

Course Project

Ju Sun

Computer Science & Engineering

University of Minnesota, Twin Cities

March 19, 2026

Timeline & L^AT_EX template

- Teaming up: Mar 26
<https://docs.google.com/spreadsheets/d/1dKlKW7dailnLtcrTu9Cyn1lZeute97QvuLYJM5yV6oM/edit?usp=sharing>
- Proposal (5%, 1–2 pages): Mar 29
- Recorded progress lightning talk (5%, 5 mins): Apr 18
- Progress report (5%, 3–4 pages): Apr 18
- Final report (25%, 7–8 pages): May 14 (Final grade: May 19)

All page counts exclude references

Template for all writeups: ICLR 2026 L^AT_EX style

[https:](https://github.com/ICLR/Master-Template/raw/master/iclr2026.zip)

[//github.com/ICLR/Master-Template/raw/master/iclr2026.zip](https://github.com/ICLR/Master-Template/raw/master/iclr2026.zip)

Add `\iclrfinalcopy` to the L^AT_EX preamble to make your names visible

Groups

2026-Spring-CSCI5527-Project-Teams

File Edit View Insert Format Data Tools Extensions Help

100% Helvet... 10

	A	B	C	D	E
1	Group ID	Student 1 (Name, Email ID)	Student 2 (Name, Email ID)	Student 3 (Name, Email ID)	Student 4 (Name, Email ID)
2	Instruction Group	Ju Sun, jusun	Jiandong Chen, chen8111	Wenjie Zhang, zhan7867	
3	Name Your Group Here!				
4	Name Your Group Here!				
5	Name Your Group Here!				
6	Name Your Group Here!				
7	Name Your Group Here!				
8	Name Your Group Here!				
9	Name Your Group Here!				
10	Name Your Group Here!				
11	Name Your Group Here!				
12	Name Your Group Here!				
13	Name Your Group Here!				
14	Name Your Group Here!				
15	Name Your Group Here!				
16	Name Your Group Here!				
17	Name Your Group Here!				
18	Name Your Group Here!				
19	Name Your Group Here!				
20	Name Your Group Here!				
21	Name Your Group Here!				
22	Name Your Group Here!				
23	Name Your Group Here!				
24	Name Your Group Here!				
25	Name Your Group Here!				
26	Name Your Group Here!				
27	Name Your Group Here!				

- Each team: 3 ~ 4 students; get my approval for exceptions
- All submissions as a team (in Canvas as group assignment); the team gets the same score

Computing resources

- Prototyping
 - * Colab Pro <https://colab.research.google.com/>
 - * Local installation of Jupyter Notebook
<https://jupyter.org/>
 - * MSI notebook notebooks.msi.umn.edu
(<https://www.msi.umn.edu/support/faq/how-do-i-get-started-jupyter-notebooks>)
- Large-scale jobs: **submit them to MSI GPU queues**
 - * MSI quick start
<https://www.msi.umn.edu/quick-start-guides>
 - * Slurm scheduler tutorial
<https://www.msi.umn.edu/slurm>
 - * MSI access notes (from Canvas home)

Five necessary components

- What problem?
- Why interesting?
- Previous work
- Your goals
- Plan and milestones

We encourage exploration and allow failures

Project ideas

Roughly by ascending level of difficulty

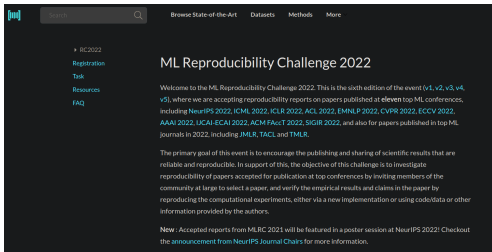
- Literature survey/review (least favorable given the good summarization capabilities of AI tools nowadays)
- Novel applications
- Novel methods
- Novel theories

Excerpt from a research project is fine, but you should describe your own contributions

A coherent account of recent **papers** in a focused topic

- Description and comparison of main ideas, or
- Implementation and comparison of performance, or
- Both of the above

should **complement** the topics we cover in the course



<https://paperswithcode.com/rc2022>

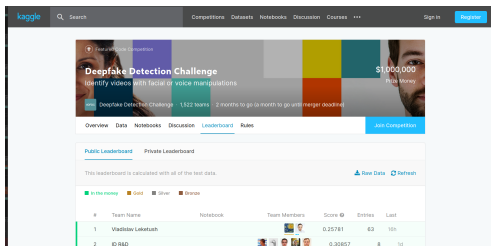
Random topics

- DL for noneuclidean data (e.g., graph NN, manifold NN)
- transformer models for sequential data
- generative models (e.g., GAN, VAE, normalization flow, diffusion models)
- 2nd order methods for deep learning
- constrained optimization for deep learning
- differential programming
- universal approximation theorems
- DL for 3D reconstruction
- DL for video understanding and analysis
- DL for solving PDEs
- DL for material discovery
- DL for inverse problems
- RL for games
- RL for robotics
- DL for medical imaging
- DL for (astro)physics
- DL for chemistry
- adversarial attacks; robustness of DL
- privacy, fairness in DL
- visualization for DNN
- network quantization and compression
- hardware/software platforms for DL
- automated ML; architecture search
- optimization/generalization theory of DL
- large vision-language models

Novel applications

Apply DL to **new** application problems

- A good place to start: Kaggle <https://www.kaggle.com/>



- Think about data availability

Google dataset search

<https://datasetsearch.research.google.com/>

- Think about GPUs

Where to find inspirations

Talk to AI, before anything below:

- arXiv machine learning
<https://arxiv.org/list/cs.LG/recent>
- Recent conference papers
 - ML: NeurIPS, ICML, ICLR, etc
 - CV: ICCV, ECCV, CVPR, etc
 - NLP: ACL, EMNLP, etc
 - Robotics: ICRA, etc
 - Graphics: SIGGRAPH, etc
- Talk to researchers (including TAs and me)!

Novel methods

Create new **NN models or training algorithms** to improve the state-of-the-art

Where to start: **Talk to AI**, before anything below

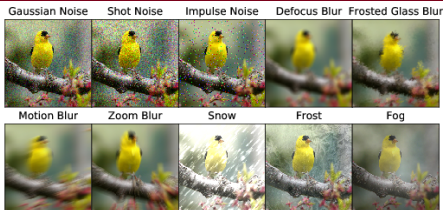
- Kaggle (again)!
- arXiv machine learning and recent conference papers
- MLRC



The screenshot shows the homepage of the ML Reproducibility Challenge 2022. The page has a dark theme with a navigation bar at the top containing a search bar and links for 'Browse State-of-the-Art', 'Datasets', 'Methods', and 'More'. On the left side, there is a sidebar menu with links for 'RC2022', 'Registration', 'Task', 'Resources', and 'FAQ'. The main content area features the title 'ML Reproducibility Challenge 2022' and a welcome message. The welcome message states: 'Welcome to the ML Reproducibility Challenge 2022. This is the sixth edition of the event (v1, v2, v3, v4, v5), where we are accepting reproducibility reports on papers published at eleven top ML conferences, including NeurIPS 2022, ICML 2022, ICLR 2022, ACL 2022, EMNLP 2022, CVPR 2022, ECCV 2022, AAAI 2022, UCAI-ECAI 2022, ACM FAccT 2022, SIGIR 2022, and also for papers published in top ML journals in 2022, including JMLR, TACL and TMLR.' Below this, a paragraph explains the primary goal: 'The primary goal of this event is to encourage the publishing and sharing of scientific results that are reliable and reproducible. In support of this, the objective of this challenge is to investigate reproducibility of papers accepted for publication at top conferences by inviting members of the community at large to select a paper, and verify the empirical results and claims in the paper by reproducing the computational experiments, either via a new implementation or using code/data or other information provided by the authors.' At the bottom, a 'New' section states: 'New : Accepted reports from MLRC 2021 will be featured in a poster session at NeurIPS 2022! Checkout the announcement from NeurIPS Journal Chairs for more information.'

<https://paperswithcode.com/rc2020>

Novel methods



Credit: ImageNet-C <https://github.com/hendrycks/robustness>

WILDS

A benchmark of in-the-wild distribution shifts spanning diverse data modalities and applications, from tumor identification to wildlife monitoring to poverty mapping.

The v2.0 update adds unlabeled data to 8 datasets. The labeled data and evaluation metrics are exactly the same, so all previous results are directly comparable. Read our [release notes](#) to find out more!

WILDS paper

Unlabeled data paper (v2)

Github

Credit: WILDS <https://wilds.stanford.edu/>

Equally interesting to fool/fail the state-of-the-art, e.g., exploring robustness of DL, finding common limitations of state-of-the-art

Novel theories

Nothing is more practical than a good theory. – V. Vapnik

- universal approximation theorems
- nonconvex optimization
- generalization

Where to start: **Talk to AI**, before anything below

- Analyses of Deep Learning (Stanford, fall 2019)
<https://stats385.github.io/>
- Theories of Deep Learning (Stanford, fall 2017)
https://stats385.github.io/stats385_2017.github.io/
- Toward theoretical understanding of deep learning (ICML 2018 Tutorial)
<https://unsupervised.cs.princeton.edu/deeplearningtutorial.html>
- <https://sunju.org/teach/TMML-Fall-2021/>

Questions?