CS 4824/ECE 4424: Machine Learning

Website: https://people.cs.vt.edu/dbhattacharya/courses/cs4824/

Piazza: https://piazza.com/vt/spring2024/cs4824ece4424/home

Canvas: https://canvas.vt.edu/courses/185332 (CS 4824)
https://canvas.vt.edu/courses/185672 (ECE 4424)

Are you in the right place?

- **This is CS 4824/ECE 4424 (ML): CRN 13430**
  - Modality is "Face-to-Face Instruction"
- **There is another section of the same course…**
  - But with different CRN

Today

- **What is Machine Learning, the discipline, about?**
  - The *trifecta* of ML
- **What is this class about?**
  - What to expect?
  - Logistics

What are we here to discuss?

A wide range of topics
in
Machine Learning
covering their
intuitions, mathematical foundations, analyses, and applications!

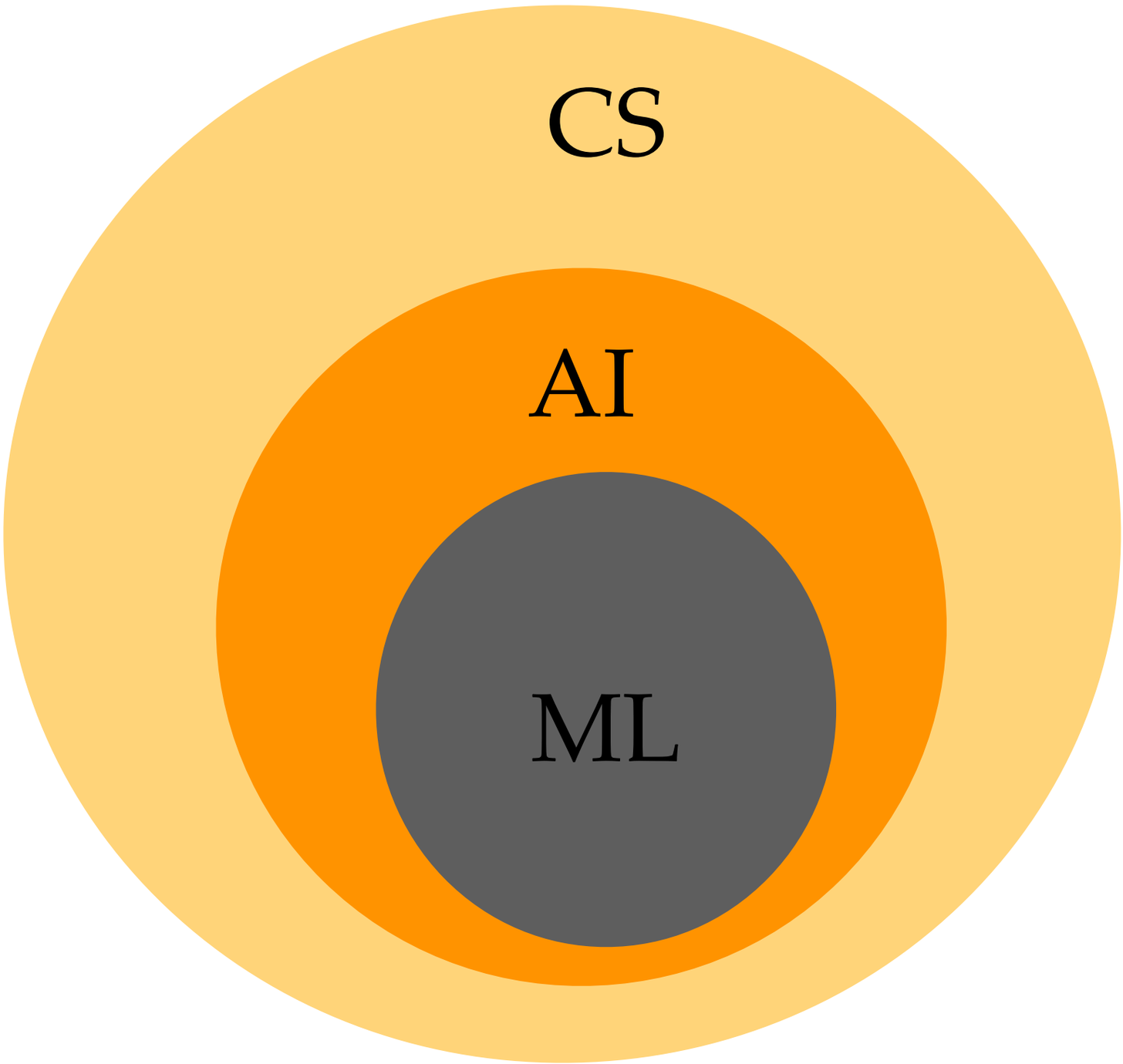
Demo time…

https://openai.com/blog/openai-codex/

Live demo:

https://youtu.be/SGUCcjHTmGY

Concepts



Intelligence

- **General Intelligence**
  - The ability to navigate in problem space.—Siddhartha Mukherjee, Columbia
- **Artificial Intelligence**
  - The science and engineering of making computers behave in ways that, until recently, we thought required human intelligence. —Andrew Moore, CMU

What is machine learning?

- **Study of algorithms that**
  - improve their performance (P)
  - at some task (T)
  - with experience (E)

Well-defined learning task: <P, T, E>

The ML niche is expanding…

- **Machine learning is central to**
  - Computer Vision
  - Natural Language Processing
  - Speech Recognition
  - Biology
  - Robotics
  - …

- **Machine learning has a conducive environment to grow in the 21st century**
  - Accumulation of "Big Data"
  - Demand for Self-customization to User
  - Post Moore's Law Computing
  - …

Spam classification

From: cheapsales@buystufffromme.com
To: ang@cs.stanford.edu
Subject: Buy now!

Deal of the week! Buy now!
Rolex w4tchs - $100
Med1cine (any kind) - $50
Also low cost M0rgages
available.

Spam

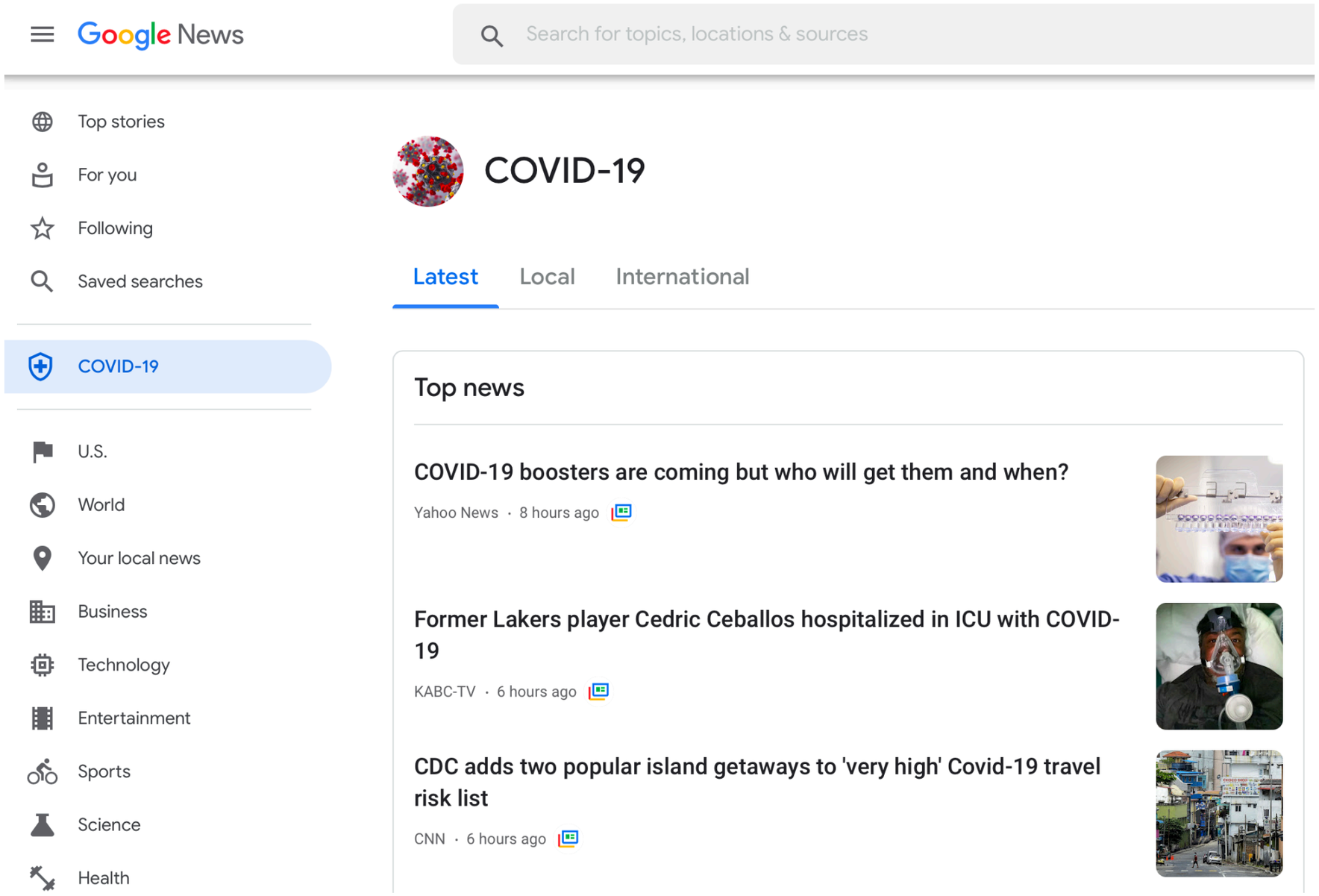
From: Alfred Ng
To: ang@cs.stanford.edu
Subject: Christmas dates?

Hey Andrew,
Was talking to Mom about plans
for Xmas. When do you get off
work. Meet Dec 22?
Alf

Non-spam

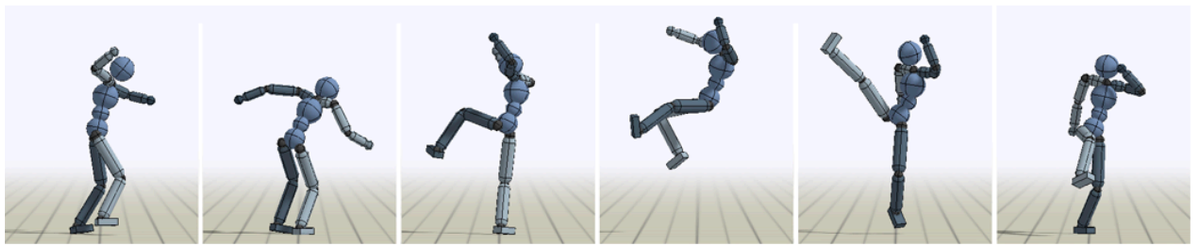
Supervised learning

News clustering

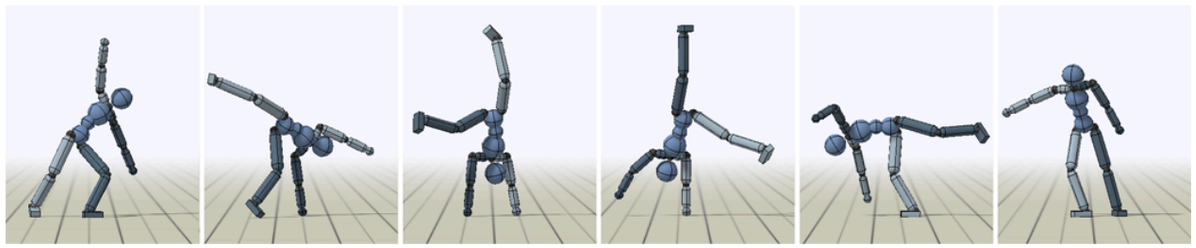


Unsupervised learning

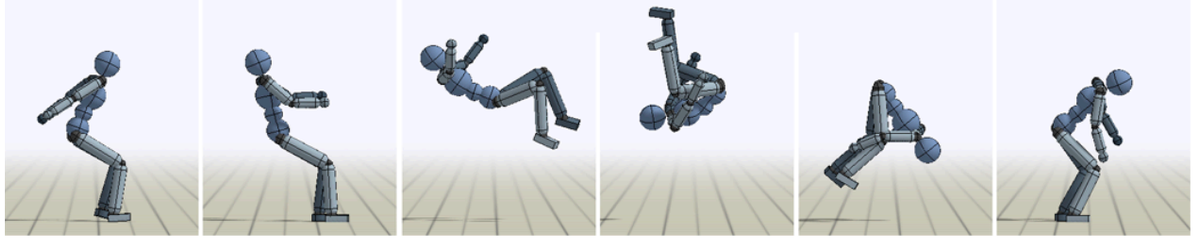
Motion planning



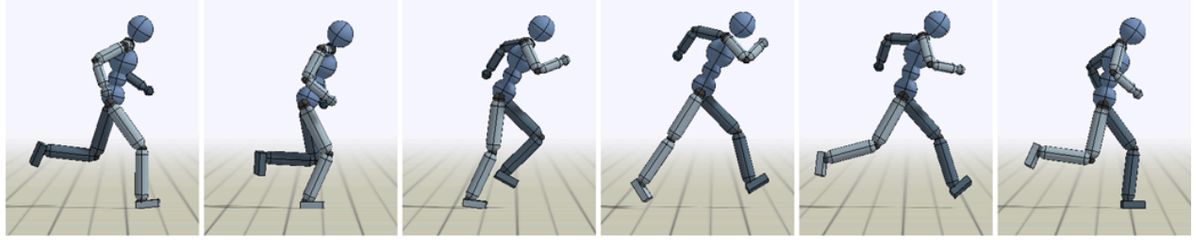
Spinkick



Cartwheel



Backflip



Run

Reinforcement learning

The *trifecta* of ML

- **Supervised learning**
  - Perform inductive inference and produce an inferred function to emulate mapping between input and output
  - Supervisory signal in the form of example input-output pairs (a.k.a. labeled training data)

- **Unsupervised learning**
  - Group together data into categories having similar features without requiring pre-assigned category labels
  - Self-discover any naturally occurring patterns in (unlabeled) data sets

- **Reinforcement learning**
  - Collect rewards for "good" actions while avoiding penalty for "bad" actions
  - Take action to explore uncharted territory and exploit experience

What this class is about?

- **Introduction to Machine Learning**

- **Goal**
  - After finishing this course, you should be ready to get started on your first ML research project

- **Target audience**
  - Senior undergrads, BS/MS students

What this class is NOT

- **NOT the target audience**
  - Grad students already working in ML area
  - Folks looking to understand the most recent breakthroughs (e.g., AlphaGo, AlphaFold, etc.)

- **NOT the goal**
  - Teaching a toolkit (e.g., TensorFlow/PyTorch)

Topics*

*Tentative and subject to change

- **Basics of Statistical Learning**
  - Loss functions, MLE, MAP, Bayesian estimation, bias-variance tradeoff, overfitting, regularization, cross-validation
- **Supervised Learning**
  - Decision Trees, Naïve Bayes, Logistic Regression, Linear Regression, Kernels and Kernel Regression, Support Vector Machines, Neural Networks
- **Unsupervised Learning**
  - EM, Clustering
- **Graphical Models**
  - Bayesian Networks, Hidden Markov Models
- **Deep Learning**
  - Convolutional Neural Networks, Recurrent Neural Networks, Attention and Transformer Networks, Autoencoders, Variational Autoencoders, Generative Adversarial Networks
- **Advanced Topics**
  - Generative AI, Diffusion Probabilistic Models

Textbook

- **None required.**

- **Optional reference books (freely available online):**
  - Machine Learning: a Probabilistic Perspective, Kevin Murphy, MIT Press, 2012
  - Pattern Recognition and Machine Learning, Christopher Bishop, Springer, 2006
  - The Elements of Statistical Learning, Trevor Hastie, Robert Tibshirani, and Jerome Friedman, Springer, 2009
  - Deep Learning, Ian Goodfellow, Yoshua Bengio, and Aaron Courville, MIT Press, 2016

Prerequisites

- **Probability and Statistics**
  - Distributions, densities, marginalization, moments, typical distributions.
  - Calculus and Linear Algebra
  - Matrix multiplication, eigenvalues, positive semi-definiteness, multivariate derivates.

- **Algorithms**
  - Dynamic programming, basic data structures, complexity.

- **Programming**
  - This is a demanding class in terms of programming skills. HWs will involve a mix of Python and libraries. You are free to choose any programming language for the project.

- **Ability to deal with abstract mathematical concepts.**

Organization and Deliverables

- **Homeworks**
  - Hands-on implementation and analysis, covering various topics that complement and supplement the lecture topics.
  - 4 individual HWs (**15% x 4 = 60%**). Involve a mix of Python and libraries.
  - 2 HWs before Spring break, 2 after Spring break.
  - Start early, Start early, Start early, Start early, Start early, Start early, Start early, Start early, Start early

- **Course Project**
  - The course project is meant for students to (1) gain experience implementing machine learning models; and (2) try machine learning on problems that interest them.
  - The project must be done individually in this semester (i.e., no double counting).
  - Project Proposal (**10%**): Due on **March 1** (i.e., before Spring break)
  - Project Final Report (**20%**): Due on **May 1** (i.e., on the last day of classes)
  - Project midway progress check (not graded, but submission is mandatory): Due on **April 24**

- **Class participation and Pop Quiz**
  - Contribute to class discussions on Piazza
  - Engage in class: ask questions, answer questions
  - In-class pop quizes (**10%**) requiring your class presence and overall engagement in the classroom.

- **Final letter grade**
  - Computed based on ceiling of the final percentage of points earned. The grade ranges are as follows:
    - A: 90%-100% B: 80%-89%. C: 70%-79%. D: 60%-69%. F: Below 60%

Late policy for deliverables

- **Late homework policy is as follows:**
  - Full credit when due.
  - Half credit next 48 hours.
  - Zero credit after that.

  Avoid invoking penalties by starting early and seeking help.
  No penalties for medical reasons or emergencies.

- **Late submissions are NOT allowed for the project proposal, midway or the final report.**

Project

- **Goal**
  - Chance to try Machine Learning
  - Encouraged to try out interesting applications of machine learning in various domains such as vision, NLP, speech, computational biology, etc.

- **Computing**
  - Google colab: jupyter-notebook + free GPU instance

Policies

- **Collaboration Policy**
  - All assignments are individual assignments
  - You may discuss the questions
  - Each student writes their own answer/code for all assignments
  - Project proposal/report will be plagiarism checked

- **Regarding Policy**
  - Within 1 week of receiving the grades

- **Academic integrity**
  - Students enrolled in this course are responsible for abiding by the Honor Code
  - Zero-tolerance philosophy regarding plagiarism or other forms of cheating

- **Principles of Community**
  - The course will include in-class discussions, and we will adhere to Virginia Tech Principles of Community.

- **Accessibility**
  - If any student needs special accommodations because of any disabilities, please contact the instructor during the first week of classes.
  - Such students are encouraged to work with The Office of Services for Students with Disabilities to help coordinate accessibility arrangements.

- **COVID-19 Policy**
  - Please follow the instructions posted at the University and public health guidelines for the latest COVID-19 Policy.

Course Team — TA

Weijie (Jack) Guan
Ph.D. Student
Computer Science
Virginia Tech

Email: skjguan@vt.edu

Research interest lies in Graph Machine Learning, Trustworthy AI, General Out-of-distribution Detection

Office Hours: Tuesday and Thursday 1:00 pm - 2:00 pm **McBryde Hall 106** or via Zoom https://virginiatech.zoom.us/skype/9248463068

Course Team — TA

Jianan Nie
Ph.D. Student
Computer Science
Virginia Tech

Email: jianan@vt.edu

Research interest lies in AI for Science (Chemistry), which includes autonomous discovery, reaction prediction, molecular design, and drug discovery.

Office Hours: Tuesday and Thursday 2:00 pm - 3:00 pm at Gilbert Place 4112 or via Zoom https://virginiatech.zoom.us/j/2649120190

Course Team

- **Instructor**
  - Debswapna Bhattacharya (Office Hours: Monday and Wednesday 1:00 pm - 2:00 pm at **Torgersen 3120B**)

- **TAs**
  - Weijie (Jack) Guan (Office Hours: Tuesday and Thursday 1:00 pm - 2:00 pm **McBryde Hall 106** or via Zoom https://virginiatech.zoom.us/skype/9248463068

  - Jianan Nie (Office Hours: Tuesday and Thursday 2:00 pm - 3:00 pm at Gilbert Place 4112 or via Zoom https://virginiatech.zoom.us/j/2649120190)

How to stay in touch?

- **Primary means of communication —Piazza**
  - No direct email to instructor unless private information
  - Instructor/TAs can provide answers to everyone on forum
  - Extra credit for regularly (and correctly) answering questions on forum!
  - No posting answers on forum. We will monitor.
  - If you have a personal matter, create a private piazza post or send an email to the course staff.

- **Staff Mailing List**
  - cs-4824-ece-4424-s24-staff-g@vt.edu

Todo: before next class

Go through the course webpage at:
https://people.cs.vt.edu/dbhattacharya/courses/cs4824/

…and ask any questions in the next class.

Get into Piazza:
https://piazza.com/vt/spring2024/cs4824ece4424