

HUGGING GPT INTRO

THANKS TO:

HUGGINGGPT: SOLVING AI TASKS WITH
CHATGPT AND ITS FRIENDS IN
HUGGINGFACE, YONGLIANG SHEN, KAITAO
SONG, XU TAN, DONGSHENG LI, WEIMING LU
AND YUETING ZHUANG (THE FIRST TWO
AUTHORS CONTRIBUTE EQUALLY)

INTRODUCTION:

WE INTRODUCE A COLLABORATIVE SYSTEM
THAT CONSISTS OF AN LLM AS THE
CONTROLLER AND NUMEROUS EXPERT MODELS
AS COLLABORATIVE EXECUTORS (FROM
HUGGINGFACE HUB). THE WORKFLOW OF OUR
SYSTEM CONSISTS OF FOUR STAGES:

HTTPS://GITHUB.COM/INSIGHTBUILDER

example.jpg. Then please describe the new image with your voice. Request Stage #1: Task Planning Stage #2: Model Selection task: pose-control task: image-to-text task: image-class. In-context task-model assignment: task: pose-control W hustvl/yolos-tiny image: <resource-2> image: <resource-2> image: [Image-1] facebook/detr-resnet-101 Query TahaDouaji/detr-doc-table-detection task dependency task: object-det. Model Cards in HuggingFace task: pose-to-image task: object-det. task: text-to-speech image: <resource-2>i text: a reading girl image: <resource-2>1 text: <resource-5 task: image-class. image: <resource-1> Stage #4: Response Generation Stage #3: Task Execution -[T4] object-det. r[T1] pose-control facebook/detr-resnet openpose control Hybrid Endpoints **Bounding boxes** [{bounding box}] [image-2] HuggingFace Endpoint with probabilities [T2] pose-to-image [T5] image-to-text obj-det. img: <Slot-2> lym/sd-ctl-pose nlpconnect/vit-gpt2 Predictions "a girl sitting on ..." [image-3] -[T3] image-class. [T6] text-to-speech Local Endpoint facebook/fastspeech google/vit [classes with prob] [audio-1] Response a girl sitting on a bed reading a book [ext-1] = <resource-5> [Image-2] = <resource-1> [Image-3] = <resource-2> [Audio-1] [Image-1] [Image-4] The image you gave me is of "boy". The first thing I did was use to analyze the pose of the boy in the image. I fed that image, specifically located at [Image-1], into the model to get an output image of the pose, located at [Image-2]. Then I take that pose and generate a new image, located at [Image-3], using 4 and the instruction "a girl is reading". Next, I ran object detection on the image generated in the previous step, to get predictions of the objects present in the image, including the box location. These predictions were run through the model. The box was drawn on the image, which is located at [Image-4]. Then, I used the predictions of a image classification model 😕 and a image caption model model to generate the caption for newly generated image. It generated text: "a girl sitting on a bed reading a book". Lastly, I ran the model to convert the generated text into audio, located at [Audio-1].

please generate an image where a girl is reading a book, and her pose is the same as the boy in the image

HUGGING GPT UNDER THE HOOD: PROMPTS

ADVANTAGES OF HUGGINGGPT

- TASK PLANNING FROM PROMPT
- INTEGRATING WITH MULTIPLE MODELS AND SELECTION
- EXECUTE THE TASK FROM HUMAN INPUT
- TAKE THE OUTPUT, MAKE IT HUMAN READABLE/HEARABLE.
- MODELS ARE LOCAL

HUGGING GPT CHALLENGES

- 1. NEED LARGE LANGUAGE MODEL
- 2. LIMITATION IN CONTEXT MEMORY
- 3. CONTROLLING THE MODEL & TASK
 - SELECTION
- 4.PROVIDING A PROGRAMMABLE DATA
 - **INGESTION ROUTE**
- **5.NEEDS A HUMAN TO WRITE THE**
 - **INITIAL PROMPT**
- CHAT GPT MODEL IS PROMPTED AT TASK SELECTION STEP
- FINAL RESPONSE IS CONTROLLED BY THE PROMPT

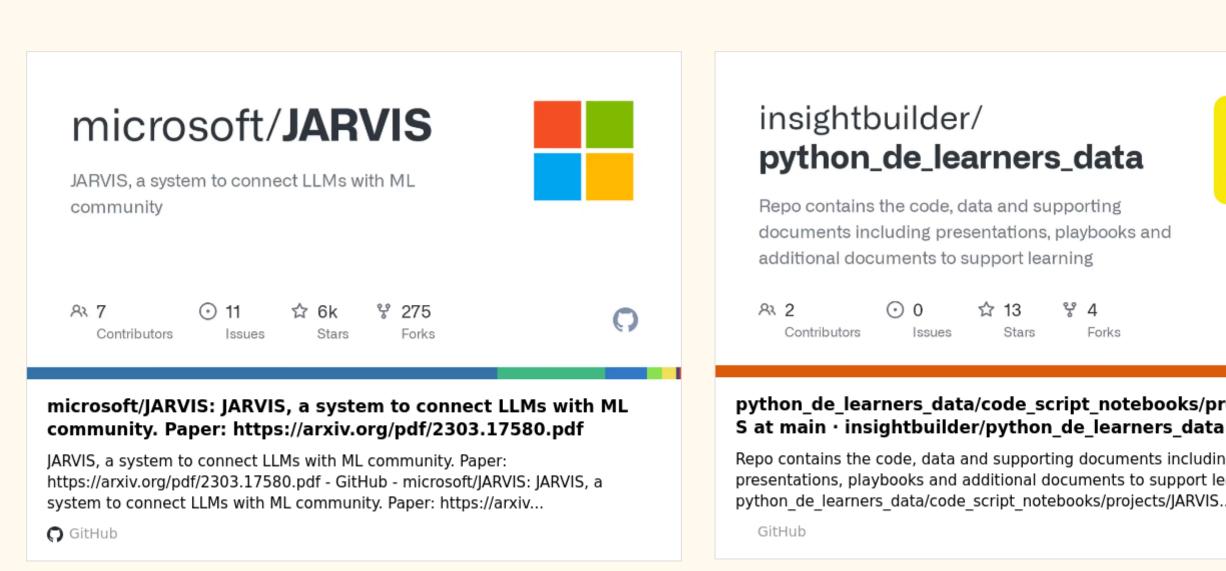
HTTPS://GITHUB.COM/INSIGHTBUILDER

LETS HEAD TO GIT REPO

NOW WE ARE TALKING. PRACTICE???

CODE WILL DOWNLOAD ALL THE MODEL!!!! REQUIRE OPENAI &

HF KEY.





Repo contains the code, data and supporting documents including

python de learners data/code script notebooks/projects/JARVIS...

GitHub

presentations, playbooks and additional documents to support learning -

THANKS FOR WATCHING

