

## Instructor & TAs



Zsolt Kira

## CS 8803 VLM Vision-Language Foundation Models

[Info](#)      [Schedule](#)

[Logistics](#)      [Materials](#)

[Request a Permit](#)

## Course Information

It is an exciting time in AI: We now have models that are able to perform some level of language comprehension and reasoning (even passing some human-level tests), and separately perceive the world through images or audio. Some combinations of these have led to cool applications such as language-conditioned image generation. Yet, if we are to have a general intelligent agent, it is important to integrate information across these modalities in a seamless manner. Since many of the advances have been driven by the Transformer architectures, which can take any arbitrary input as long as it can be tokenized, many in the community have begun to develop multi-modal models such as GPT4o and Gemini.

This course will cover the foundations of these models, covering some of the fundamentals (Transformers, etc.), Multi-Modal Models (LLaVA, BLIP, etc.), as well as emerging areas including Open-Vocabulary Perception, Vision-Language Reasoning, Multi-Modal Decision-Making Agents, and Embodied AI. A key component is reading and synthesizing papers in this area, both seminal papers as well as current state-of-art. Since this a rapidly evolving field, it will be a seminar-style course, with a focus on reading and discussing these techniques. The other important component of the course will be the project, where students can dive into specific areas of interest and advance upon the state of art.

# Course Outcomes

This course will provide students with knowledge and experience with the history and state of art of vision-language (and in general multi-modal) Foundation Models, and students will be able to:

- Explain various design decisions in terms of architectures, training, and data involved in developing VLMs.
- Differentiate between various classes of multi-modal models
- Design and/or use multi-modal to solve novel problems

## Recommended Background

We will assume that you already have some background in machine learning, deep learning/neural networks, and potentially some of the modalities (e.g. computer vision, natural language processing, etc.). We expect you to be familiar with state of art architectures such as Transformers, and to a lesser degree Recurrent Neural Networks (RNNs)/Long- Short Term Memory (LSTMs), and Convolutional Neural Networks (CNNs). If you need to brush up on these, we recommend the following resources:

- Guest Lecture from CS7643 Deep Learning on Transformers
- Justin Johnson's notes on Convolutional Neural Networks and Transformers
- CS231n: Convolutional Neural Networks for Visual Recognition
- CS224n: Natural Language Processing with Deep Learning

Note that we will not expect that you have deep experience in all of the modalities - it is OK if you are strong in one modality and weaker in another. We will also assume that you are comfortable in executing an interesting project in this area.

## Required Materials

There is no required textbook or materials not freely available online.

## Grading Components

See the materials section for details and templates for these deliverables.

- **15% Class Participation**
  - Attendance and engagement during our class discussion, posting questions and answers on Ed, and adding materials/resources for the papers discussed.
- 20% Paper Reviews
- 15% Paper Discussion
- 50% Project
  - 5% Proposal
  - 10% Midterm Presentation
  - 15% Final Presentation & Video
  - 20% Final Report

-

[Info](#)      [Schedule](#)

[Logistics](#)      [Materials](#)

[Request a Permit](#)

## Grading Scale

There will not be a curve for this course, the standard grading scale will be used (we may round depending on the scores):

### Grade Percentage Range

A	90% and above
B	80% - 89%
C	70% - 79%
D	60% - 69%
F	Below 60%

## Course Materials

All materials will be provided on this website. The slides for any lectures (in the first few sessions) will be provided right before class, and papers discussed will be linked in the schedule. There is no book for this course. Logistics to be determined. The grading will

include paper reviews and presentations, participation, and project (which will include various aspects including the project proposal, report, presentation/video, etc.). Participation will be required given that the course is largely discussion-driven.

## Communication Policy

You are responsible for knowing the following information:

1. Anything posted to this syllabus.
2. Anything emailed directly to you by the teaching team (including announcements via Ed Discussion), 24 hours after receiving such an email or post.

Because Ed announcements are emailed to you as well, you need only to check your Georgia Tech email once every 24 hours to remain up to date on new information during the semester. Georgia Tech generally recommends students to check their Georgia Tech email once every 24 hours. So, if an announcement or message is time sensitive, you will not be responsible for the contents of the announcement until 24 hours after it has been sent.

## Late and Make-up Work Policy

There will be no make-up work provided for missed assignments. Of course, emergencies (illness, family emergencies) will happen. In those instances, please [contact the Dean of Students office](#). Let us know as soon as possible (do not send us personal/medical information!), and if you ask for accommodations or late submission as a result notify us before the due date. The Dean of Students is equipped to verify emergencies and pass confirmation on to all your classes. For consistency, we ask all students to do this in the event of an emergency. Do not send any personal/medical information to the instructor or TAs; all such information should go through the Dean of Students.

## Online Student Conduct and (N)etiquette

Communicating appropriately in the online classroom can be challenging. It is especially important for a discussion-based course. All communication, whether by email, Ed,

Canvas, or otherwise, must be professional and respectful. In order to minimize this challenge, it is important to remember several points of **internet etiquette** that will smooth communication for both students and instructors.

1. Read first, Write later. Read the ENTIRE set of posts/comments on a discussion board before posting your reply, to prevent repeating commentary or asking questions that have already been answered.
2. Avoid language that may come across as strong or offensive. Language can be easily misinterpreted in written electronic communication. Review email and discussion board posts BEFORE submitting. Humor and sarcasm may be easily misinterpreted by your reader(s). Try to be as matter of fact and as professional as possible.
3. Follow the language rules of the Internet. Do not write using all capital letters, because it will appear as shouting. Also, the use of emoticons can be helpful when used to convey nonverbal feelings. 😊
4. Consider the privacy of others. Ask permission prior to giving out a classmate's email address or other information.
5. Keep attachments small. If it is necessary to send pictures, change the size to an acceptable 250kb or less (one free, web-based tool to try is [picsize.com](http://picsize.com)).
6. No inappropriate material. Do not forward virus warnings, chain letters, jokes, etc. to classmates or instructors. The sharing of pornographic material is forbidden.

**NOTE:** The instructor reserves the right to remove posts that are not collegial in nature and/or do not meet the Online Student Conduct and Etiquette guidelines listed above.

## Plagiarism & Academic Integrity

Georgia Tech aims to cultivate a community based on trust, academic integrity, and honor. Students are expected to act according to the highest ethical standards. All students enrolled at Georgia Tech, and all its campuses, are to perform their academic work according to standards set by faculty members, departments, schools and colleges of the university; and cheating and plagiarism constitute fraudulent misrepresentation for which no credit can be given and for which appropriate sanctions are warranted and will be

applied. For information on Georgia Tech's Academic Honor Code, please visit <http://www.catalog.gatech.edu/policies/honor-code/> or <http://www.catalog.gatech.edu/rules/18/>.

For this class, the following policy will be in effect:

- **Paper Reviews, Presentations, Ed Posts/Discussions:** All materials submitted for these deliverables must be entirely your own. It is OK to discuss the papers with others, of course, as discussions and collaborations are an important part of science. However, all submitted materials (text, code, pseudo-code, figures, etc.) must be your own or (if you're using snippets for illustration) directly quoted/cited from the original sources. Of course, we expect you to not just copy-paste from the papers, but to synthesize the information in your own words. You cannot use AI assistants for ANY part of the reviews.
- **Projects:** It is OK to discuss the projects with others (of course including your team) and you are free to use whatever online codebases, blogs, resources, and AI coding assistants (e.g. Copilot) if you wish. Explicitly acknowledge any and all resources used (including who you discussed with), including AI assistants, and for the latter specifically describe how they were used (which parts, whether you prompted with your own materials or something else, etc.). Separate from this, the proposed problem and/or approach should be novel and your own, all experimental implementation and running of experiments should be your own and your group's. All presentations, reports, and videos should be completely and wholly your own and/or your group's and you may NOT use AI assistants to generate any part of these deliverables.

We will actively check for cheating, and any act of dishonesty will result in a Fail grade. Any student suspected of cheating or plagiarizing on any deliverable will be reported to the Office of Student Integrity, who will investigate the incident and identify the appropriate penalty for violations.

## **Illness, Disability, Mental Health Resources and Support Services**

### **Students with Disabilities**

If you are a student with learning needs that require special accommodation, contact the Office of Disability Services at 404.894.2563 or <http://disabilityservices.gatech.edu/>, as soon as possible, to make an appointment to discuss your special needs and to obtain an accommodations letter. Please also e-mail me as soon as possible to set up a time to discuss your learning needs.

## Illness and Other Ailments

If you are a student that is negatively impacted by a health-related matter, please contact the Office of Disability Services or the Office of Dean of Students at 404.894.6367 or [studentlife@studentlife.gatech.edu](mailto:studentlife@studentlife.gatech.edu). Do NOT send us any personal health information. They will provide you with an accommodation letter that will allow us to try to find a suitable schedule for completing all assignments. You MUST submit this and inform us that you did so on Ed before the due date for the deliverable.

## Campus Resources

### Georgia Tech Police Department

Emergency: Call 911 |  
404-894-2500

### Dean of Students Office

404-894-2565 |  
[studentlife.gatech.edu](http://studentlife.gatech.edu)  
Afterhours Assistance Line &  
Dean on Call: 404-894-2204

### Center for Assessment, Referral and Education (CARE)

404-894-3498 | [care.gatech.edu](http://care.gatech.edu)

### Collegiate Recovery Program

404-894-2575 |  
[counseling.gatech.edu](http://counseling.gatech.edu)

### Counseling Center

404-894-2575 |  
[counseling.gatech.edu](http://counseling.gatech.edu)

### Health Initiatives

404-894-9980  
[healthinitiatives.gatech.edu](http://healthinitiatives.gatech.edu)

### LGBTQIA Resource Center

404-385-4780 |  
[lgbtqia.gatech.edu](http://lgbtqia.gatech.edu)

### Stamps Psychiatry Center

404-894-1420

### VOICE

404-385-4464 |  
404-385-4451  
24/7 Info Line:  
404-894-9000 |  
[voice.gatech.edu](http://voice.gatech.edu)

### Women's Resource Center

404-385-0230 |  
[womenscenter.gatech.edu](http://womenscenter.gatech.edu)

### Veterans Resource Center

404-894-4953 |  
[veterans.gatech.edu](http://veterans.gatech.edu)

## Community Resources

### **Georgia Crisis and Access Line**

1-800-715-4225

The crisis line is staffed with professional social workers and counselors 24 hours per day, every day, to assist those with urgent and emergency needs.

### **Trevor Project**

1-866-488-7386

Trained counselors are available to support anyone in need.

### **National Suicide Prevention Hotline**

1-800-273-8255

A national network of local crisis centers that provides free and confidential emotional support to people in suicidal crisis or emotional distress 24/7.

### **Georgia State Psychology Clinic**

404-413-2500

The clinic offers high quality and affordable psychological services to adults, children, adolescents, families and couples from the greater Atlanta area.

## Student-Faculty Expectations Agreement

At Georgia Tech we believe that it is important to strive for an atmosphere of mutual respect, acknowledgement, and responsibility between faculty members and the student body. See <http://www.catalog.gatech.edu/rules/22/> for an articulation of some basic expectation that you can have of me and that I have of you. In the end, simple respect for knowledge, hard work, and cordial interactions will help build the environment we seek. Therefore, I encourage you to remain committed to the ideals of Georgia Tech while in this class.

## Subject to Change Statement

The syllabus and course schedule may be subject to change. Changes will be communicated via the Ed announcement tool. It is the responsibility of students to check Ed Discussions, email messages, and course announcements to stay current in their online courses.

Credits: This course was designed with materials and inspiration from great courses such as Learning from Limited Labels and Internet Data Science