

Lecture 10: Multimodality - X and Language



Sara Beery | 4/22/25



Language in deep learning

- Transformers flexible, non-local representations with long context windows

 First in NLP, then ViT
- Auto-regressive training at internet scale
 - I.e. "next word prediction"
- Accessible/usable without AI skills
 - Large-scale deployment
 - Prompt engineering
 - Human correction/verification
- Hallucination, bias, mode collapse...



*****Claude



Joint embeddings

- CLIP contrastive image/language training
- SigLIP introduces sigmoid loss to handle normalization issues with softmax
- What about other modalities? Other structured knowledge?
 - CLAP, BioCLIP, TaxaBind



Generative vision-language models

- Text-conditional image/video generation
 - Stable diffusion, DALL-E, Sora
- Image-conditioning in generative language models
 - Ie "Does this image contain an American Robin?"
 - Use image encoders that are accessible/interpretable to the language models, but more tokens per image than CLIP (ie 170 tokens for an image in GPT-40)



Prompt engineering

- Prompts are brittle. "A photo of a dog" vs "A picture of a dog" will give very different results
- In-context or "chain of thought" prompting
- It can be hard to track/understand errors (reasoning is wrong, answer is right?)



Code assistants

- Can make AI more accessible
- Doesn't require foundational knowledge
- What are the pros and cons?



GitHub Copilot: Al that builds with you

8 GitHub Copilot

Ask a question or type '/' for topics

Agentic Al

Query: How many muffins can each kid have for it to be fair?



Generated Code

def execute_command(image): image_patch = ImagePatch(image) muffin_patches = image_patch.find("muffin") kid_patches = image_patch.find("kid") return str(len(muffin_patches) // len(kid_patches))



- AI that writes and executes code (ie Viper GPT, many more)
- Enables more complex reasoning
- Provides structure to force reliance on real data and functions/reduce hallucinations

https://arxiv.org/abs/2303.08128

Agentic Al



The AI co-scientist system design

https://research.google/blog/accelerating-scientific-breakthroughs-with-an-ai-co-scientist/