HOMEWORK 623
PAPER PRESENTATION *

10-423/10-623 GENERATIVE AI
http://423.mlcourse.org

OUT: 11/04/24
DUE: 12/02/24

1 Paper Presentation

For this assignment you will read a recent generative AI paper and video record a (short, 5-10 minute) presentation about it.

We will follow the role-playing methodology from Collin Raffel and Alec Jacobson.

https://colinraffel.com/blog/role-playing-seminar.html

You may select one of the following roles and present your paper in that format. (The instructions below were copied and adapted from 10-719's use of the role-playing paper presentation approach.)

1. **Scientific Peer Reviewer:** Complete a full review of the paper as if it were submitted to a conference. Follow the guidelines for NeurIPS reviewers to produce your review. In particular, please answer Questions 1 to 10 under "Review Form", including assigning an overall score.
2. **Archaeologist:** Determine where this paper sits in the context of previous and subsequent work. Find and briefly report on both: (1) a prior paper that substantially influenced the current paper, and (2) a more recent paper that cites this current paper.
3. **Academic Researcher:** You're an academic researcher working on a new project in this area. Propose an imaginary follow-up project that builds on the current paper. Pretend that this new project has been successful, and write up a brief introduction for a paper about your project using the five-point structure provided here (under "The Introduction"). You do not need to actually *write* the introduction, but instead should present in the style of an introduction to this new project.
4. **Industry Practitioner:** You work at a company or organization developing an application or product of your choice (one that has not already been suggested in a prior session). Describe the application/product in detail, and bring a convincing pitch for why you should be paid to implement the method in the paper for this particular application.
5. **Private Investigator:** You are a detective who needs to run a background check on one of the paper's authors. Where have they worked? What did they study? What previous projects might have led to working on this one? What motivated them to work on this project?
6. **Social Impact Assessor:** Identify how this paper self-assesses its positive or negative impact on the world. Have any additional positive social impacts been left out? What are possible negative social

*Compiled on Monday 4th November, 2024 at 19:37

impacts that were overlooked or omitted? Please read this short paper to see examples.

A few notes:

- For each of the above roles, your presetation *must* start with a summary of the paper.
- You are welcome to record with your webcam turned off (Khan Academy style).
- After submission, we will share all the HW623 video presentations with the rest of class via Piazza, so others can learn from you as well.
- You will make slides and submit them alongside your video recording.

2 Select a paper

To select a paper, you should pick of the ones below. If you would like to present a generative AI paper not on this list, send a Private Note to the "Instructors" on Piazza with the paper and 1-2 sentences about why you are interested in it.

[1] Diederik P. Kingma and Prafulla Dhariwal. *Glow: Generative Flow with Invertible 1x1 Convolutions*. 2018. arXiv: 1807.03039 [stat.ML].

[2] Yang Song et al. *Score-Based Generative Modeling through Stochastic Differential Equations*. 2021. arXiv: 2011.13456 [cs.LG].

[3] Yuntao Bai et al. *Constitutional AI: Harmlessness from AI Feedback*. 2022. arXiv: 2212.08073 [cs.CL].

[4] Arpit Bansal et al. *Cold Diffusion: Inverting Arbitrary Image Transforms Without Noise*. 2022. arXiv: 2208.09392 [cs.CV]. URL: https://arxiv.org/abs/2208.09392.

[5] Cheng Lu et al. *DPM-Solver: A Fast ODE Solver for Diffusion Probabilistic Model Sampling in Around 10 Steps*. 2022. arXiv: 2206.00927 [cs.LG]. URL: https://arxiv.org/abs/2206.00927.

[6] Narek Tumanyan, Michal Geyer, Shai Bagon, and Tali Dekel. *Plug-and-Play Diffusion Features for Text-Driven Image-to-Image Translation*. 2022. arXiv: 2211.12572 [cs.CV]. URL: https://arxiv.org/abs/2211.12572.

[7] Xiaohui Zeng et al. *LION: Latent Point Diffusion Models for 3D Shape Generation*. 2022. arXiv: 2210.06978 [cs.CV]. URL: https://arxiv.org/abs/2210.06978.

[8] Amanda Bertsch, Uri Alon, Graham Neubig, and Matthew R. Gormley. *Unlimiformer: Long-Range Transformers with Unlimited Length Input*. 2023. arXiv: 2305.01625 [cs.CL]. URL: https://arxiv.org/abs/2305.01625.

[9] Keqin Chen et al. *Shikra: Unleashing Multimodal LLM's Referential Dialogue Magic*. 2023. arXiv: 2306.15195 [cs.CV].

[10] Zhengyang Geng, Ashwini Pokle, and J. Zico Kolter. *One-Step Diffusion Distillation via Deep Equilibrium Models*. 2023. arXiv: 2401.08639 [cs.CV].

[11] Rohit Girdhar et al. *ImageBind: One Embedding Space To Bind Them All*. 2023. arXiv: 2305.05665 [cs.CV].

[12] Yutong He, Ruslan Salakhutdinov, and J. Zico Kolter. *Localized Text-to-Image Generation for Free via Cross Attention Control*. 2023. arXiv: 2306.14636 [cs.CV].

[13] Shaohan Huang et al. *Language Is Not All You Need: Aligning Perception with Language Models*. 2023. arXiv: 2302.14045 [cs.CL]. URL: https://arxiv.org/abs/2302.14045.

[14] Siyuan Huang et al. *Diffusion-based Generation, Optimization, and Planning in 3D Scenes*. 2023. arXiv: 2301.06015 [cs.CV]. URL: https://arxiv.org/abs/2301.06015.

[15] Huiqiang Jiang, Qianhui Wu, Chin-Yew Lin, Yuqing Yang, and Lili Qiu. *LLMLingua: Compressing Prompts for Accelerated Inference of Large Language Models*. 2023. arXiv: 2310.05736 [cs.CL]. URL: https://arxiv.org/abs/2310.05736.

[16] Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. *BLIP-2: Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models*. 2023. arXiv: 2301.12597 [cs.CV].

[17] Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. *Visual Instruction Tuning*. 2023. arXiv: 2304.08485 [cs.CV]. URL: https://arxiv.org/abs/2304.08485.

[18] William Peebles and Saining Xie. *Scalable Diffusion Models with Transformers*. 2023. arXiv: 2212.09748 [cs.CV].

[19] Dustin Podell et al. *SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis*. 2023. arXiv: 2307.01952 [cs.CV].

[20] Yujia Qin et al. *ToolLLM: Facilitating Large Language Models to Master 16000+ Real-world APIs*. 2023. arXiv: 2307.16789 [cs.AI]. URL: https://arxiv.org/abs/2307.16789.

[21] Rafael Rafailov et al. *Direct Preference Optimization: Your Language Model is Secretly a Reward Model*. 2023. arXiv: 2305.18290 [cs.LG].

[22] Zineng Tang et al. *CoDi-2: In-Context, Interleaved, and Interactive Any-to-Any Generation*. 2023. arXiv: 2311.18775 [cs.CV].

[23] Qingyun Wu et al. *AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation*. 2023. arXiv: 2308.08155 [cs.AI]. URL: https://arxiv.org/abs/2308.08155.

[24] Yilun Xu et al. *PFGM++: Unlocking the Potential of Physics-Inspired Generative Models*. 2023. arXiv: 2302.04265 [cs.LG].

[25] Jianwei Yang et al. *Set-of-Mark Prompting Unleashes Extraordinary Visual Grounding in GPT-4V*. 2023. arXiv: 2310.11441 [cs.CV].

[26] Shunyu Yao et al. *Tree of Thoughts: Deliberate Problem Solving with Large Language Models*. 2023. arXiv: 2305.10601 [cs.CL].

[27] Lijun Yu et al. *MAGVIT: Masked Generative Video Transformer*. 2023. arXiv: 2212.05199 [cs.CV]. URL: https://arxiv.org/abs/2212.05199.

[28] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. *Adding Conditional Control to Text-to-Image Diffusion Models*. 2023. arXiv: 2302.05543 [cs.CV].

[29] Ziwei Zhang et al. *Graph Meets LLMs: Towards Large Graph Models*. 2023. arXiv: 2308.14522 [cs.LG]. URL: https://arxiv.org/abs/2308.14522.

[30] Wentao Zhu et al. *Human Motion Generation: A Survey*. 2023. arXiv: 2307.10894 [cs.CV]. URL: https://arxiv.org/abs/2307.10894.

[31] Shoufa Chen et al. *GenTron: Diffusion Transformers for Image and Video Generation*. 2024. arXiv: 2312.04557 [cs.CV]. URL: https://arxiv.org/abs/2312.04557.

[32] Yali Du, Hui Sun, and Ming Li. *A Joint Learning Model with Variational Interaction for Multilingual Program Translation*. 2024. arXiv: 2408.14515 [cs.SE]. URL: https://arxiv.org/abs/2408.14515.

[33] Wenqi Fan et al. *Graph Machine Learning in the Era of Large Language Models (LLMs)*. 2024. arXiv: 2404.14928 [cs.LG]. URL: https://arxiv.org/abs/2404.14928.

[34] Tsu-Jui Fu et al. *Guiding Instruction-based Image Editing via Multimodal Large Language Models*. 2024. arXiv: 2309.17102 [cs.CV]. URL: https://arxiv.org/abs/2309.17102.

[35] Zhibin Gou et al. *CRITIC: Large Language Models Can Self-Correct with Tool-Interactive Critiquing*. 2024. arXiv: 2305.11738 [cs.CL]. URL: https://arxiv.org/abs/2305.11738.

[36] Daya Guo et al. *DeepSeek-Coder: When the Large Language Model Meets Programming – The Rise of Code Intelligence*. 2024. arXiv: 2401.14196 [cs.SE]. URL: https://arxiv.org/abs/2401.14196.

[37] Zahra Kadkhodaie, Florentin Guth, Eero P. Simoncelli, and Stéphane Mallat. *Generalization in diffusion models arises from geometry-adaptive harmonic representations*. 2024. arXiv: 2310.02557 [cs.CV].

[38] Aditya Kusupati et al. *Matryoshka Representation Learning*. 2024. arXiv: 2205.13147 [cs.LG].

[39] Bin Lin et al. *Video-LLaVA: Learning United Visual Representation by Alignment Before Projection*. 2024. arXiv: 2311.10122 [cs.CV]. URL: https://arxiv.org/abs/2311.10122.

[40] Ji Lin et al. *VILA: On Pre-training for Visual Language Models*. 2024. arXiv: 2312.07533 [cs.CV].

[41] Shuming Ma et al. *The Era of 1-bit LLMs: All Large Language Models are in 1.58 Bits*. 2024. arXiv: 2402.17764 [cs.CL].

[42] Tsendsuren Munkhdalai, Manaal Faruqui, and Siddharth Gopal. *Leave No Context Behind: Efficient Infinite Context Transformers with Infini-attention*. 2024. arXiv: 2404.07143 [cs.CL]. URL: https://arxiv.org/abs/2404.07143.

[43] Yotam Nitzan et al. *Lazy Diffusion Transformer for Interactive Image Editing*. 2024. arXiv: 2404.12382 [cs.CV]. URL: https://arxiv.org/abs/2404.12382.

[44] Baptiste Rozière et al. *Code Llama: Open Foundation Models for Code*. 2024. arXiv: 2308.12950 [cs.CL]. URL: https://arxiv.org/abs/2308.12950.

[45] Yutao Sun et al. *You Only Cache Once: Decoder-Decoder Architectures for Language Models*. 2024. arXiv: 2405.05254 [cs.CL]. URL: https://arxiv.org/abs/2405.05254.

[46] Vikram Voleti et al. *SV3D: Novel Multi-view Synthesis and 3D Generation from a Single Image using Latent Video Diffusion*. 2024. arXiv: 2403.12008 [cs.CV]. URL: https://arxiv.org/abs/2403.12008.

[47] Fei Wang, Xingchen Wan, Ruoxi Sun, Jiefeng Chen, and Sercan Ö. Arık. *Astute RAG: Overcoming Imperfect Retrieval Augmentation and Knowledge Conflicts for Large Language Models*. 2024. arXiv: 2410.07176 [cs.CL]. URL: https://arxiv.org/abs/2410.07176.

[48] Junlin Wang, Jue Wang, Ben Athiwaratkun, Ce Zhang, and James Zou. *Mixture-of-Agents Enhances Large Language Model Capabilities*. 2024. arXiv: 2406.04692 [cs.CL]. URL: https://arxiv.org/abs/2406.04692.

[49] Xingyao Wang et al. *Executable Code Actions Elicit Better LLM Agents*. 2024. arXiv: 2402.01030 [cs.CL]. URL: https://arxiv.org/abs/2402.01030.

[50] Jia-Yu Yao et al. *LLM Lies: Hallucinations are not Bugs, but Features as Adversarial Examples*. 2024. arXiv: 2310.01469 [cs.CL]. URL: https://arxiv.org/abs/2310.01469.

[51] Zhiqiang Yuan et al. *TRANSAGENT: An LLM-Based Multi-Agent System for Code Translation*. 2024. arXiv: 2409.19894 [cs.SE]. URL: https://arxiv.org/abs/2409.19894.

[52] Lvmin Zhang and Maneesh Agrawala. *Transparent Image Layer Diffusion using Latent Transparency*. 2024. arXiv: 2402.17113 [cs.CV].

[53] Peiyuan Zhang et al. *Long Context Transfer from Language to Vision*. 2024. arXiv: 2406.16852 [cs.CV]. URL: https://arxiv.org/abs/2406.16852.

3 Video Recording and Submission

Video Recording For the presentation, you will record yourself and then send us the link to the recording. Here are the steps you should take:

1. Open Zoom and start your "Personal Meeting". Be sure to turn on your microphone.
2. Click "Record" and then "Record to the Cloud".
3. Introduce yourself by name. Then go ahead and present for 5-10 minutes.
4. *Extremely important:* Now click "End the Meeting". Your recording will be uploaded to the cloud and processed.
5. Go to `https://cmu.zoom.us/recording` and wait ($\sim$ 10 minutes) for your recording to finish processing.
6. Click the "Share..." button on the left of the recording, then click "Copy Sharing Information" and paste the clipboard text into Gradescope.

Submit to Gradescope You should submit your presentation video and slides to Gradescope.

7. Open the HW623 assignment in Gradescope.
8. Paste the video recording information that you copy/pasted above into the free-text field on the assignment.
9. Then upload your slides as a PDF on the file-upload question.
10. In Gradescope, save and submit this assignment.