AI Models:

Text to Video – CogVideoX-2b

An open-source video generating AI model that will generate a video from text input at 720x480 resolution 8 fps for 6 seconds. Even for high end consumer GPUs inference times can exceed 30 min and most GPUs will likely take hours to generate.

Text to Speech – SpeechT5 (Microsoft)

A text-to-speech model that is proficient at a variety of speech functionalities but is utilised for its text-to-speech ability. Multiple voices are available to use, and the default has been changed to speaker 7306.

OOP:

Encapsulation-

The AI codes were wrapped using private variables prefixed with underscores such as \_pipe, \_speaker\_embedding and \_is\_loaded to hide implementation details. This makes it easier for GUI developer and OOP architect to read and debug their coding that uses these wrappers due to the reduced code length. The methods load\_model(), generate\_speech() and generate\_video() allow a simple interface for implementation of additional OOP patterns (such as polymorphism, method overriding, multiple decorators) without having to understand or code the speaker embeddings, tensor operations and pipeline configurations.