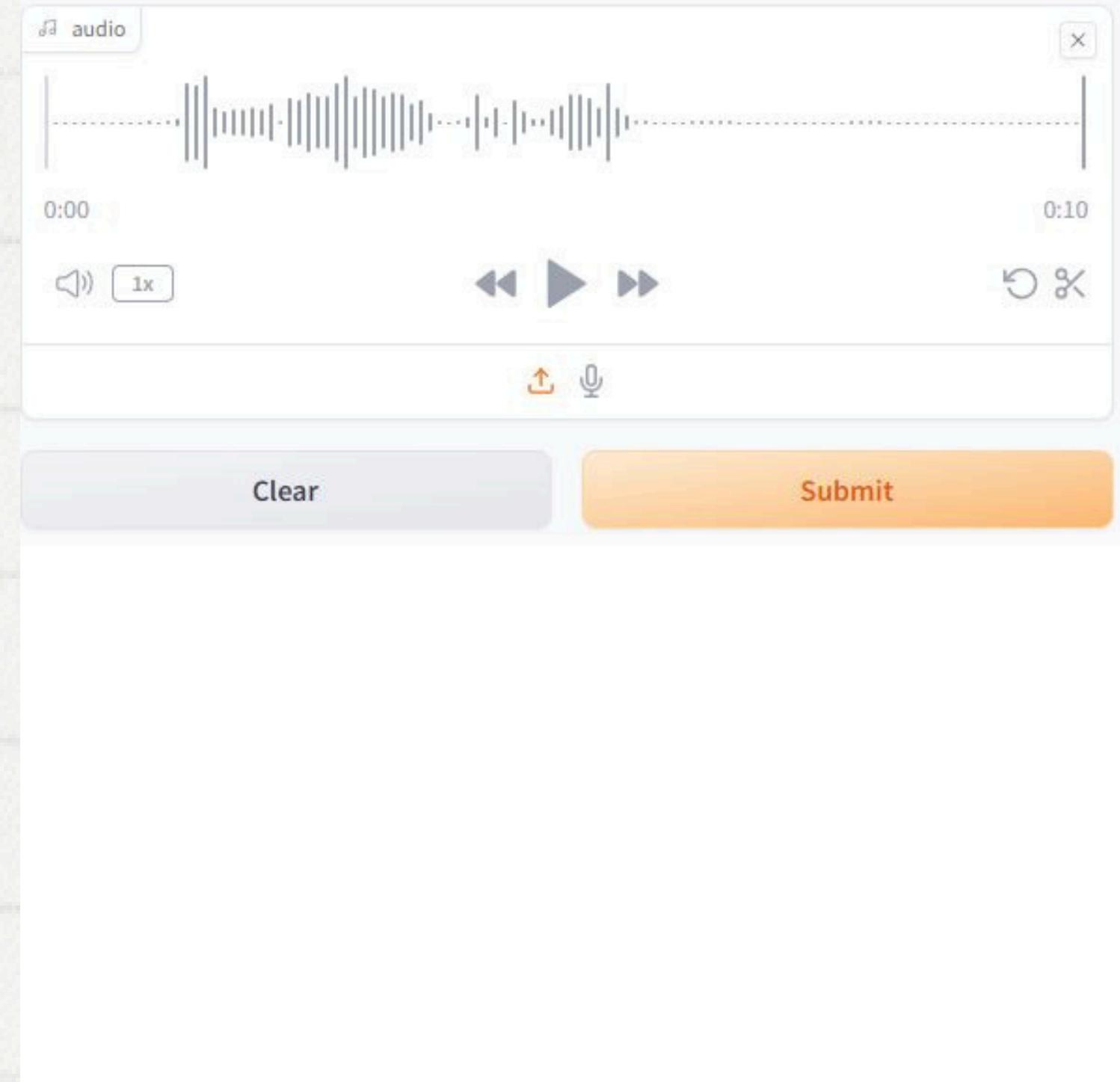
Workflow:

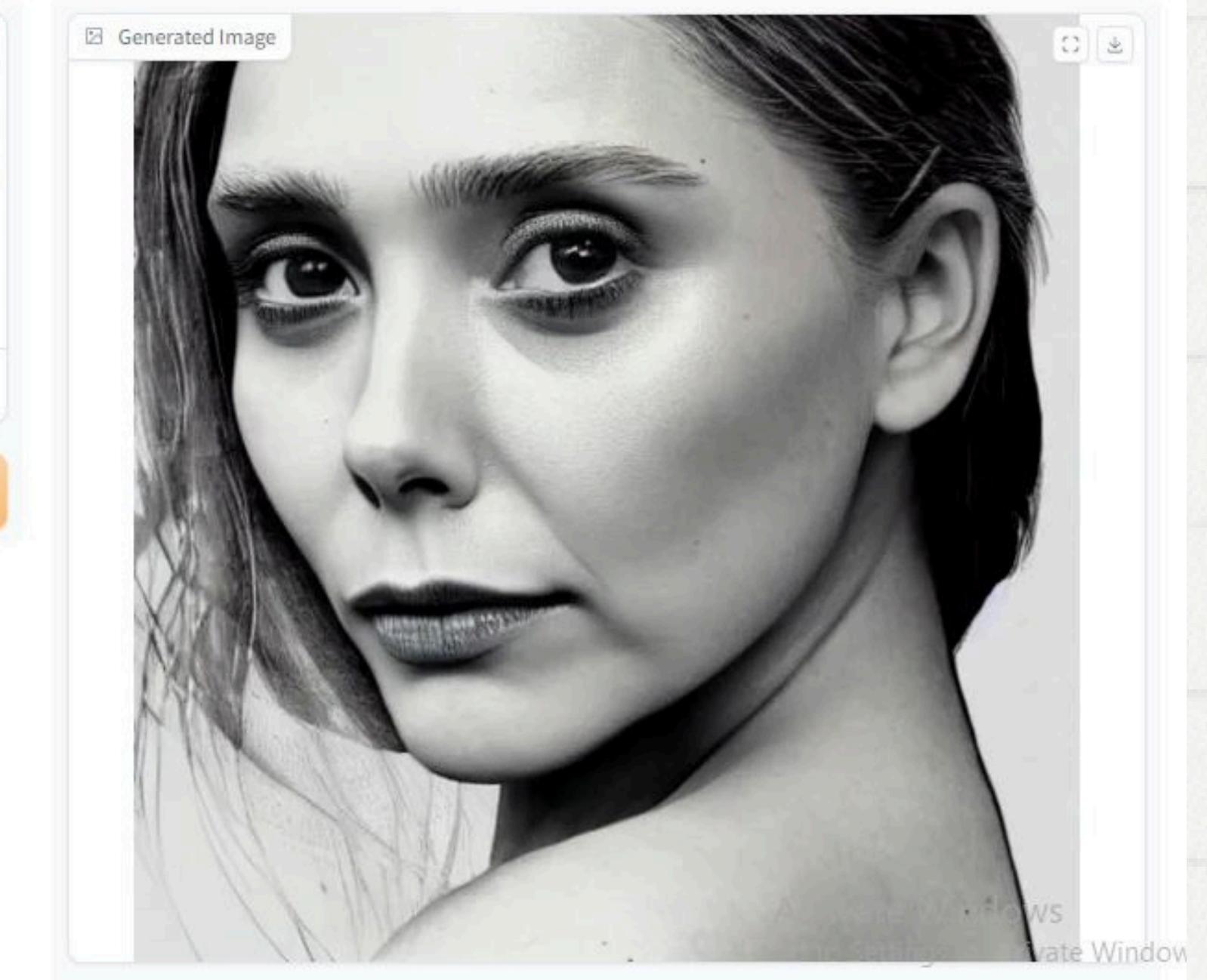
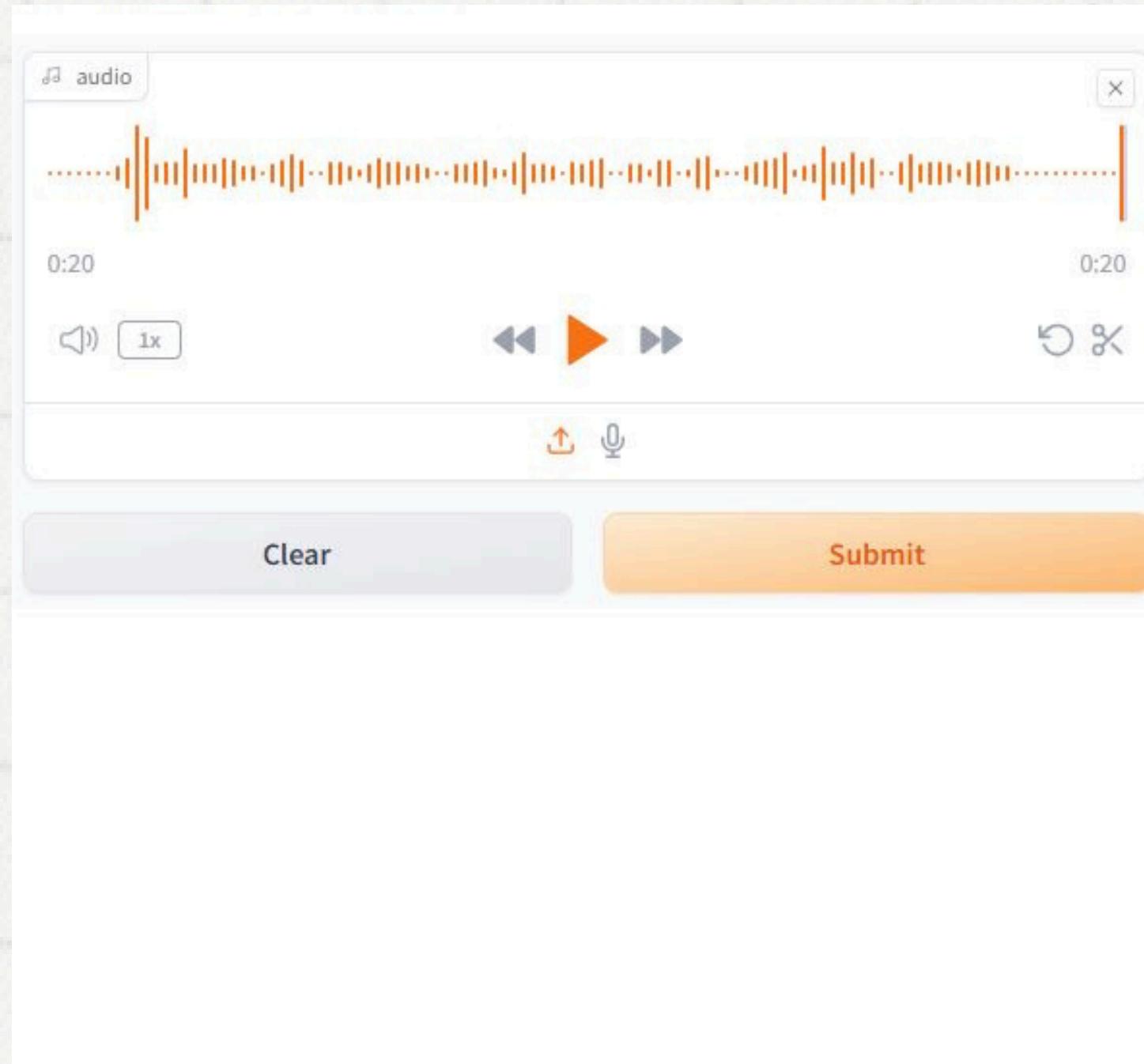
- Input: English text
- Text-to-Image Model: Uses diffusion models to create visuals
- Output: Generated image

Challenges:

Achieving accurate visual representation, model tuning.







Question & Answer

Thank you very much!