



University of Colorado
Boulder

CSCI 6502

**Big Data Analytics: Systems,
Algorithms, and Applications**

Spring 2021
Lecture 01 (Jan 14)

Agenda

- ◆ Introduction: Instructor, TA, class
- ◆ Course overview
- ◆ Administrative information
- ◆ Policies
- ◆ Summary



About the Instructor (I)

- ◆ Qin (Christine) Lv
- ◆ Associate Professor
- ◆ Co-Associate Chair for Graduate Education
- ◆ Department of Computer Science
- ◆ Contact information
 - ◆ office: ECCR 1B24, phone: (303) 492 - 8821
 - ◆ email: Qin.Lv@Colorado.EDU
 - ◆ <https://www.cs.colorado.edu/~lv>



About the Instructor (2)

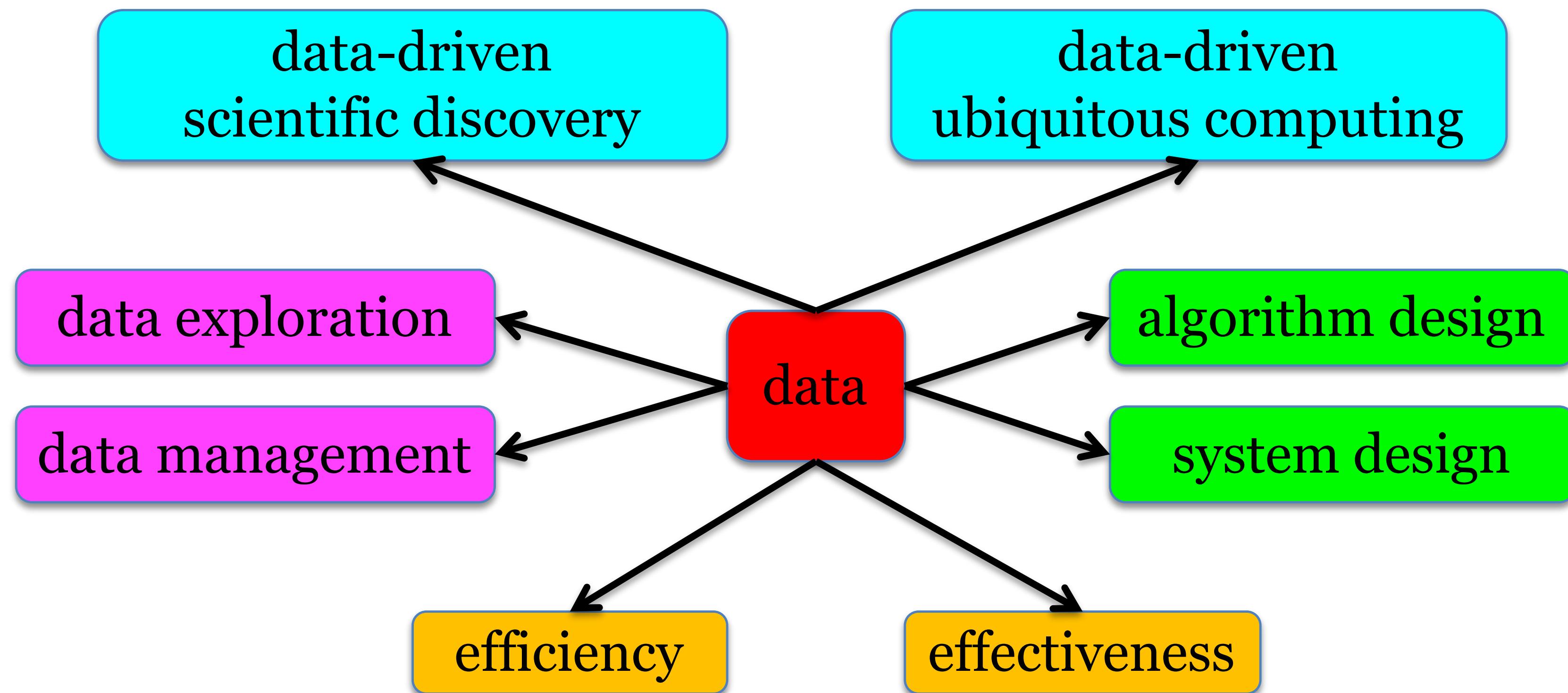
- ◆ **B.E. Tsinghua University**
 - ◆ computer systems, wireless networks
- ◆ **Ph.D. Princeton University**
 - ◆ peer-to-peer networks, distributed storage systems, scalable similarly search
- ◆ **University of Colorado Boulder**
 - ◆ data-driven scientific discovery & ubiquitous computing
 - ◆ full-stack data analytics



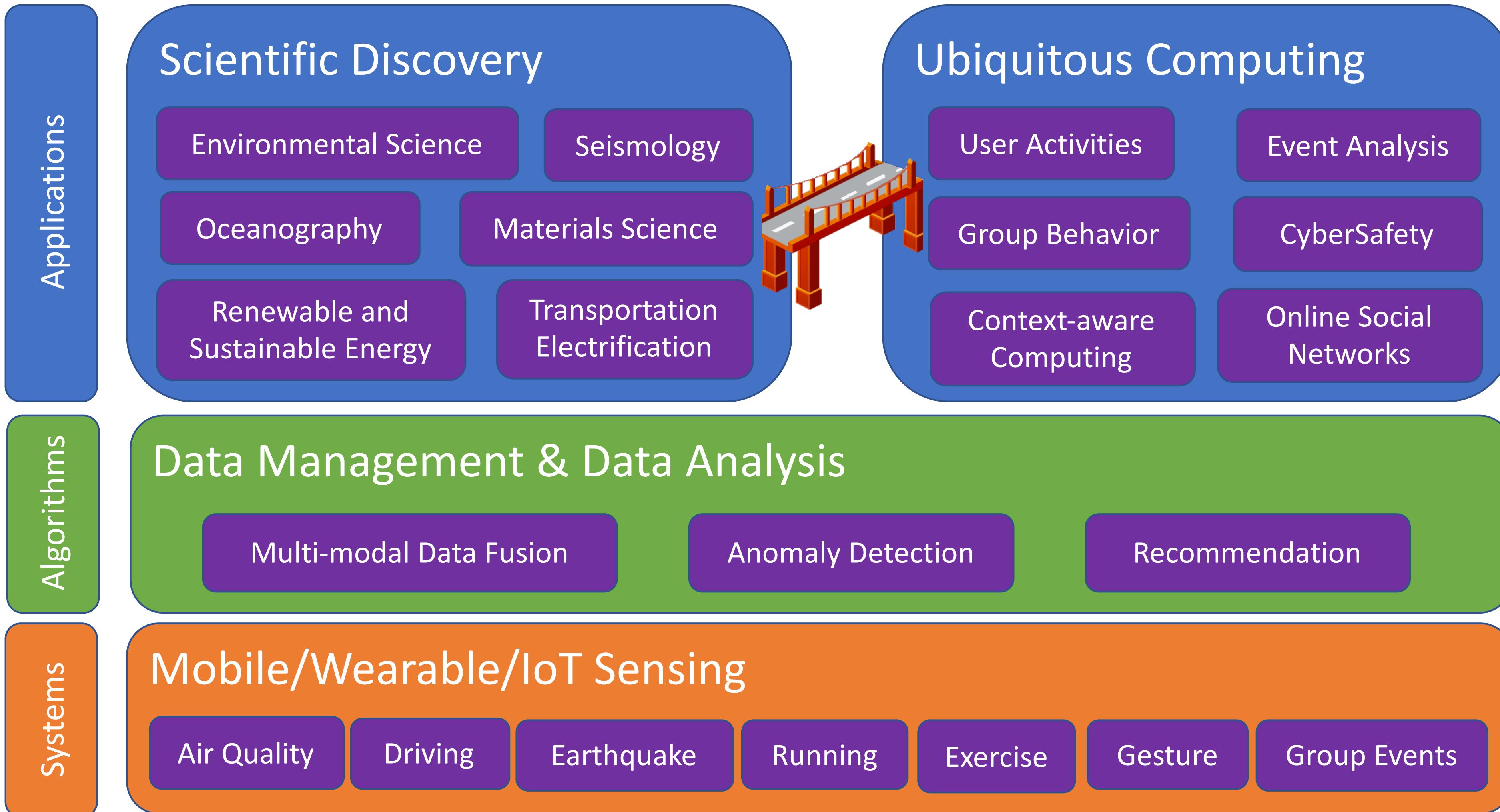
About the Instructor (3)

Research Overview

interdisciplinary



About the Instructor (4)



About the Instructor (5)

- ◆ Sample research projects
 - ◆ air quality sensing & data analysis
 - ◆ remote sensing, ocean sensing, earthquake early warning
 - ◆ PHEV, solar/wind farm, transportation electrification
 - ◆ Twitter, Reddit: event analysis, user location, NBA fans
 - ◆ recommendation: group, social, stability
 - ◆ cybersafety: flashers, cyberbullying, IRA
 - ◆ ...



About the TA

- ◆ Rahul Aedula
- ◆ PhD student
- ◆ Department of Computer Science
- ◆ Contact information
- ◆ rahul.Aedula@colorado.edu



About the Class

- ◆ CSCI 6502-001 / CSCI 6502-001B
- ◆ Master's? PhD (advisor)?
- ◆ CS? ECEE? Other?
- ◆ C/C++? Java? Python? R? MATLAB?
- ◆ Hadoop? Cassandra? MongoDB? Spark? TensorFlow?
- ◆ AWS? GCP? Azure? Other?
- ◆ Data-related background & interests



Agenda

- ◆ Introduction: Instructor, TA, class
- ◆ Course overview
- ◆ Administrative information
- ◆ Policies
- ◆ Summary



Motivations

- ◆ Massive amounts of digital data (Big Data)
 - ◆ the 3Vs, 4Vs, 5Vs
 - ◆ volume, variety, velocity, veracity, value
- ◆ Two key tasks
 - ◆ data management, data exploration
- ◆ Three integral components
 - ◆ systems, algorithms, applications
- ◆ State-of-the-art research & practices



Course Overview

- ◆ **Systems**
 - ◆ cloud computing, GFS, MapReduce, NoSQL, stream processing, mobile/wearable/IoT sensing, deep learning platforms, ...
- ◆ **Algorithms**
 - ◆ sketching, indexing, large graph analysis, recommendation, deep learning algorithms, ...
- ◆ **Applications**
 - ◆ online social networks, health informatics, business intelligence, urban computing, scientific discovery, ...



Agenda

- ◆ Introduction: Instructor, TA, class
- ◆ Course overview
- ◆ Administrative information
- ◆ Policies
- ◆ Summary



Administrative Information (I)

- ◆ Big Data Analytics: Systems, Algorithms, and Applications
 - ◆ CSCI 6502-001 and -001B, Spring 2021
 - ◆ hybrid in-person/remote/online
-
- ◆ Class time and location
 - ◆ TuTh, 12:45pm -- 2:00pm
 - ◆ <https://cuboulder.zoom.us/j/99541523104>
 - ◆ ECCS IBI4 (after Feb 15) — will send out a survey for in-person learning preference and assignment



Administrative Information (2)

- ◆ Course website
- ◆ <https://canvas.colorado.edu>
- ◆ lecture slides: posted after each class
- ◆ recordings: posted within a couple of hours after each class
- ◆ reading materials
- ◆ all course submissions
- ◆ general course discussion, private messaging
- ◆ **FOR COURSE USE ONLY, DO NOT DISTRIBUTE**



Administrative Information (3)

- ◆ **Textbook**
 - ◆ no required textbook
 - ◆ a list of reading materials (some require CU access on campus or via VPN, e.g., ACM digital library)
- ◆ **Office hours**
 - ◆ Instructor: Tu 3:00pm-4:00pm or by appointment
 - ◆ TA: Fr 11:00am-12:00pm or by appointment
 - ◆ via class Zoom



Course Schedule (tentative)

- ◆ Week I (1/14) Class introduction
- ◆ Week 2-6 Systems
- ◆ Week 7 (2/23, 2/25) Project Proposal
- ◆ Week 8-11 Algorithms
- ◆ Week 11 (3/25) Wellness Day
- ◆ Week 12 (3/30, 4/1) Project Checkpoint
- ◆ Week 13-15 Applications
- ◆ Week 16 (4/27, 4/29) Final Project Report



Agenda

- ◆ Introduction: Instructor, TA, class
- ◆ Course overview
- ◆ Administrative information
- ◆ Policies
- ◆ Summary



Policies

- ◆ Check course syllabus in Canvas
- ◆ Academic integrity
 - ◆ Honor Code Pledge
- ◆ Classroom behavior
- ◆ Disability
- ◆ Discrimination and harassment
- ◆ Religious observance
- ◆ COVID-19 compliance



Academic Integrity

- ◆ **WORK ALONE**, unless specified explicitly as a group assignment
- ◆ All submitted work should include the [Honor Code Pledge](#)
- ◆ Properly acknowledge other people's work, including information you find on the Web
- ◆ Cheating or plagiarism will not be tolerated!



Grading

- ◆ Class participation (10%)
- ◆ Paper presentation (10%)
 - ◆ I presentation, work as a pair
- ◆ Paper review (40%)
 - ◆ 8 paper reviews, work alone
- ◆ Course project (40%)
 - ◆ work alone or as a group, scope appropriate for group size
- ◆ No midterm exam or final exam



Class Participation (10%)

- ◆ Class attendance
- ◆ Finish reading assignment before class
- ◆ Participate in in-class discussions
 - ◆ listen to others
 - ◆ ask questions
 - ◆ answer questions
 - ◆ tell us your thoughts
- ◆ Opt out for online-only students. Please confirm with TA.



Paper Presentation (10%)

- ◆ One presentation, work in pairs
- ◆ Pick your topic from the reading list, or suggest new topic
- ◆ 10-minute presentation + discussion
 - ◆ Research paper: problem, key contributions, evaluation, future work
 - ◆ Tool: key capabilities, use cases
- ◆ Email slides to instructor one day before class



Paper Review (40%)

- ◆ 8 paper reviews per student, work alone
- ◆ Pick the papers from the reading list
 - ◆ not covered in your presentation
- ◆ Due one day before corresponding class
- ◆ Submit in Canvas in PDF format
- ◆ Include paper title, authors, affiliations, venue
- ◆ Include your name, ID, Honor Code Pledge



Paper Review (40%)

- ◆ How to read a research paper?
 - ◆ title, authors, affiliations, publication venue
 - ◆ abstract, introduction
 - ◆ related work
 - ◆ problem formulation, proposed work
 - ◆ evaluation
 - ◆ discussion, conclusion, future work



Paper Review (40%)

- ◆ What may be contained in a paper review?
- ◆ What is the paper about?
- ◆ What are the strengths and weaknesses?
- ◆ Are the evaluations convincing?
- ◆ Can you improve their technique?
- ◆ Can the technique be applied to other domains/problems?
- ◆ Any other observations or questions?



Course Project (40%)

- ◆ A self-contained project related to topics covered in this course
- ◆ Work alone or as a group
 - ◆ project scope appropriate for group size
- ◆ Pick your own project
 - ◆ related to your own research interests
- ◆ Discuss with instructor and other students
- ◆ Start early!



Project Proposal (Week 7)

- ◆ A 5-10 minute presentation
 - ◆ motivation, literature survey
 - ◆ your approach
 - ◆ how to evaluate
 - ◆ milestones
- ◆ Submit presentation slides &
 - ◆ project proposal (~3 pages)



Project Checkpoint (Week 12)

- ◆ A 5-10 minute presentation
 - ◆ proposal review: motivation, your technique, evaluation, milestones
 - ◆ what you have accomplished so far
 - ◆ what remains to be done
- ◆ Submit presentation slides &
 - ◆ report highlighting progress (~6 pages)
 - ◆ updated version of your initial proposal



Project Final (Week 16)

- ◆ A 10-15 minute presentation
 - ◆ motivation, related work, your techniques, evaluation, conclusions, future work
- ◆ Submit presentation slides &
 - ◆ final project report (ACM format, title, abstract, ..., references, 10-12 pages)
 - ◆ related source code and results



Agenda

- ◆ Introduction: Instructor, TA, class
- ◆ Course overview
- ◆ Administrative information
- ◆ Policies
- ◆ Summary



Todos

- ◆ Familiarize yourself with our Canvas course
- ◆ Check the “Modules” page regularly for updates
- ◆ Finish corresponding reading materials before each lecture
- ◆ Pick papers to present or review
- ◆ Suggest other topics or reading materials
- ◆ Questions?

