Machine Learning Research and Senior Design Project

Updated 09/03/2024

Xiaodong Qu

Table of Contents

Introduction to Machine Learning in Academia

- Brief overview of ML's role and significance in academic research and innovation.
- Key milestones in ML academic history to showcase its evolution and impact.

Defining Machine Learning:

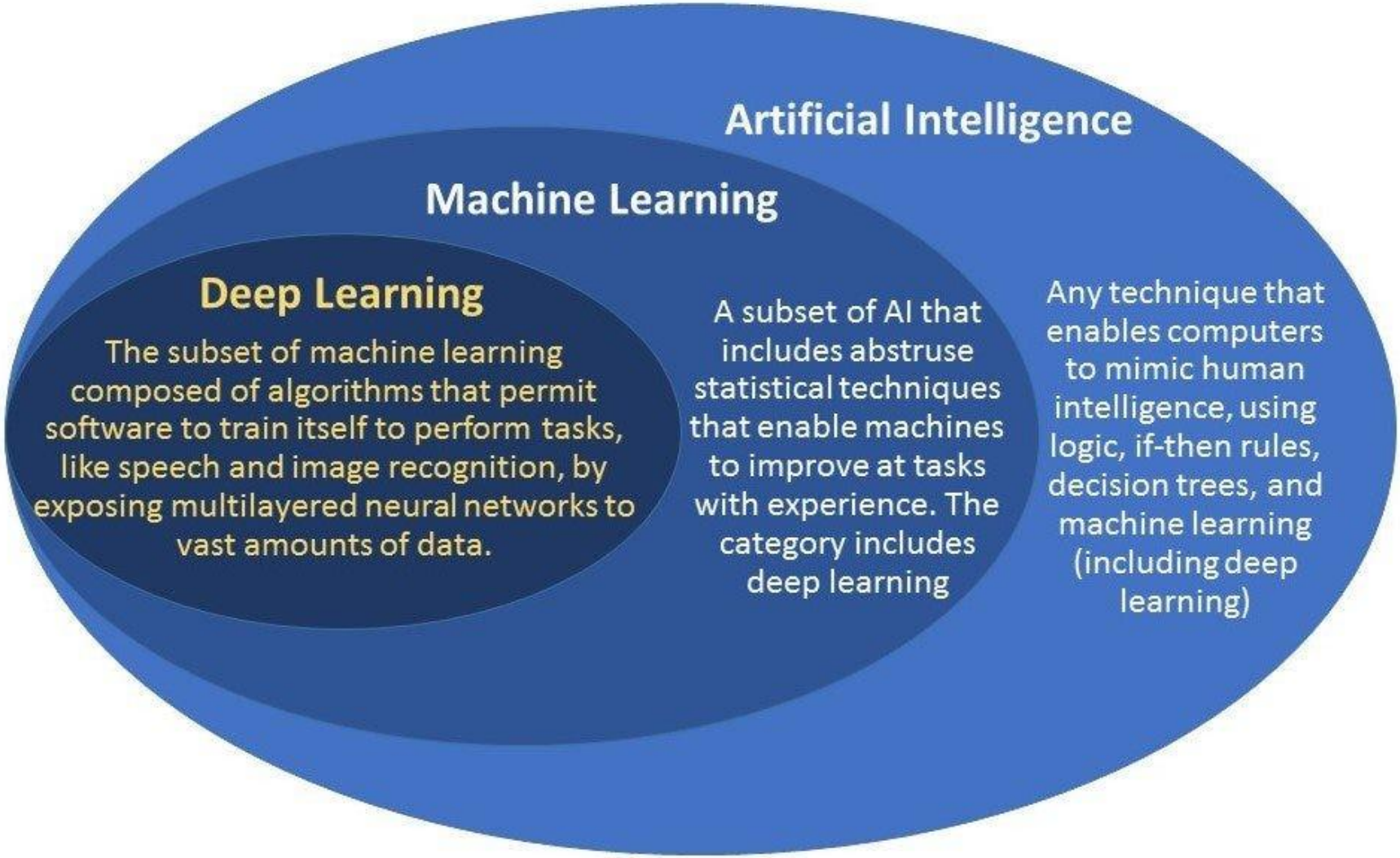
Briefly define machine learning as a subset of artificial intelligence that enables systems to learn and improve from experience without being explicitly programmed.

Emphasize ML's capability to uncover insights from data, leading to innovative solutions across various domains.

ML's Role in Academic Research:

Discuss how ML drives advancements in numerous academic fields, including healthcare (predictive diagnostics, personalized medicine), environmental science (climate modeling, conservation efforts), and engineering (autonomous systems, smart infrastructure).

Highlight the interdisciplinary nature of ML, fostering collaboration between computer science, statistics, biology, economics, and more.


Artificial Intelligence
Any technique that enables computers to mimic human intelligence, using logic, if-then rules, decision trees, and machine learning (including deep learning)
Machine Learning
A subset of AI that includes abstruse statistical techniques that enable machines to improve at tasks with experience. The category includes deep learning
Deep Learning
The subset of machine learning composed of algorithms that permit software to train itself to perform tasks, like speech and image recognition, by exposing multilayered neural networks to vast amounts of data.

Key Milestones in ML Academic History:

1950s: The inception of ML with Alan Turing's question, "Can machines think?", leading to the development of the Turing Test.

1967: The introduction of the "nearest neighbor" algorithm, laying foundational work for pattern recognition.

1980s: The resurgence of neural networks and the backpropagation algorithm, enabling the training of multi-layer networks.

1997: IBM's Deep Blue defeats world chess champion Garry Kasparov, demonstrating the potential of AI and ML.

2012: The success of AlexNet in the ImageNet competition, marking the beginning of the deep learning revolution in computer vision.

2018: Turing Awards, Nobel Prize of computing given to 'godfathers of AI'

Present: Continuous breakthroughs in natural language processing, generative models, and reinforcement learning, pushing the boundaries of what ML can achieve.

Fathers of the Deep Learning Revolution Receive ACM A.M. Turing Award

Bengio, Hinton and LeCun Ushered in Major Breakthroughs in Artificial Intelligence

ACM named Yoshua Bengio, Geoffrey Hinton, and Yann LeCun recipients of the 2018 ACM A.M. Turing Award for conceptual and engineering breakthroughs that have made deep neural networks a critical

Prerequisites and Foundational Knowledge

Outline necessary mathematical and programming skills (e.g., linear algebra, probability, Python).

Suggest prerequisite courses or resources to bridge knowledge gaps.

My ML courses in the last three years, and students Publications.

Core ML Courses and Advanced Electives

Core ML courses available,

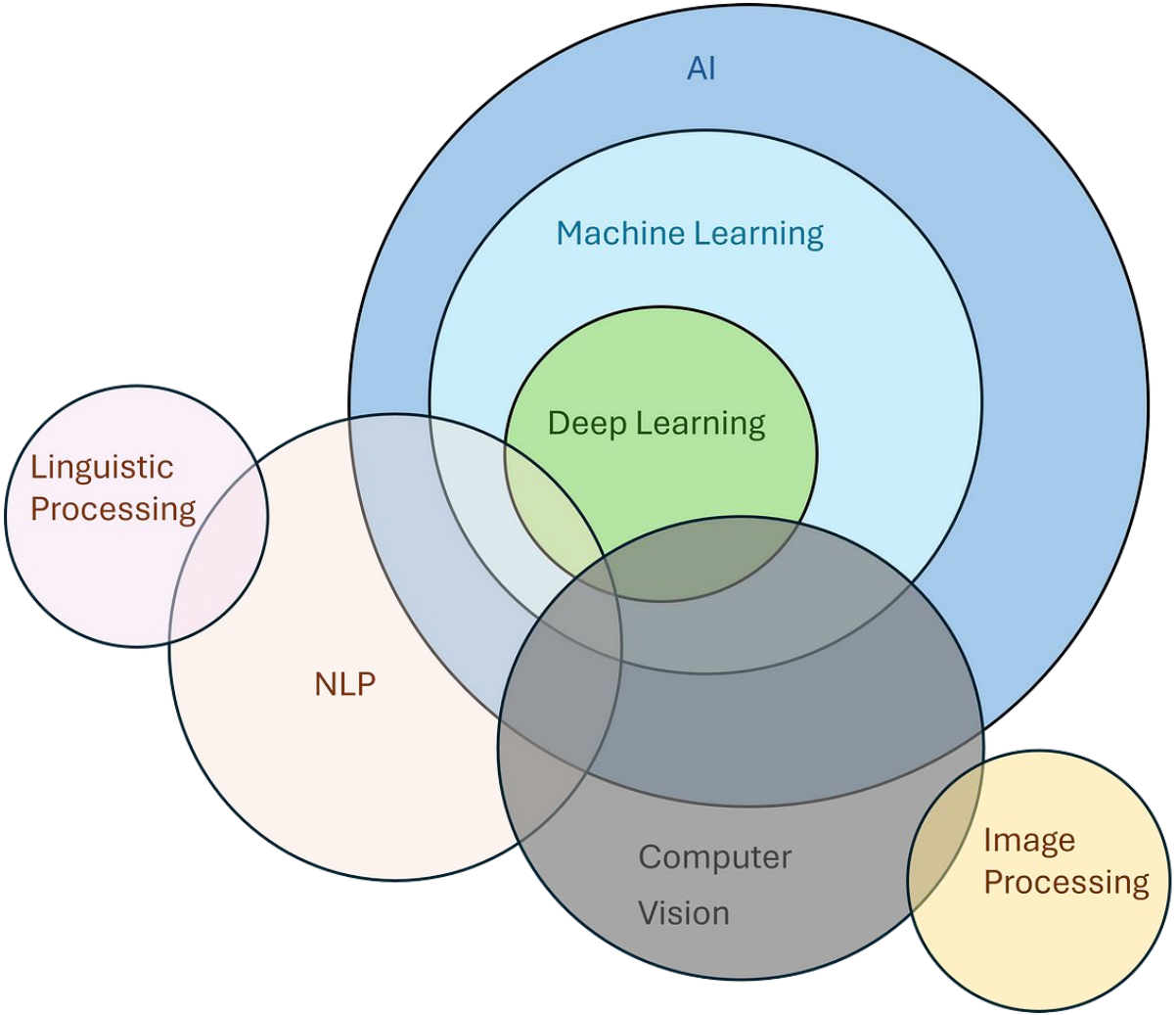
From Universities, or From Industry

Learning outcomes.

Introduce advanced electives that complement senior design projects, such as deep learning, natural language processing, or computer vision.

6 Best Machine Learning Courses in 2024

Course Title	Provider	Length	Cost	Certification
Machine Learning Specialization	Stanford Online & DeepLearning.AI on Coursera	2 months	Free to audit ($49 per month for a certificate)	✓
AWS Machine Learning Engineer Nanodegree	Udacity & Amazon Web Services	5 months	$249 per month	✓
Machine Learning A-Z: AI, Python & R + ChatGPT Prize	Udemy	42.5 hours	$119	✓
A Practical Guide to Machine Learning with Python	Educative	72.5 hours	$16.66 per month	✓
Machine Learning for All	University of London on Coursera	3 weeks	Free to audit ($49 for a certificate)	✓
Introduction to Machine Learning	Massachusetts Institute of Technology's Open Learning Library	13 weeks	Free	✗


AI
Machine Learning
Deep Learning
Linguistic Processing
NLP
Computer Vision
Image Processing

My ML course Fall 2024, 2023, and 2022

For Undergraduate and Graduate students

Using real world data to solve problems and publish research papers

Datasets

Algorithms

Experiments

Publications

Datasets

NeurIPS 2024 Datasets and Benchmarks Track

Examples: EEG Eye Net (2021), Amazon M2 (2023)

Top AI conferences

NeurIPS

ICLR

ICML

AAAI

Google Scholar

Top publications

Categories > Engineering & Computer Science > **Artificial Intelligence**

	Publication	h5-index	h5-median
1.	Neural Information Processing Systems	337	614
2.	International Conference on Learning Representations	304	584
3.	International Conference on Machine Learning	268	424
4.	AAAI Conference on Artificial Intelligence	220	341
5.	Expert Systems with Applications	165	228
6.	IEEE Transactions On Systems, Man And Cybernetics Part B, Cybernetics	155	212
7.	IEEE Transactions on Neural Networks and Learning Systems	149	215
8.	Neurocomputing	136	210
9.	International Joint Conference on Artificial Intelligence (IJCAI)	136	192
10.	Neural Computing and Applications	135	184

NeurIPS 2024 Datasets and Benchmarks Track

If you'd like to become a reviewer for the track, or recommend someone, please use this **form**.

The **Datasets and Benchmarks track** serves as a venue for high-quality publications, talks, and posters on highly valuable machine learning datasets and benchmarks, as well as a forum for discussions on how to improve dataset development. Datasets and benchmarks are crucial for the development of machine learning methods, but also require their own publishing and reviewing guidelines. For instance, datasets can often not be reviewed in a double-blind fashion, and hence full anonymization will not be required. On the other hand, they do require additional specific checks, such as a proper description of how the data was collected, whether they show intrinsic bias, and whether they will remain accessible. The Datasets and Benchmarks track is proud to support the open source movement by encouraging submissions of open-source libraries and tools that enable or accelerate ML research.

The previous editions of the Datasets and Benchmarks track were highly successful; you can view the accepted papers from 2021, 2002, and 2023, and the winners of the best paper awards 2021, 2022 and 2023

EEGEyeNet: a Simultaneous Electroencephalography and Eye-tracking Dataset and Benchmark for Eye Movement Prediction

Ard Kastrati, Martyna Beata Plomecka, Damian Pascual, Lukas Wolf, Victor Gillioz, Roger Wattenhofer, Nicolas Langer

Published: 29 Jul 2021, Last Modified: 14 Jul 2024 NeurIPS 2021 Datasets and Benchmarks Track (Round 1) Readers: Everyone Show Bibtex Show Revisions

Keywords: electroencephalography, eye-tracking, deep learning, dataset, benchmark

TL;DR: We present a large dataset of synchronised EEG and eye tracking data, together with a benchmark for evaluating eye tracking from EEG.

Abstract: We present a new dataset and benchmark with the goal of advancing research in the intersection of brain activities and eye movements. Our dataset, EEGEyeNet, consists of simultaneous Electroencephalography (EEG) and Eye-tracking (ET) recordings from 356 different subjects collected from three different experimental paradigms. Using this dataset, we also propose a benchmark to evaluate gaze prediction from EEG measurements. The benchmark consists of three tasks with an increasing level of difficulty: left-right, angle-amplitude and absolute position. We run extensive experiments on this benchmark in order to provide solid baselines, both based on classical machine learning models and on large neural networks. We release our complete code and data and provide a simple and easy-to-use interface to evaluate new methods.

Supplementary Material: zip

URL: http://www.eegeye.net

Community Implementations: 1 code implementation

Amazon-M2: A Multilingual Multi-locale Shopping Session Dataset for Recommendation and Text Generation

Wei Jin, Haitao Mao, Zheng Li, Haoming Jiang, Chen Luo, Hongzhi Wen, Haoyu Han, Hanqing Lu, Zhengyang Wang, Ruirui Li, Zhen Li, Monica Xiao Cheng, Rahul Goutam, Haiyang Zhang, Karthik Subbian, Suhang Wang, Yizhou Sun, Jiliang Tang, Bing Yin, Xianfeng Tang

Published: 25 Sept 2023, Last Modified: 02 Nov 2023 NeurIPS 2023 Datasets and Benchmarks Poster Everyone Revisions BibTeX

Keywords: session based recommendation, multilingual language models, recommender system, pre-training & fine-tuning, transfer learning, text generation

Abstract:

Modeling customer shopping intentions is a crucial task for e-commerce, as it directly impacts user experience and engagement. Thus, accurately understanding customer preferences is essential for providing personalized recommendations. Session-based recommendation, which utilizes customer session data to predict their next interaction, has become increasingly popular. However, existing session datasets have limitations in terms of item attributes, user diversity, and dataset scale. As a result, they cannot comprehensively capture the spectrum of user behaviors and preferences. To bridge this gap, we present the Amazon Multilingual Multi-locale Shopping Session Dataset, namely Amazon-M2. It is the first multilingual dataset consisting of millions of user sessions from six different locales, where the major languages of products are English, German, Japanese, French, Italian, and Spanish. Remarkably, the dataset can help us enhance personalization and understanding of user preferences, which can benefit various existing tasks as well as enable new tasks. To test the potential of the dataset, we introduce three tasks in this work: (1) next-product recommendation, (2) next-product recommendation with domain shifts, and (3) next-product title generation. With the above tasks, we benchmark a range of algorithms on our proposed dataset, drawing new insights for further research and practice. In addition, based on the proposed dataset and tasks, we hosted a competition in the KDD CUP 2023 https://www.aicrowd.com/challenges/amazon-kdd-cup-23-multilingual-recommendation-challenge and have attracted thousands of users and submissions. The winning solutions and the associated workshop can be accessed at our website~https://kddcup23.github.io/.

Integrating ML into Senior Design Projects

Discuss how ML can enhance various project domains (e.g., healthcare, finance, robotics).

Showcase exemplary senior projects that successfully integrated ML, highlighting challenges and solutions.

The phrase **"garbage in, garbage out" (GIGO)** in machine learning

- Data quality
- Data preprocessing
- Feature selection

Practical ML Tools and Frameworks

Provide an overview of essential ML libraries and frameworks (e.g., TensorFlow, PyTorch, Scikit-learn).

Offer guidance on selecting appropriate tools based on project needs.

- Assess the Project’s Goals and Objectives
- Evaluate Dataset Availability Early
  - **Suggested Tip**: "Explore existing datasets in the field as early as possible.
- Encourage Flexibility in Project Ideas

Ethical Considerations and Responsible AI

Address ethical considerations in ML projects, such as data privacy and algorithmic bias.

Discuss the importance of responsible AI and how to implement it in design projects.

Research Methodologies in ML

Outline common research methodologies used in ML projects, including data collection, model training, and evaluation.

1. What's the problem?
2. What's my proposed solution?
3. What compelling experiments can I run to demonstrate the effectiveness of my solution?

Offer tips on writing effective ML research proposals and reports.

Navigating ML Publications and Resources

Guide on finding and utilizing ML academic publications, preprints, and conferences for project inspiration and literature review.

Recommend databases, journals, and digital libraries for ML research.

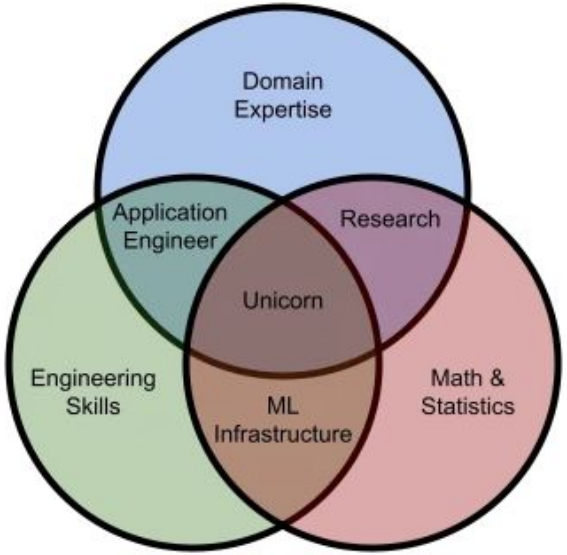
GW IT for Machine Learning

SEAS Careers for Machine Learning

Collaboration and Interdisciplinary Projects

Encourage interdisciplinary collaboration, highlighting how ML can be applied across different fields.

Suggest strategies for forming balanced teams with complementary skills.



Future Trends and Opportunities in ML

Discuss emerging trends in ML and potential future applications.

Highlight opportunities for further study, research, or career paths in ML.

Q&A and Project Brainstorming Session

Allocate time for students to ask questions and brainstorm project ideas with instructors and TAs.

Schedule your weekly meeting with a professor!

Instructions for Scheduling Weekly Project Meetings

Select Your Meeting Time: Choose a **30-minute time slot** for your weekly meeting

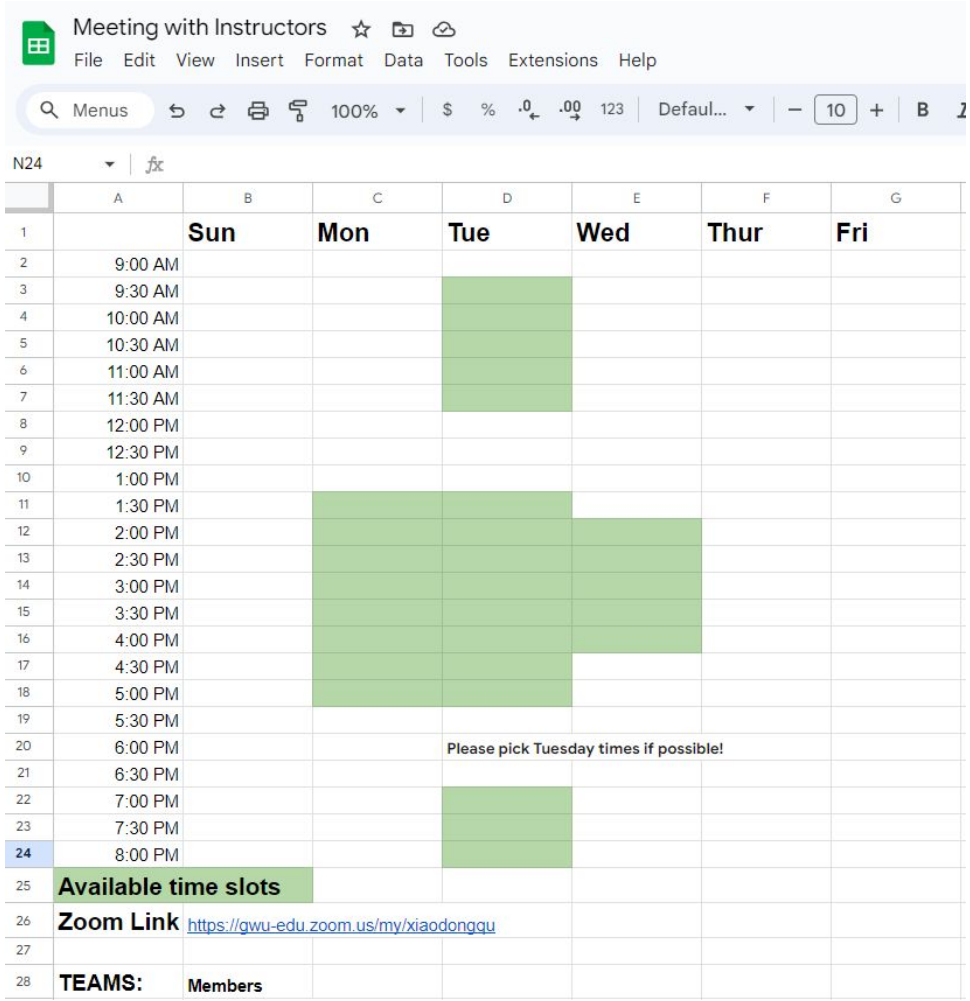
Navigate to the tab for your assigned instructor, Tim or Xiaodong
Find an available **green time slot** that works for all team members.
Enter your **team name** in the chosen time slot.

Important Notes:

Do not edit entries made by other teams.
If no available time slots work for your team, contact your instructor.

Recurring Meetings:

The selected time slot will be your recurring meeting time each week during the fall semester.
We may ask some teams to adjust their slots later to consolidate meetings.
Your flexibility is appreciated, and we will do our best to accommodate your schedules.



	A	B	C	D	E	F	G
1		**Sun**	**Mon**	**Tue**	**Wed**	**Thur**	**Fri**
2	9:00 AM						
3	9:30 AM						
4	10:00 AM						
5	10:30 AM						
6	11:00 AM						
7	11:30 AM						
8	12:00 PM						
9	12:30 PM						
10	1:00 PM						
11	1:30 PM						
12	2:00 PM						
13	2:30 PM						
14	3:00 PM						
15	3:30 PM						
16	4:00 PM						
17	4:30 PM						
18	5:00 PM						
19	5:30 PM						
20	6:00 PM			**Please pick Tuesday times if possible!**			
21	6:30 PM						
22	7:00 PM						
23	7:30 PM						
24	8:00 PM						
25	**Available time slots**						
26	**Zoom Link**	https://gwu-edu.zoom.us/my/xiaodongqu					
27							
28	**TEAMS:**	**Members**					

No more full-class lectures on Tuesdays

Starting next week, there will be **no more full-class lectures on Tuesdays**. Instead, each student team (4 members per team) will need to schedule a **30-minute Zoom meeting** with a professor to discuss the details of your senior design project.

Key Points:

No more **Tuesday lectures** with the full class.

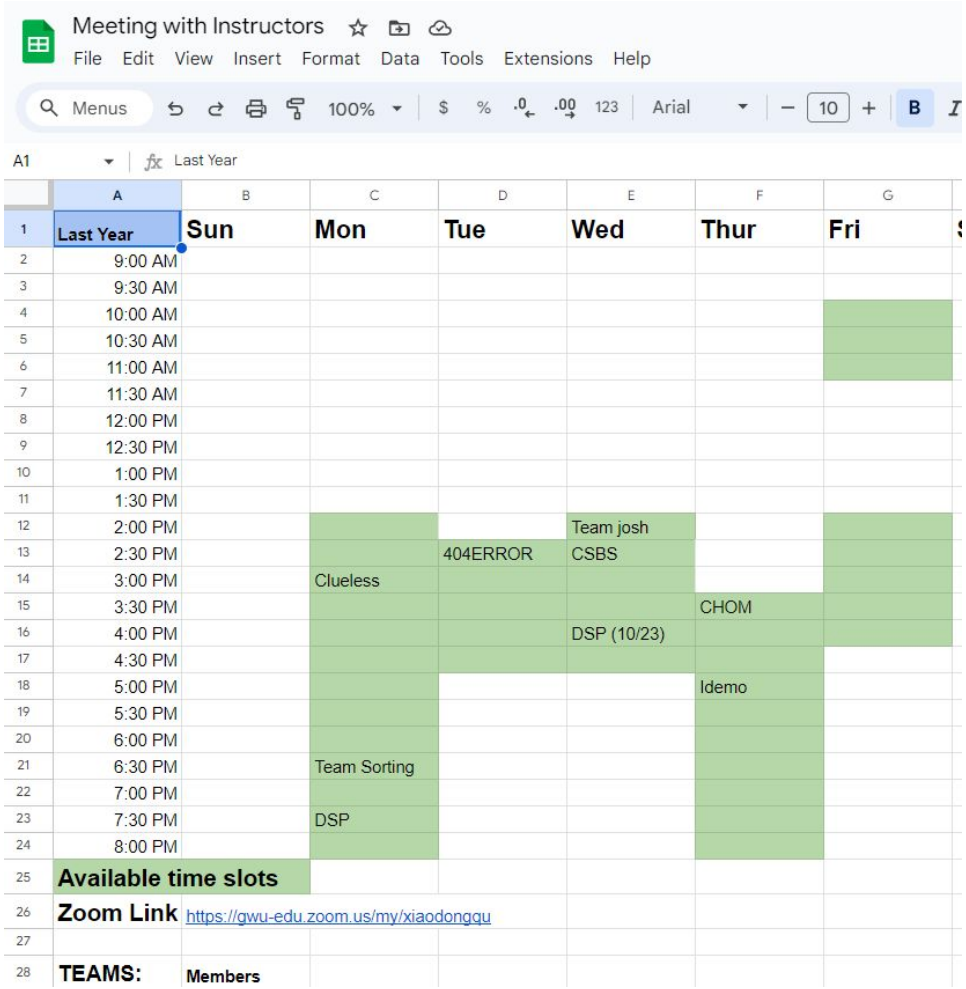
Each team (4 students) must schedule a 30-minute Zoom meeting with a professor.

Meetings will focus on discussing the specifics and progress of your senior design project.

Wednesday Lab Remains Unchanged: All students are still required to attend the regular Wednesday lab session.

Please coordinate with your team and **choose a time** that works best for everyone. **By Sunday 09/08/2024.**

We look forward to hearing about your projects in more detail!



	A	B	C	D	E	F	G
1	Last Year	Sun	Mon	Tue	Wed	Thur	Fri
2	9:00 AM						
3	9:30 AM						
4	10:00 AM						
5	10:30 AM						
6	11:00 AM						
7	11:30 AM						
8	12:00 PM						
9	12:30 PM						
10	1:00 PM						
11	1:30 PM						
12	2:00 PM				Team josh		
13	2:30 PM			404ERROR	CSBS		
14	3:00 PM		Clueless				
15	3:30 PM					CHOM	
16	4:00 PM				DSP (10/23)		
17	4:30 PM						
18	5:00 PM					Idemo	
19	5:30 PM						
20	6:00 PM						
21	6:30 PM		Team Sorting				
22	7:00 PM						
23	7:30 PM		DSP				
24	8:00 PM						
25	**Available time slots**						
26	**Zoom Link**	https://gwu-edu.zoom.us/my/xiaodongqu					
27							
28	**TEAMS:**	**Members**					