

CS 411 System for Artificial Intelligence

Fall 2023

Instructor: Jongse Park

Time: TU/TH 14:30 – 16:00

Credits: 3-0-3

Room: E3-1 #1101

Course description & goals

Welcome everyone! We are all living in an insanely exciting time of artificial intelligence (AI). ChatGPT is there for you to ask any questions, the future of autonomous driving is just around the corner, and AR/VR technologies are soon going to permanently change every type of your social interaction experience. While AI algorithms are always the primary driving force behind these remarkable technical innovations, they cannot be practically utilized without efficient and performant “systems” to process them. In this course, we will delve into the fascinating world of building these AI systems, also known as Sys4AI. Get ready for an exciting journey ahead!

While you are likely to be familiar with algorithmic foundations of AI, we will still go over the basic math first, mostly focusing on their computational aspects. We will then talk about the software components of Sys4AI, which constitute middleware, runtime, compiler, and many others. Lastly, we will discuss various hardware acceleration techniques (a.k.a., AI chips), aiming for both inference and training.

This course will be heavily project-oriented. All projects will be team-based and your team will have three projects: (1) software project, (2) hardware project, and (3) team-defined final project. In software project, you will build a tiny ML framework, similar to PyTorch, learning and experiencing highly-parallel programming techniques on CPU and GPU. In hardware project, your team will be assigned with a FPGA board and will develop a very simple AI accelerator on it, using a hardware description language such as Verilog. These hands-on projects will give you invaluable experience that could be very useful for your future careers. Finally, leveraging the skillsets and knowledge acquired from the aforementioned two projects, your team will have the opportunity to define and develop your own final project, which can be exclusively software-based (SW-only), hardware-based (HW-only), or a co-designed integration of both software and hardware (SW-HW). We will have a series of presentation sessions where your team will present your project outcome in front of everyone in the class.

You will also be asked to submit critiques for selected papers in the field of AI systems. (1) All critiques should start with a short summary of the papers. (2) Then, you must list up their strengths and weaknesses. (3) Then, you must provide the discussion points derived from your critical thinking. Critical thinking is a cognitive process that involves analyzing and evaluating information, arguments, and ideas in a systematic and logical manner. It is the ability to objectively assess and question assumptions, recognize biases, consider alternative perspectives, and make reasoned judgments based on evidence and reasoning. You must have two or more discussion points (the more, the better). These discussion points are the most important parts of your critiques to get high scores, while summary is the least.

The details of lectures and projects are tentative and subject to be updated but the main theme will be more and less the same.

Communication

Discussion: The class discussion board is on Piazza. You may sign up at this link:

<https://piazza.com/kaist.ac.kr/fall2023/cs411>

I will post course-related announcements and information on the board. You must read the discussion board at least once per day, and you should post course-related questions and responses there. I expect you to make good use of the discussion group and of TA support when you have technical or administrative questions or problems. You are responsible for any and all information posted to Piazza. We will pin any important announcements to the top of the Piazza feed. You are expected to read all announcements within twelve hours of their being posted.

Email to the Instructor or TAs: Emails to course staff should begin with "CS411:" in the subject line, followed by a brief description of the purpose of your email. If you follow this rule, we will be better able to address your questions in a timely manner. If you do not, a response to your email may be delayed indefinitely.

Prerequisites

This class has three prerequisites: CS230, CS311, plus one AI course (e.g., CS376, CS470, AI501, EE331). While not meeting these prerequisites will not result in your removal from the class, it will certainly pose challenges when working on projects. It is important to note that this course is project-oriented, and failing to complete any of the three projects will have undesirable consequences. Before registering for this course, please thoroughly assess your technical capabilities to prevent unnecessary tension during the grade submission period.

Rubric

This course will include projects, critiques, and class participation.

Projects: 50%

Paper Critiques: 30%

Class participation: 20%

Cheating

Instructor and TAs will perform several forms of cheating detections under the hood.

Examples of cheating include: looking at other team's program, writing your program while talking to other team members about it, talking another team members through the solution code, allowing others to look at your team's solution code, and looking on the Internet for code to do your project, including stackoverflow and other such sites. Code generation tools, including ChatGPT, are strictly prohibited. If you have any doubts about what is allowed, ask the instructor.

Students who violate University rules on scholastic dishonesty are subject to disciplinary penalties, including the possibility of failure in the course and/or dismissal from the University. Because such dishonesty harms the individual, all students, and the integrity of the University, policies on scholastic dishonesty will be strictly enforced. **The penalty for cheating on projects and critiques is F.**