Christof Teuscher

ECE 410/510: Hardware for AI and ML

Week 7: Recap, outlook, and reminders

Portland State University
Department of Electrical and Computer Engineering (ECE)
www.teuscher-lab.com
teuscher@pdx.edu

teuscher Lab
teuscher-lab.com

Portland State UNIVERSITY

Recap

- How can we accelerate an algorithm?
- What are the most promising ways to do so today?
- What are the biggest bottlenecks in today's traditional computing architectures?
- What is a non-von Neumann architecture?
- What is in-memory computation.
- What is one way to realize in-memory computation?

What did you learn last week?

Reminder

Mid-term assessment

The goal of the mid-term assessment is for you and me to assess your knowledge about the topics of this course.

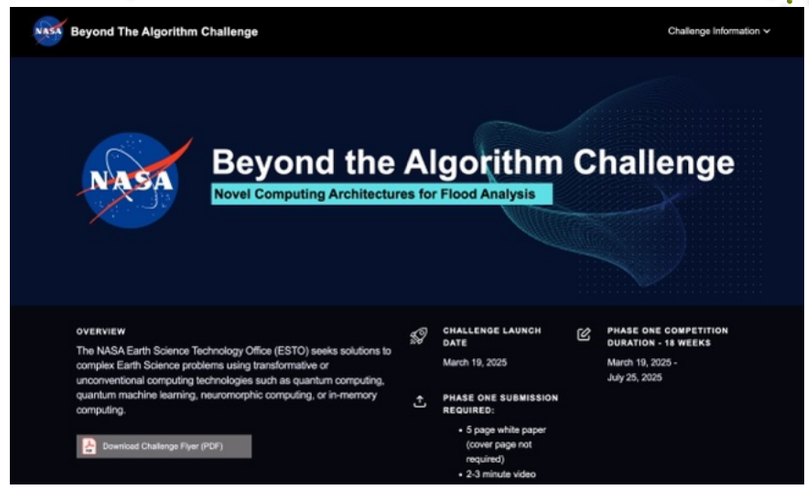
If you've been in class, paid attention, asked questions, and took notes, you should be able to answer these questions in 30min without going back to the slides.

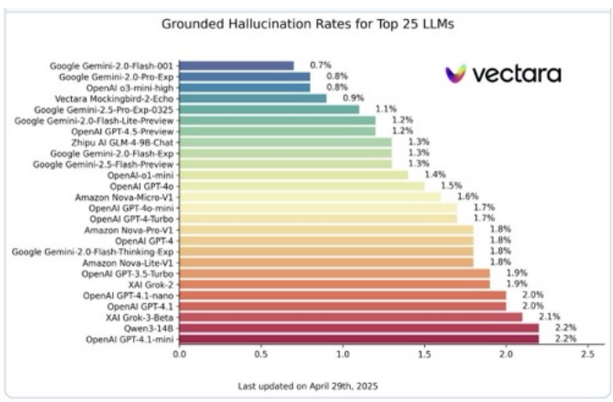
There is a total of 26 multiple-choice questions.

30:00 Time Limit **5/14/2025 11:59 PM** Due in 2 days

Week	Monday	Wednesday (Codefest)
2	HW/AI/ML overview + codesign overview	Start main project: pick workload, start analysis, benchmark, …
3	GPU architecture and programming for AI	Drafting a HW architecture, creating a model
4	Deep neural networks on GPUs	Coding HW description
5	Transformers on GPUs	First simulation + refinement
6	In-memory computation	Improving initial design
7	Neuromorphic chips: TrueNorth, Loihi, Akida	Simulation + refinement
8	Neuromorphic computing with mem-devices	Synthesizing design + benchmarking
9	Hardware accelerators for embedded systems	Final improvements
10	Emerging technologies and future directions	Final tests, validation, verification, benchmarking

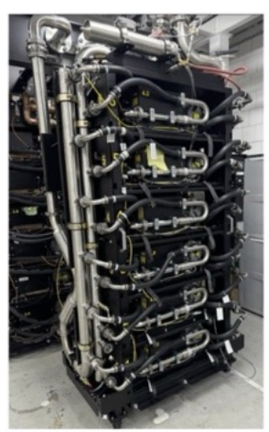
Memorial Day





Hallucination leaderboard: https://github.com/vectara/hallucination-leaderboard

"The dramatic increase in chip power consumption — from 100W chips to accelerators exceeding 1000W — has made advanced thermal management essential."

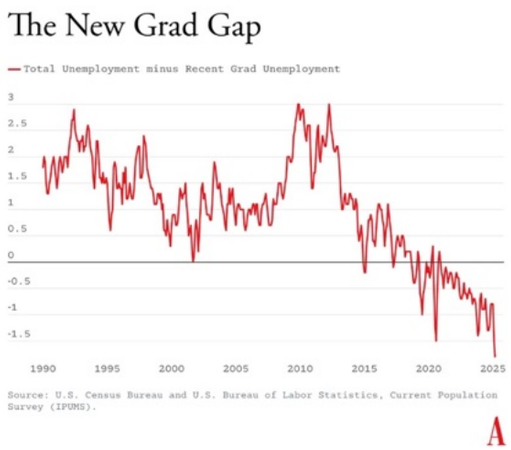


"We are excited to introduce +/-400 VDC power delivery that can support up to 1 MW per rack."

~1,200 homes!

https://cloud.google.com/blog/topics/systems/enabling-1-mw-it-racks-and-liquid-cooling-at-ocp-emea-summit

AI can do low-level work that you as an entry level grad without job experience would do.



Leaving gaps in your basic knowledge is a key ingredient in the recipe for professional failure.

"Our recommendations emphasize iterative interaction with ChatGPT and independent verification of its outputs. Considering the importance of utilizing ChatGPT judiciously and with expertise, we recommend its usage for experts who are well-versed in the respective domains."

- "The comparative analysis shows that the ChatGPT-assisted group significantly improved code quality, with fewer rule violations and reduced cyclomatic and cognitive complexities."
- "These findings suggest that ChatGPT can be beneficial in programming education by helping students write cleaner, less complex code and adhere to coding conventions."