Innovative Tech Challenge

Open Source Generative AI with Hugging Face

Rina Buoy, PhD

Applied AI Researcher, Techo Startup Center

Source: Open Source Models with Hugging Face

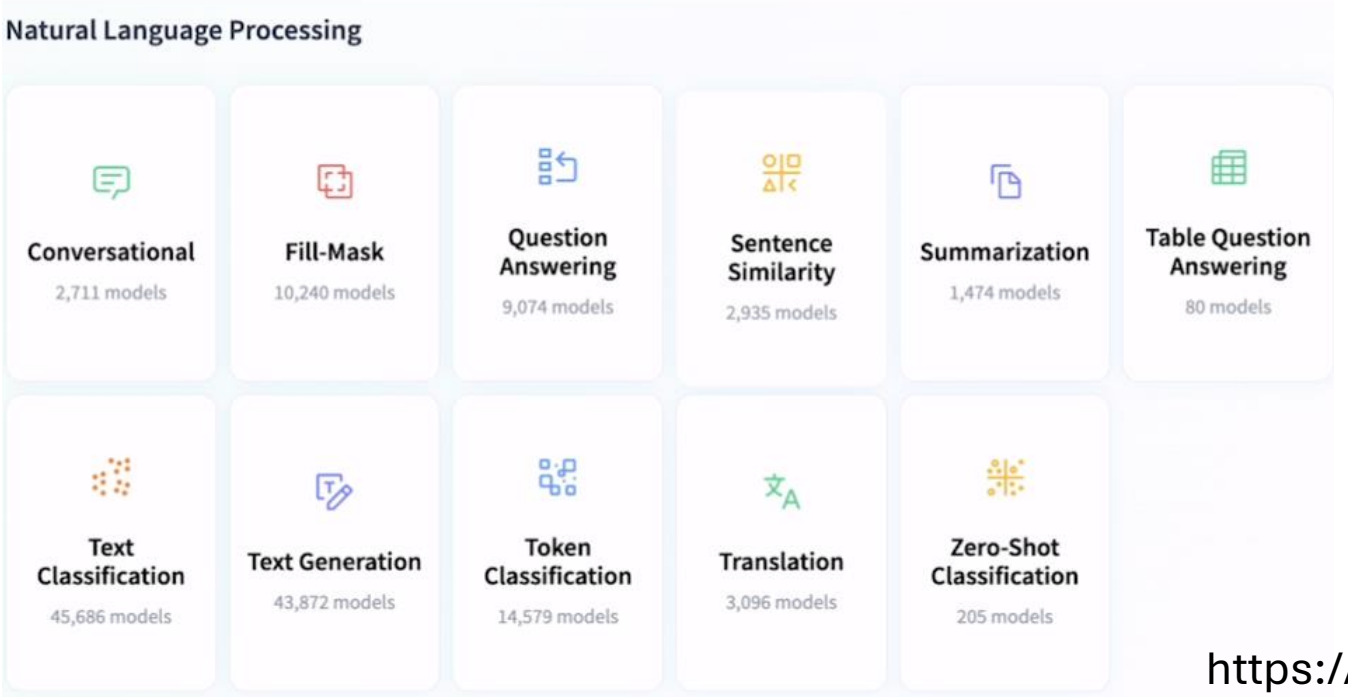
Hugging Face Library

Multimodal Data

🤗 Hugging Face Library

Here are some common NLP tasks:

- Text generation
- Sentence similarity
- Summarization
- Machine translation

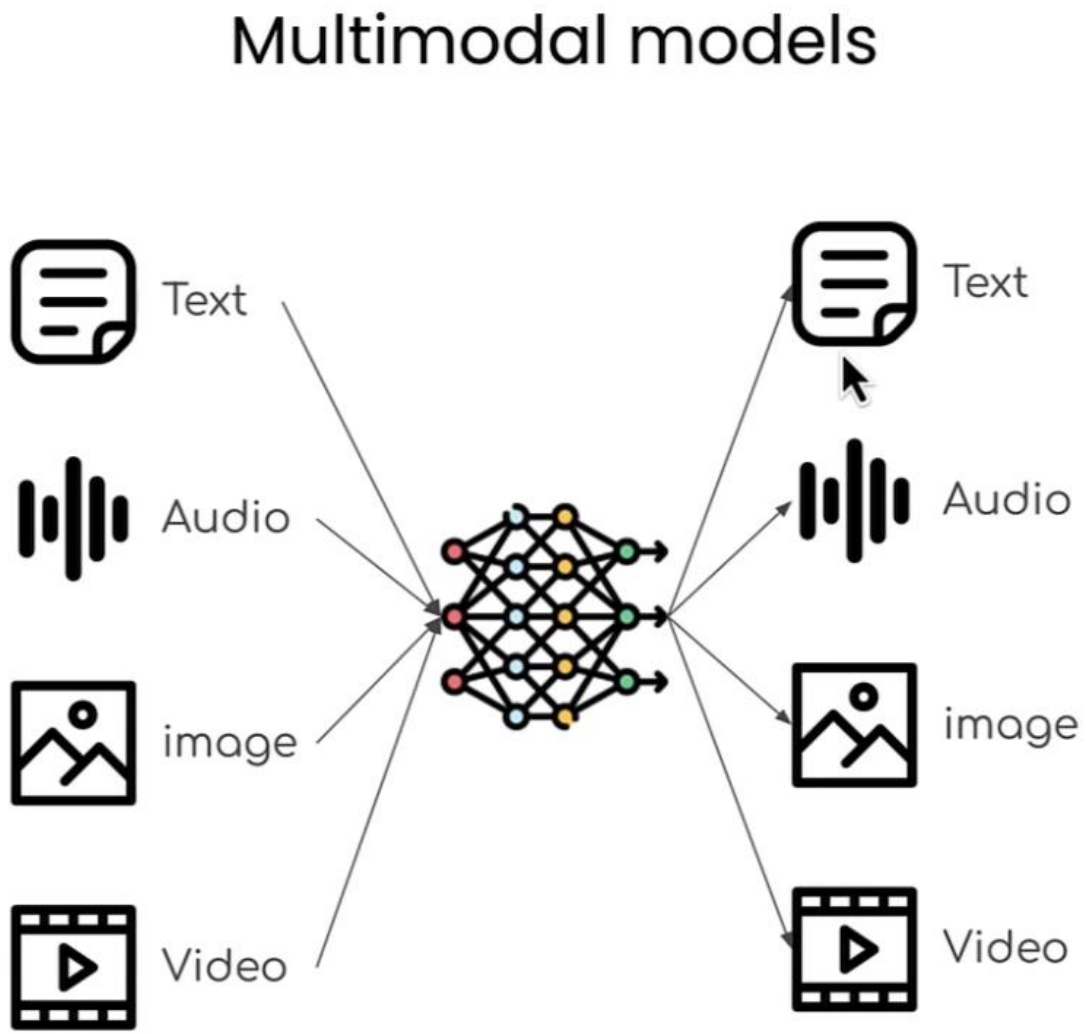


Hugging Face

- Datasets
- Models
- Deployment

https://github.com/ksm26

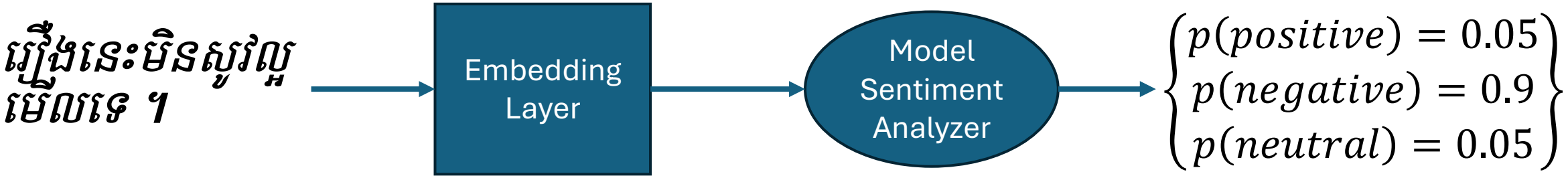
Multimodal GenAI Applications



https://github.com/ksm26

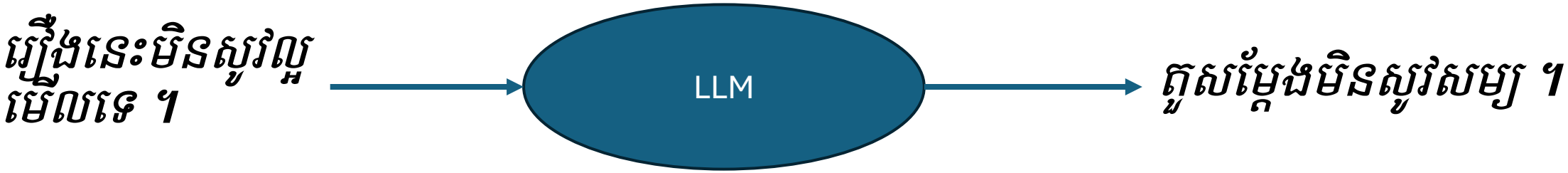
Text

- Written or spoken language.
- Using text modality to understand, analyze, and create linguistic content.
- This type of data includes natural language in various forms such as sentences, documents, conversations, etc.



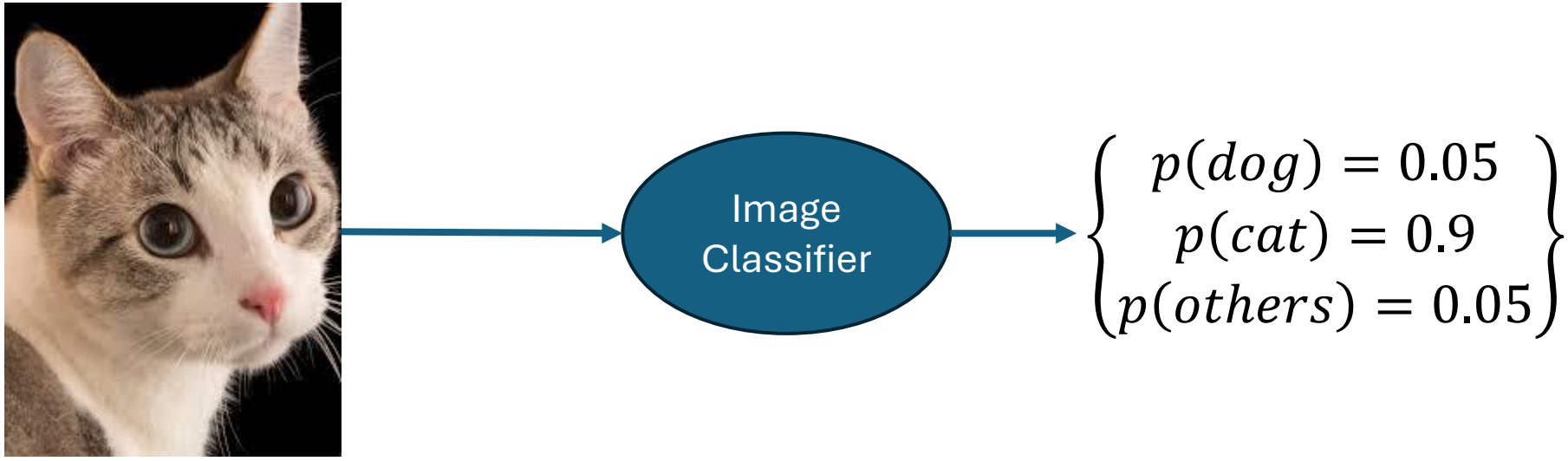
Text

- LLM takes input as text and outputs text.

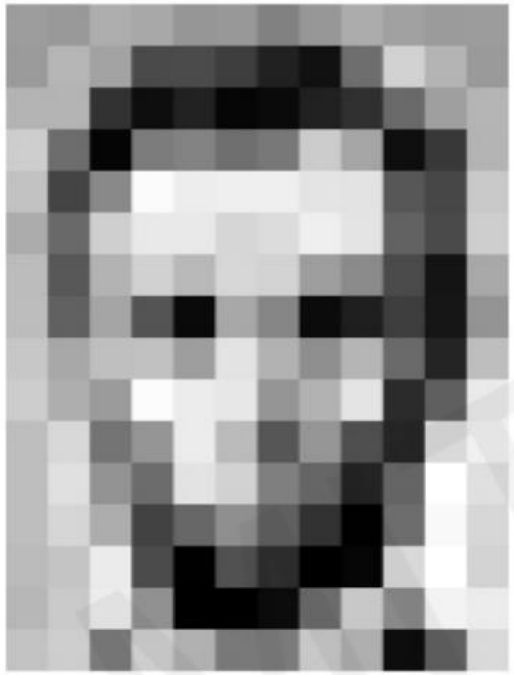


Image

- Visual or graphical in nature, specifically images.
- Using image modality to understand, analyze, and manipulate visual content, making use of techniques like computer vision (CV).



Images are numbers



157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

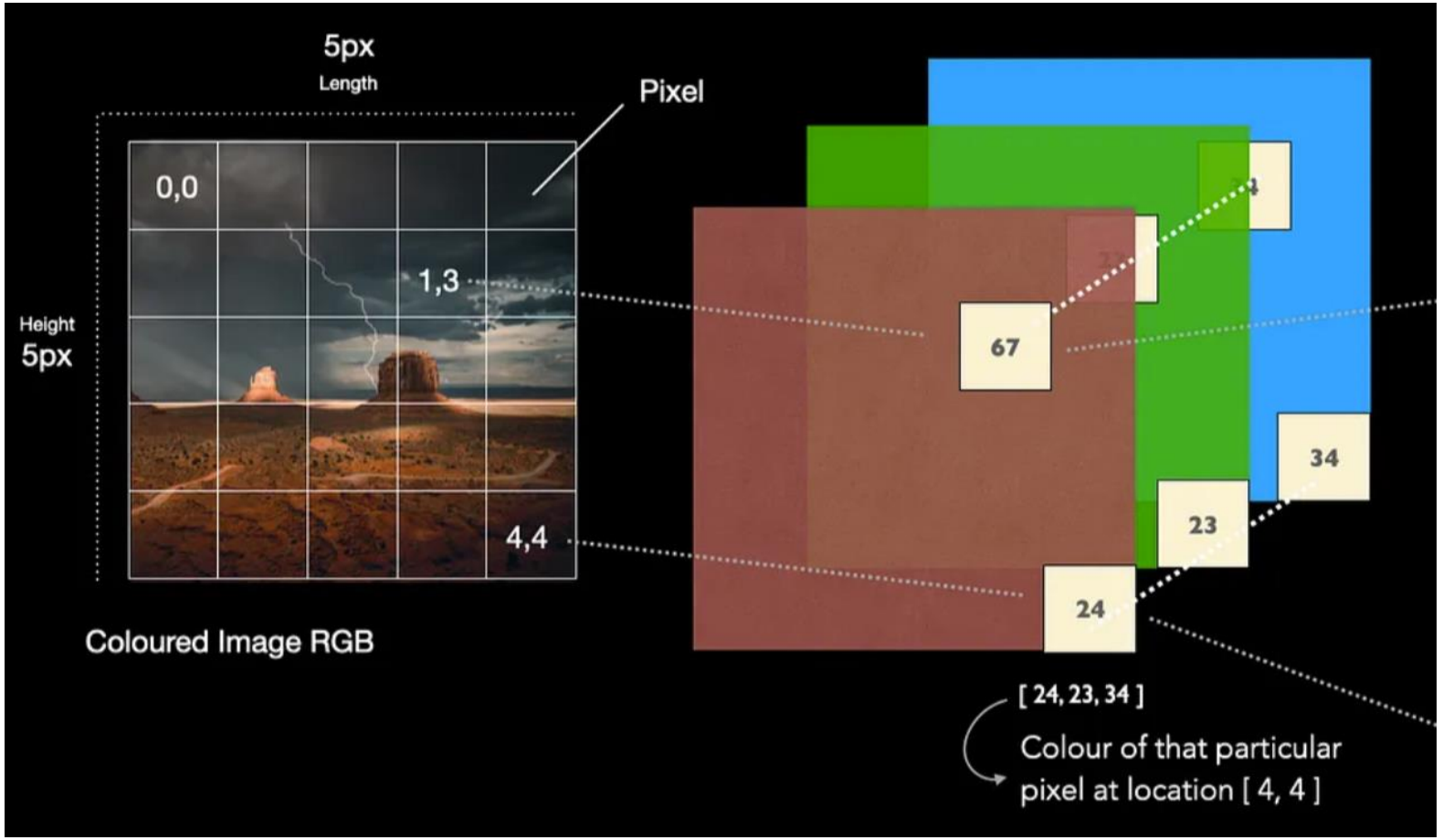
What the computer sees

157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

An image is just a matrix of numbers [0,255]!
i.e., 1080x1080x3 for an RGB image

http://introtodeeplearning.com/slides/6S191_MIT_DeepLearning_L3.pdf

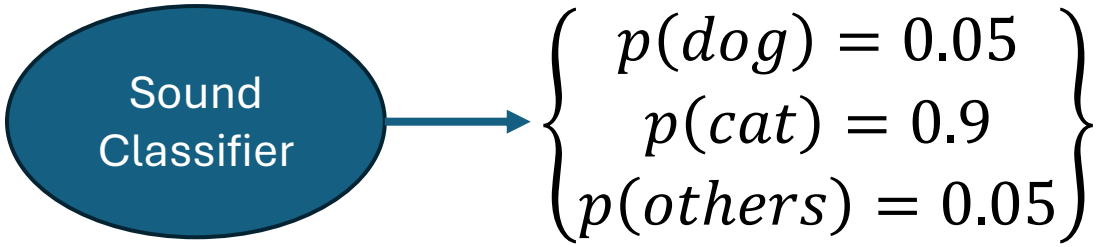
Images are numbers



https://medium.com/advanced-deep-learning/decoding-image-representation-understanding-the-structure-of-rgb-images-6a211eb8800d

Audio

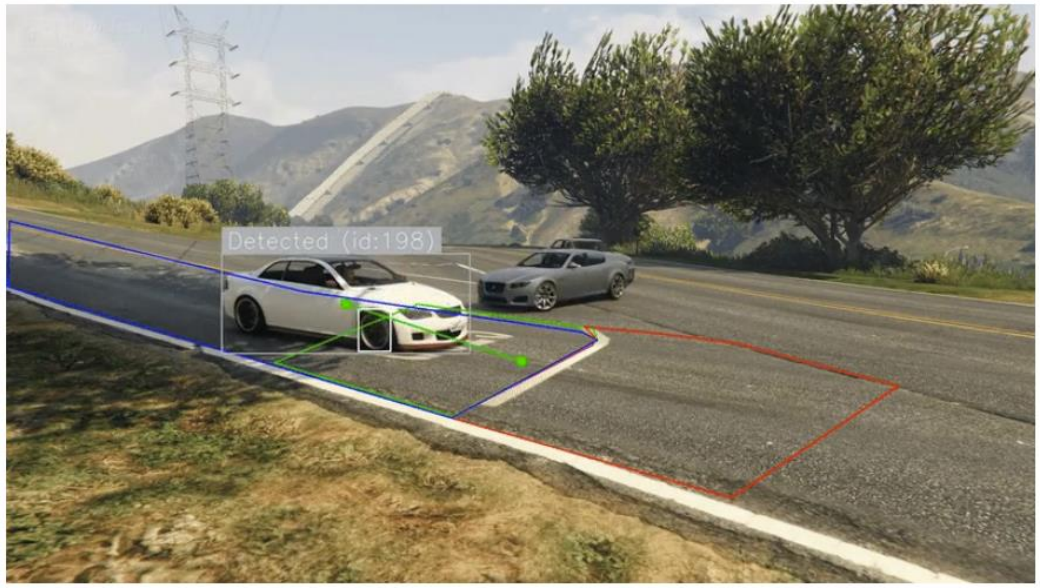
- Sound, including speech, music, and environmental noises.
- Using audio modality focus on analyzing, interpreting, and generating audio signals to perform tasks like speech recognition, sound classification, and audio synthesis.



http://introtodeeplearning.com/slides/6S191_MIT_DeepLearning_L2.pdf

Video

- A sequence of images (frames) combined with audio, capturing both visual and auditory elements over time.
- Using video modality analyze these dynamic inputs to perform a wide variety of tasks such as object tracking, action recognition, and video generation.



https://blog.roboflow.com/stop-sign-violation-detection/

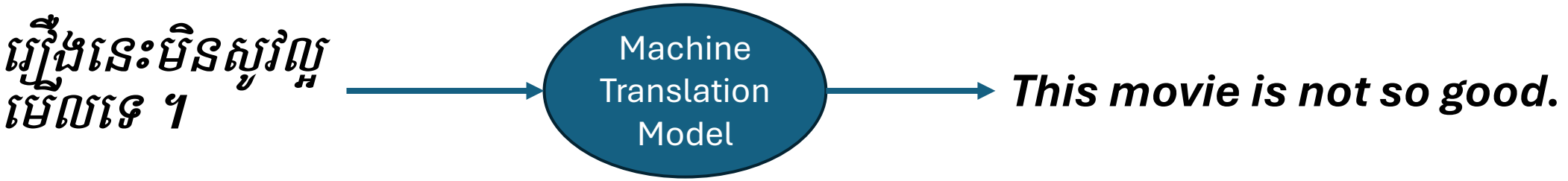
GenAI Applications

Translation and Summarization

- **Translation:**
  - Converting text from one language to another.
  - Understand the meaning of a source language and generate equivalent text in the target language.
- **Summarization**
  - Shortening long pieces of text into concise summaries while retaining the essential information.
  - Automatically extract key points from large documents, articles, or conversations.

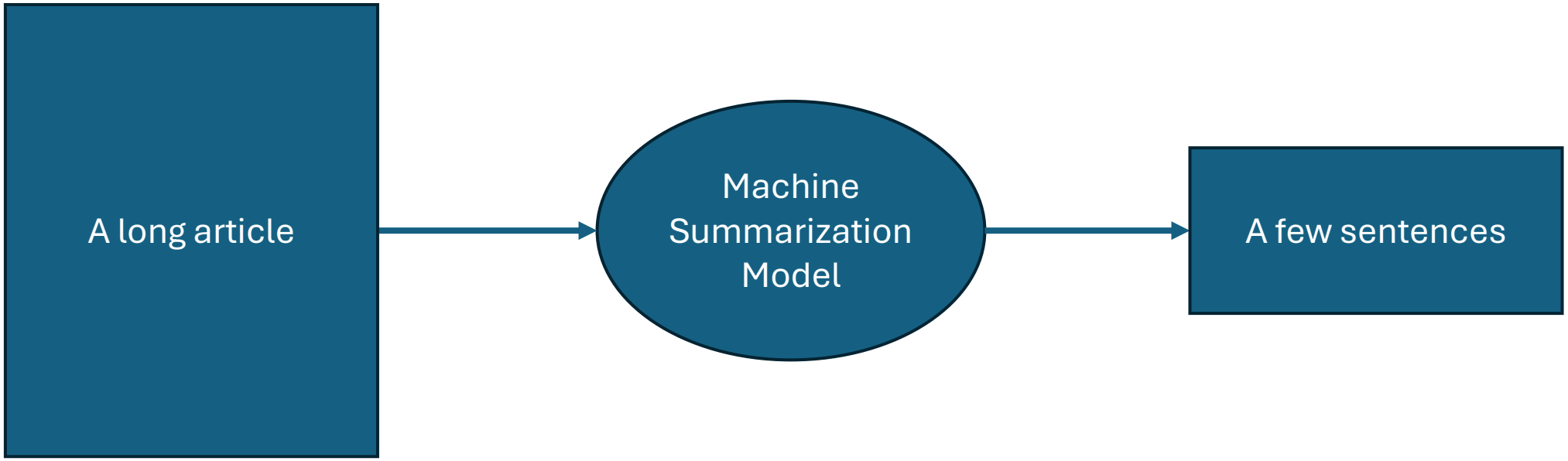
Translation

- Meta AI has built a single AI model, NLLB-200, that is the first to translate across 200 different languages with state-of-the-art quality that has been validated through extensive evaluations for each of them.



Summarization

- BART (by Facebook) is for text generation (e.g. summarization, translation).



Demo Time

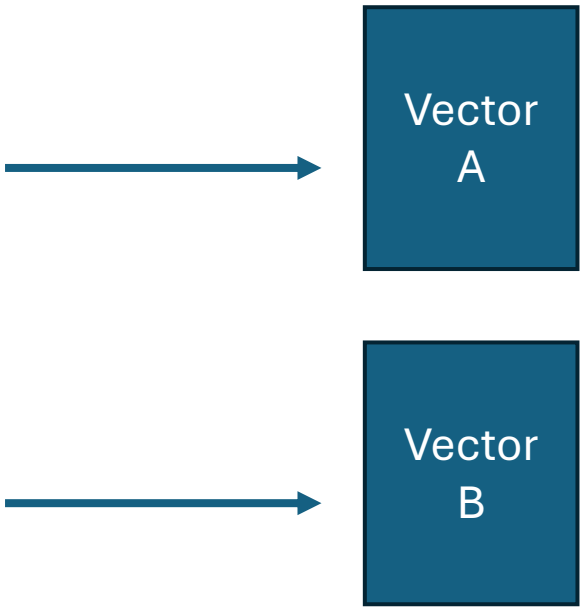
L3_Translation_and_Summarization_InnoTechChallenge.ipynb

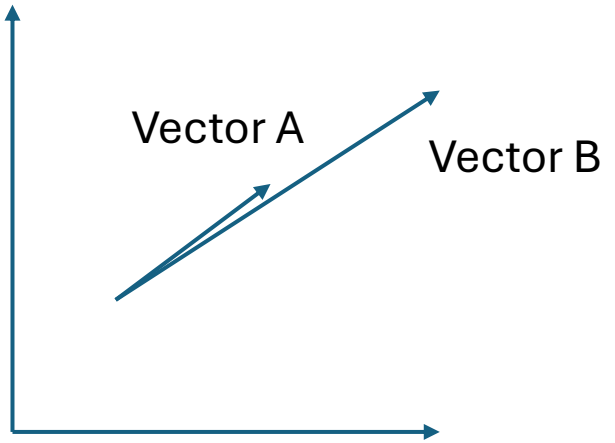
Sentence Embedding

- Transforming a sentence into a fixed-size vector (a series of numbers) that captures the semantic meaning of the sentence.
- Represent sentences in a form that is computationally manageable while preserving their key meanings and relationships.

រឿងនេះមិនសូវល្អមើលទេ ។

រឿងនេះមិនល្អទេ ។





cosine similarity is a measure of similarity between two non-zero vectors.

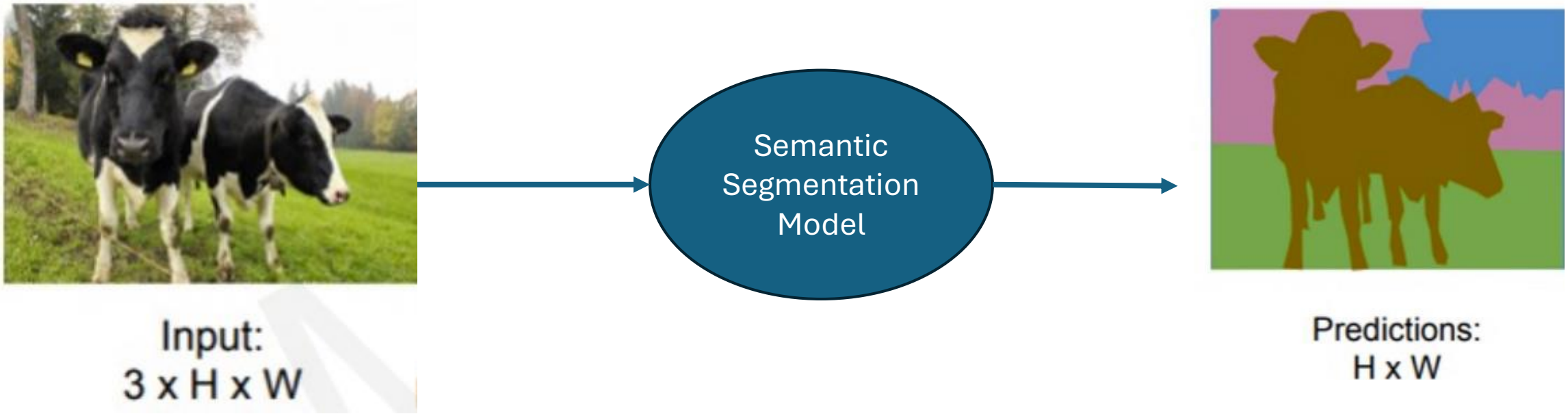
Demo Time

- L4_Sentence_Embeddings_InnoTechChallenge.ipynb

Object Detection

http://introtodeeplearning.com/slides/6S191_MIT_DeepLearning_L3.pdf

Semantic Segmentation



Depth Estimation

Depth Estimation Model

https://www.ikomia.ai/blog/depth-anything-monocular-estimation-revolution

Demo Time

- L8_object_detection_InnoTechChallenge.ipynb
- Image_Segmentation_InnovativeTechChallenge.ipynb
- Depth_Estimation_InnovativeTechChallenge.ipynb

Next, LLM, Speech etc.