Artificial Intelligence – Week 11

Instructor: Teerapong Panboonyuen

Course Repository:
https://github.com/kaopanboonyuen/SC310005_ArtificialIntelligence_2025s1

🎯 Objective

This week, you will explore **pretrained Generative AI models** across multiple modalities and experiment with your own inputs/prompts. Your goal is to understand model behavior, generate outputs, and compare models.

Extra points: Students who explore newer or more impressive models on Hugging Face.

You will:
✅ Explore text-to-image generation with Stable Diffusion
✅ Experiment with image-to-image style transfer using diffusion
✅ Chat with text generation models (GPT-2 or similar)
✅ Generate captions from images using BLIP or similar models
✅ Try to discover newer or improved models on Hugging Face
✅ Document your process, results, and observations

📚 Dataset / Inputs

You may use any publicly available images or create your own prompts.

Example starter image:

- Barack Obama Sample Face

📚 Topic Focus

- **Text-to-Image:** Stable Diffusion, DALL·E, MidJourney
- **Image-to-Image:** Style Transfer / Diffusion
- **Text-to-Text:** GPT-2, GPT-Neo, LLaMA
- **Image-to-Text:** BLIP, Flamingo
- **Experimentation:** prompt engineering, sampling strategies, fine-tuning tricks, LoRA

📝 Assignment Instructions

🎨 1. Text ➡️ Image

- Use a pretrained text-to-image model (Stable Diffusion recommended)
- Generate **at least 3 different images** with creative prompts
- Experiment with style, colors, or concept
- Optional: explore newer text-to-image models for extra points

---

🔄 2. Image ➡️ Image (Style Transfer / Diffusion)

- Take an input image (e.g., sample face) and apply style transfer
- Generate **at least 2 variations**
- Optional: Try different diffusion models to see style differences

---

✍️ 3. Text ➡️ Text (Chat with GPT-2 or similar)

- Prepare **at least 10 prompts/questions**
- Generate responses using a pretrained text model
- Analyze quality, coherence, and creativity
- Optional: experiment with **temperature, top-k, top-p** sampling

---

🖼️ 4. Image ➡️ Text (Captioning with BLIP)

- Input at least **5 images** (your own or public datasets)
- Generate captions and analyze accuracy/descriptiveness
- Optional: explore newer image-to-text models for bonus points

---

📑 Report Instructions

- Document your **inputs, outputs, and model configurations**
- Include screenshots of **generated images or text responses**
- Compare different models if multiple are used
- Highlight any **tricks, prompt strategies, or fine-tuning methods**
- Extra points for discovering **newer or more impressive models**

💡 Tips for Students

- Organize folder structure: `/images/input`, `/images/output`
- Keep **prompts organized** for easy analysis
- Try **creative prompts** to explore model limits
- Use a free GPU on Colab for large models like Stable Diffusion
- Optional: use **LoRA** or lightweight fine-tuning for custom outputs

---

🎨 Deliverables

- **Report / Notebook**: Python scripts or Colab notebook
- Include: generated images, captions, text outputs, model configs
- Bonus points if newer models outperform standard ones

🚀 Getting Started

- **Stable Diffusion:** Documentation
- **BLIP:** Documentation
- **GPT-2 / Transformers:** Documentation

📅 Submission Deadline

- To be announced in FB class.
- Be **creative** 🎨🤖 and try to push model capabilities!
