# HOMEWORK 623
# PAPER PRESENTATION *

10-423/623/723 GENERATIVE AI
http://423.mlcourse.org

OUT: 3/30/26
DUE: 4/20/26

## 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 30th March, 2026 at 15:10

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.
- **Regarding AI Use:** You may use AI to facilitate your understanding of your chosen paper. However all materials must be created and submitted must be your own with no AI assistance.

## 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] Ben Mildenhall et al. *NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis*. 2020. arXiv: 2003.08934 [cs.CV]. URL: https://arxiv.org/abs/2003.08934.

[2] Yuntao Bai et al. *Constitutional AI: Harmlessness from AI Feedback*. 2022. arXiv: 2212.08073 [cs.CL]. URL: https://arxiv.org/abs/2212.08073.

[3] Jonathan Ho et al. *Video Diffusion Models*. 2022. arXiv: 2204.03458 [cs.CV]. URL: https://arxiv.org/abs/2204.03458.

[4] Anthony Brohan et al. *RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control*. 2023. arXiv: 2307.15818 [cs.RO]. URL: https://arxiv.org/abs/2307.15818.

[5] Rohit Girdhar et al. *ImageBind: One Embedding Space To Bind Them All*. 2023. arXiv: 2305.05665 [cs.CV]. URL: https://arxiv.org/abs/2305.05665.

[6] 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]. URL: https://arxiv.org/abs/2301.12597.

[7] 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.

[8] Neel Nanda, Lawrence Chan, Tom Lieberum, Jess Smith, and Jacob Steinhardt. *Progress measures for grokking via mechanistic interpretability*. 2023. arXiv: 2301.05217 [cs.LG]. URL: https://arxiv.org/abs/2301.05217.

[9] William Peebles and Saining Xie. *Scalable Diffusion Models with Transformers*. 2023. arXiv: 2212.09748 [cs.CV]. URL: https://arxiv.org/abs/2212.09748.

[10] 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.

[11] Zineng Tang et al. *CoDi-2: In-Context, Interleaved, and Interactive Any-to-Any Generation*. 2023. arXiv: 2311.18775 [cs.CV]. URL: https://arxiv.org/abs/2311.18775.

[12] 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.

[13] Shunyu Yao et al. *Tree of Thoughts: Deliberate Problem Solving with Large Language Models*. 2023. arXiv: 2305.10601 [cs.CL]. URL: https://arxiv.org/abs/2305.10601.

[14] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. *Adding Conditional Control to Text-to-Image Diffusion Models*. 2023. arXiv: 2302.05543 [cs.CV]. URL: https://arxiv.org/abs/2302.05543.

[15] 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.

[16] 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.

[17] Albert Gu and Tri Dao. *Mamba: Linear-Time Sequence Modeling with Selective State Spaces*. 2024. arXiv: 2312.00752 [cs.LG]. URL: https://arxiv.org/abs/2312.00752.

[18] 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.

[19] 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.

[20] 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]. URL: https://arxiv.org/abs/2402.17764.

[21] Lingchen Meng et al. *DeepStack: Deeply Stacking Visual Tokens is Surprisingly Simple and Effective for LMMs*. 2024. arXiv: 2406.04334 [cs.CV]. URL: https://arxiv.org/abs/2406.04334.

[22] Rafael Rafailov et al. *Direct Preference Optimization: Your Language Model is Secretly a Reward Model*. 2024. arXiv: 2305.18290 [cs.LG]. URL: https://arxiv.org/abs/2305.18290.

[23] Zhihong Shao et al. *DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models*. 2024. arXiv: 2402.03300 [cs.CL]. URL: https://arxiv.org/abs/2402.03300.

[24] Haotian Tang et al. *HART: Efficient Visual Generation with Hybrid Autoregressive Transformer*. 2024. arXiv: 2410.10812 [cs.CV]. URL: https://arxiv.org/abs/2410.10812.

[25] 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.

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

[27] Shuyan Zhou et al. *WebArena: A Realistic Web Environment for Building Autonomous Agents*. 2024. arXiv: 2307.13854 [cs.AI]. URL: https://arxiv.org/abs/2307.13854.

[28] A. Feder Cooper et al. *Extracting memorized pieces of (copyrighted) books from open-weight language models*. 2025. arXiv: 2505.12546 [cs.CL]. URL: https://arxiv.org/abs/2505.12546.

[29] Diego Gosmar and Deborah A. Dahl. *Hallucination Mitigation using Agentic AI Natural Language-Based Frameworks*. 2025. arXiv: 2501.13946 [cs.CL]. URL: https://arxiv.org/abs/2501.13946.

[30] Katja Schwarz, Denys Rozumnyi, Samuel Rota Bulò, Lorenzo Porzi, and Peter Kontschieder. *A Recipe for Generating 3D Worlds From a Single Image*. 2025. arXiv: 2503.16611 [cs.CV]. URL: https://arxiv.org/abs/2503.16611.

[31] Jacob Mitchell Springer et al. *Overtrained Language Models Are Harder to Fine-Tune*. 2025. arXiv: 2503.19206 [cs.CL]. URL: https://arxiv.org/abs/2503.19206.

[32] Teng Wang et al. *Towards Hierarchical Multi-Step Reward Models for Enhanced Reasoning in Large Language Models*. 2025. arXiv: 2503.13551 [cs.CL]. URL: https://arxiv.org/abs/2503.13551.

[33] Jiachen Zhu, Xinlei Chen, Kaiming He, Yann LeCun, and Zhuang Liu. *Transformers without Normalization*. 2025. arXiv: 2503.10622 [cs.LG]. URL: https://arxiv.org/abs/2503.10622.

## 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 7-10 minutes. (If your presentation exceeds 11 minutes, you will be penalized.)
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.