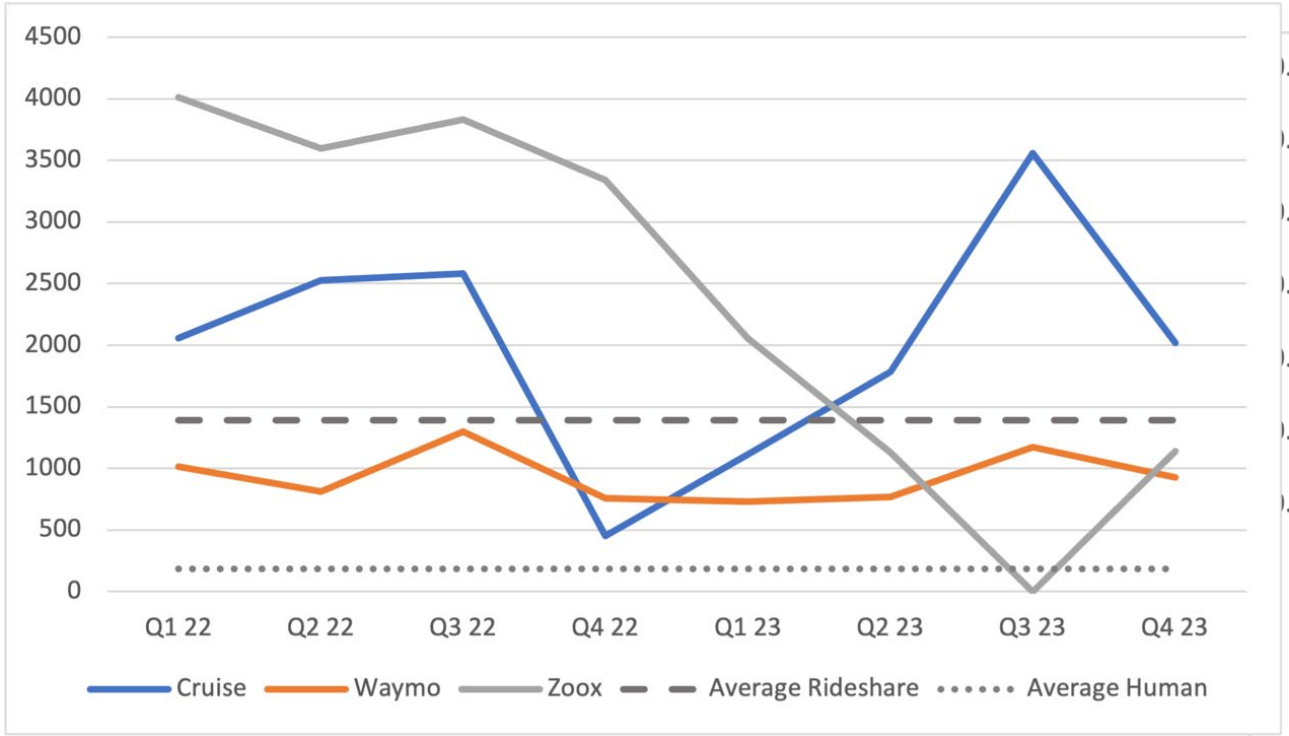
SYSTEMS ENGINEERING & SAFETY-CRITICAL AI: A REALITY CHECK

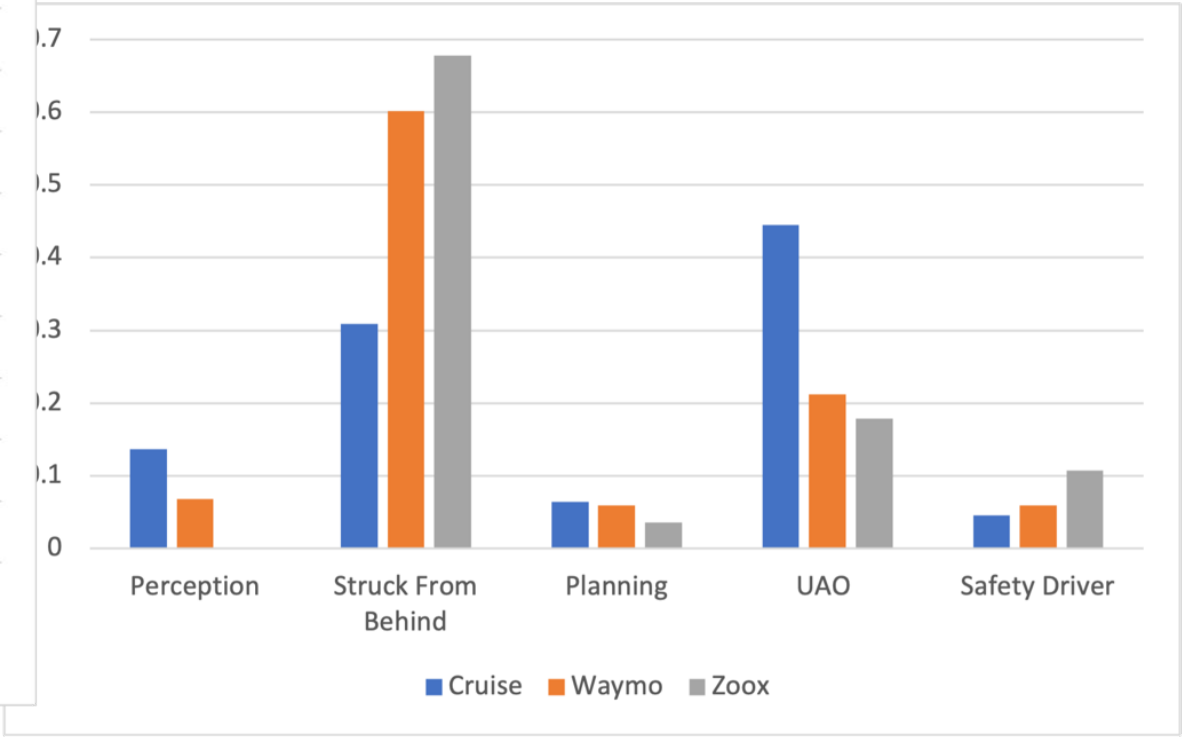
Missy Cummings, PhD

George Mason University

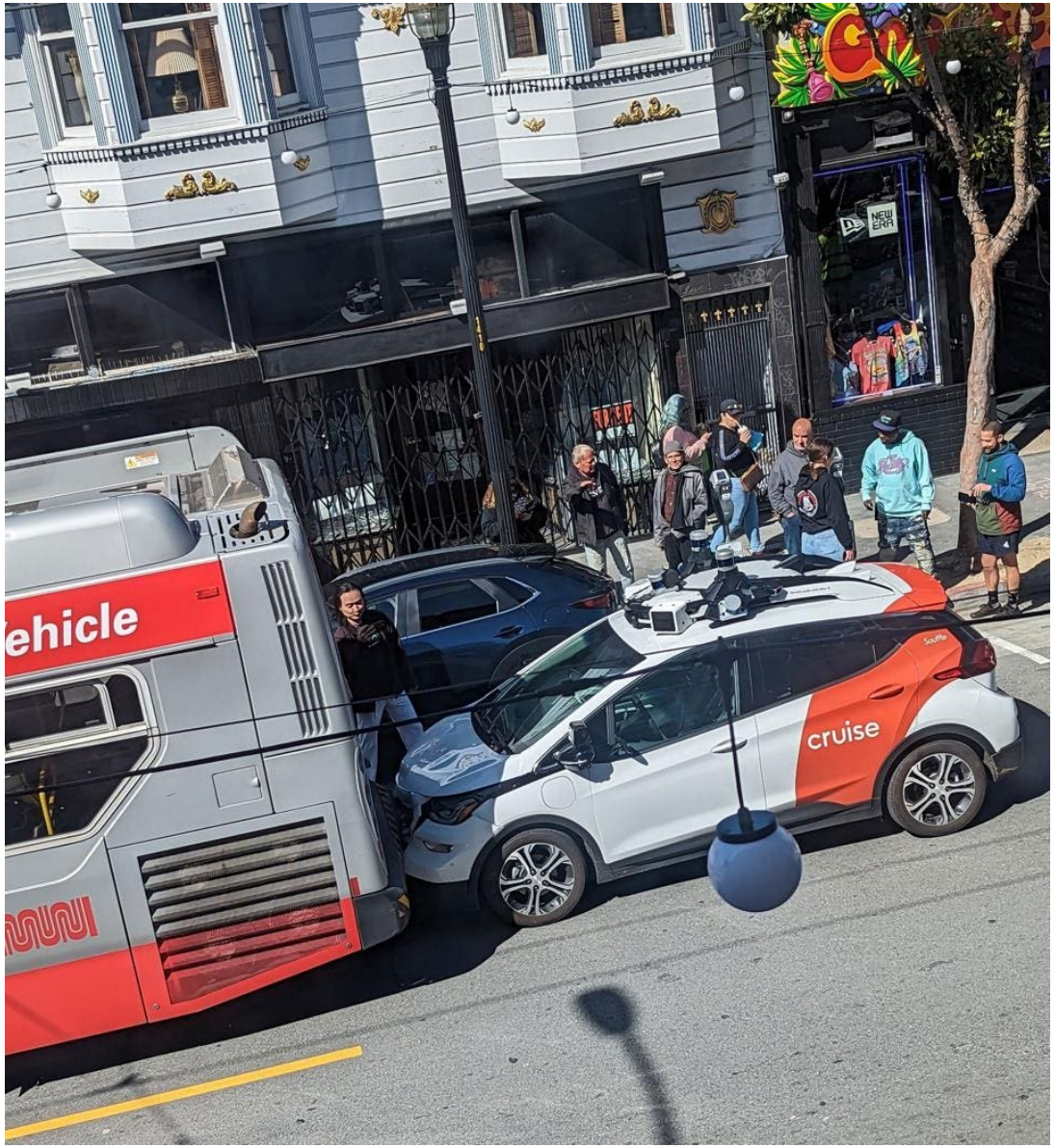
Just how well are self-driving cars performing?

- 2022-2023 CA crash data

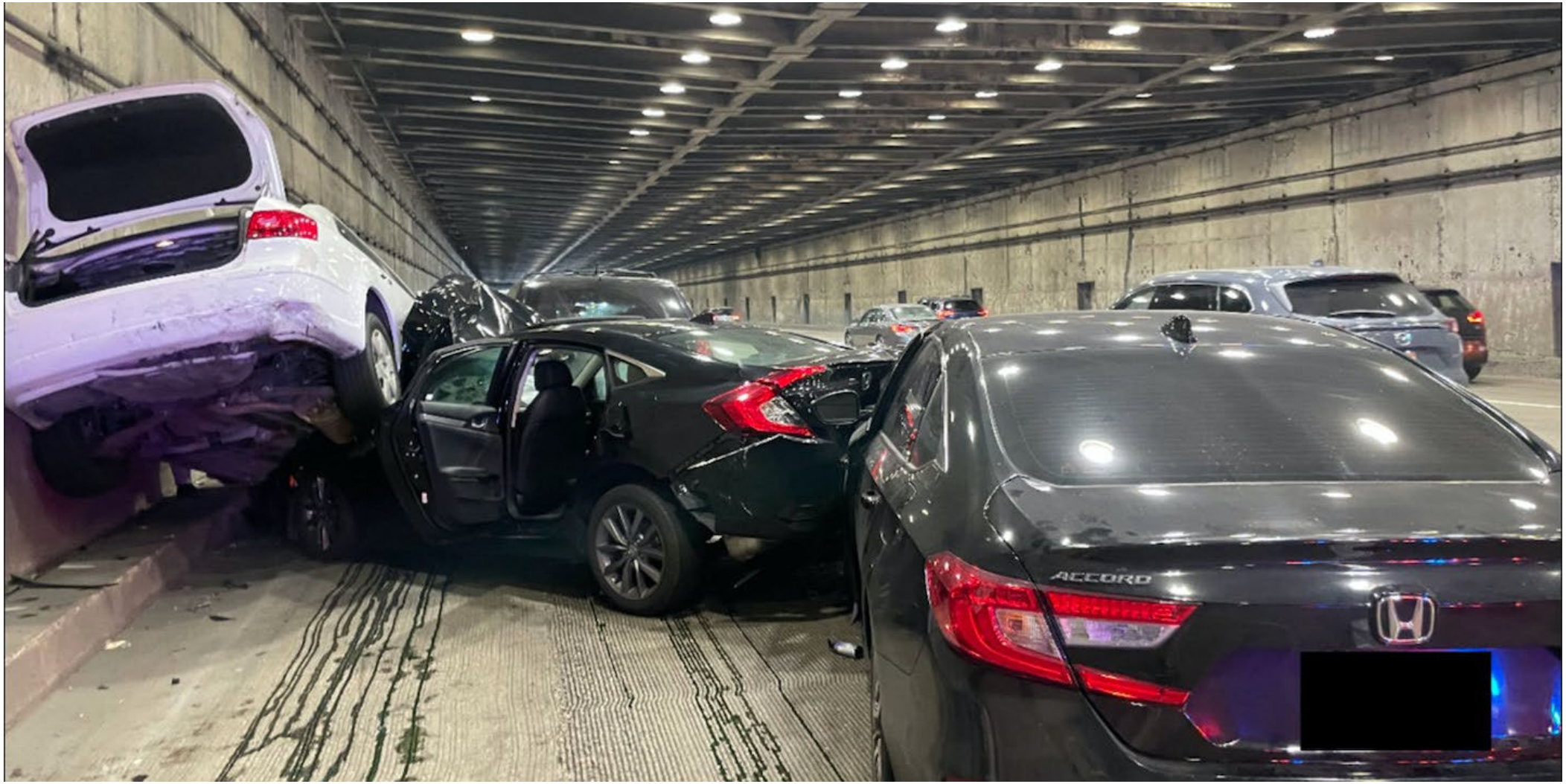




Perception

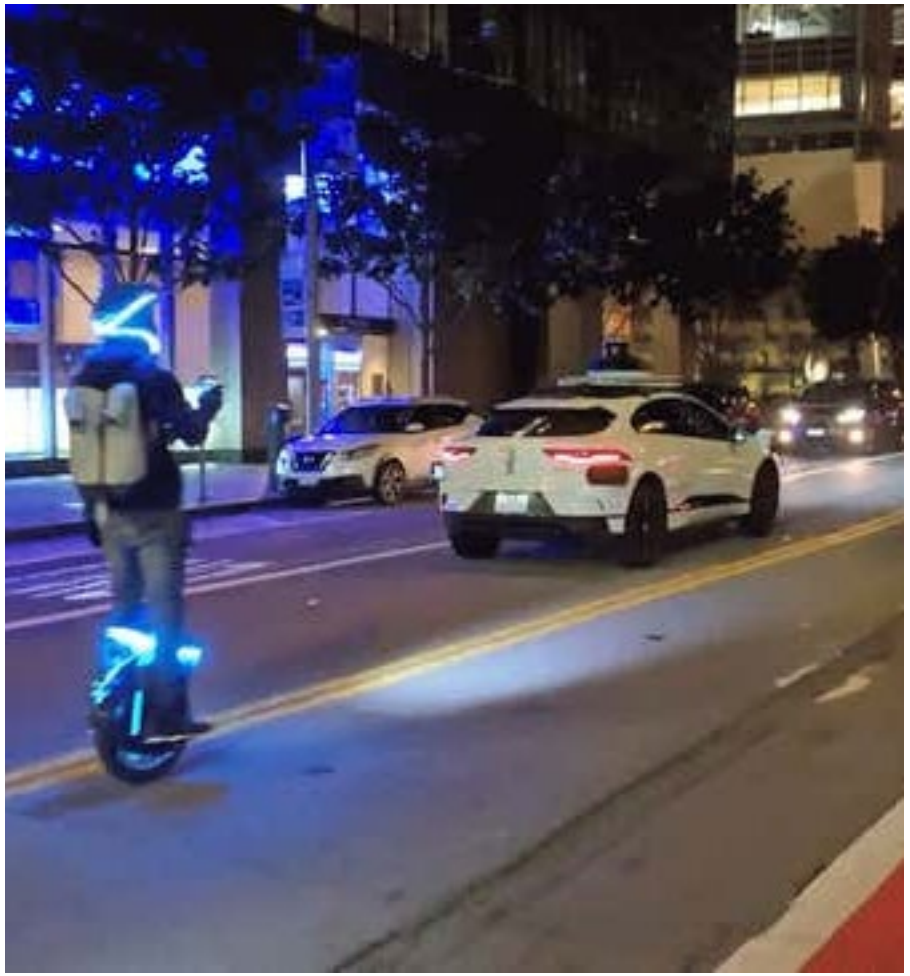


Struck-from-Behind



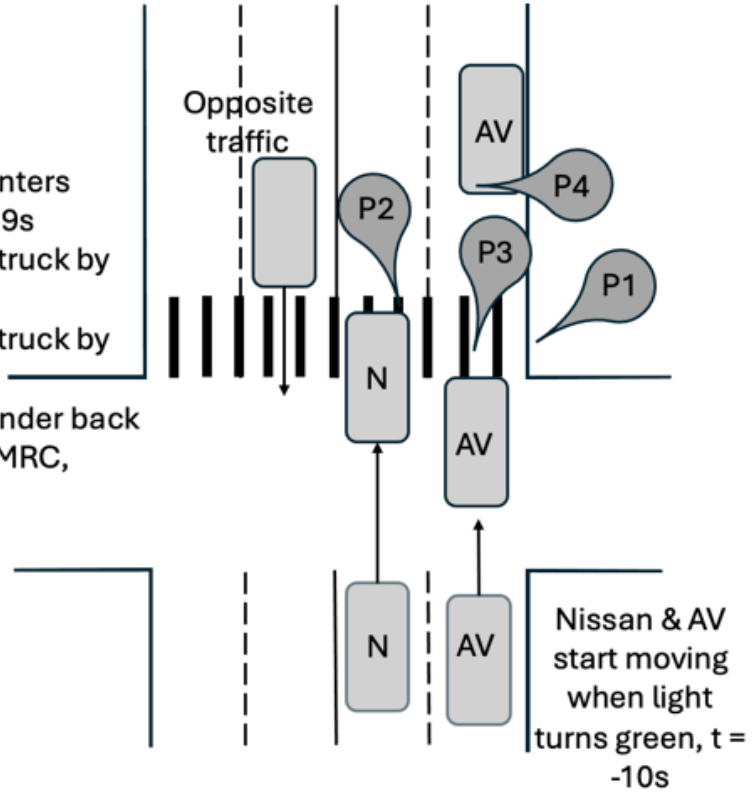
Planning


Autonomous Waymo
goes the wrong way
Autonomous Waymo
goes the wrong way

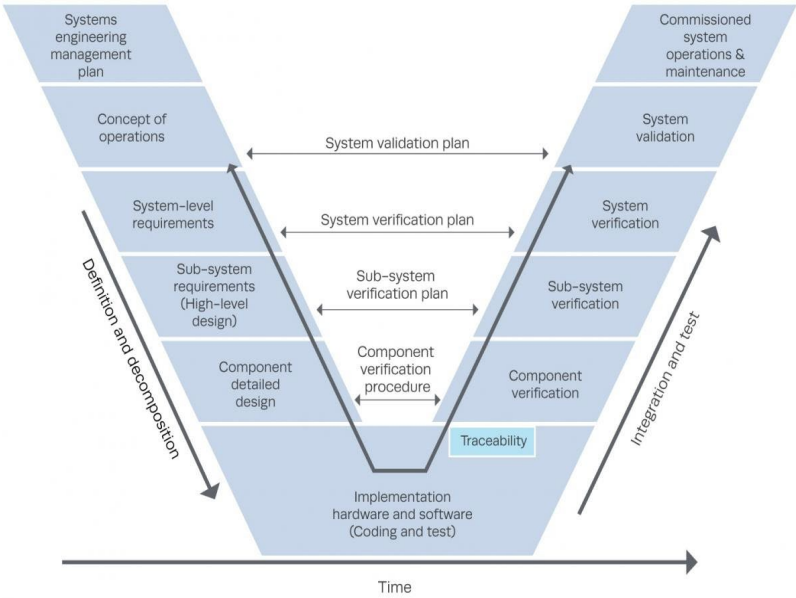


Unexpected Actions by Others

- P1: Pedestrian enters cross walk, t=-7.9s
- P2: Pedestrian struck by Nissan, t=-2.9s
- P3: Pedestrian struck by AV, t=0s
- P4: Pedestrian under back left tire after AV MRC, t=8.8s

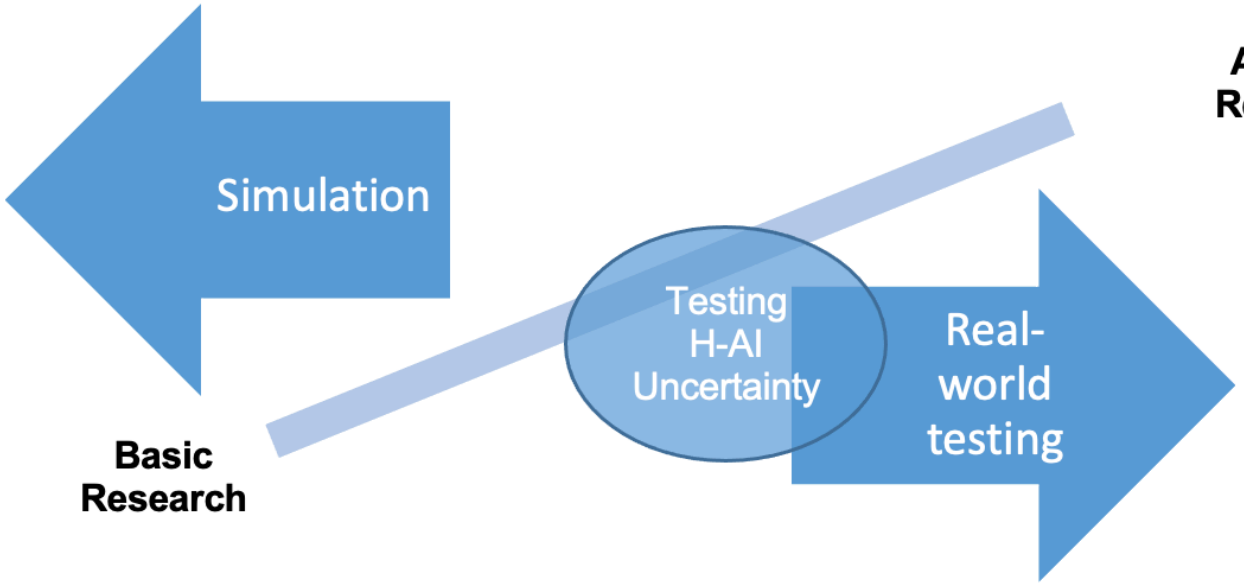


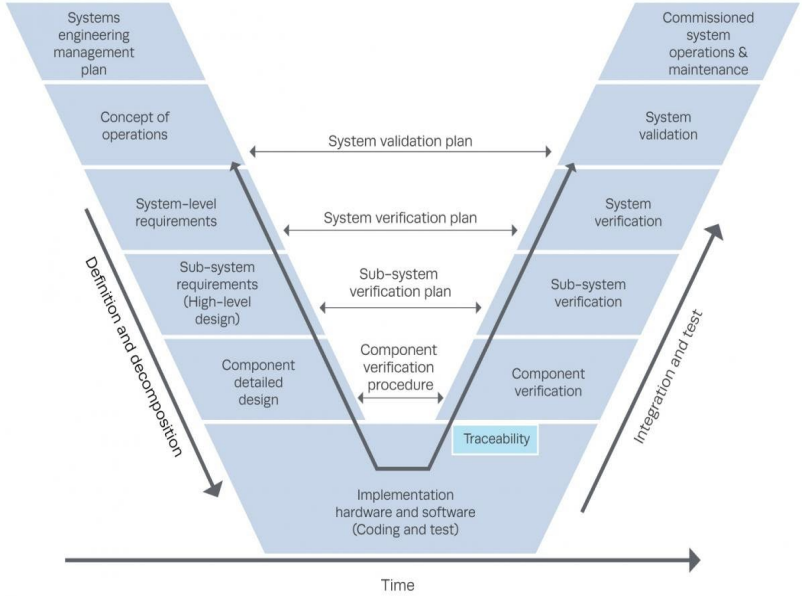
GOFSE is missing

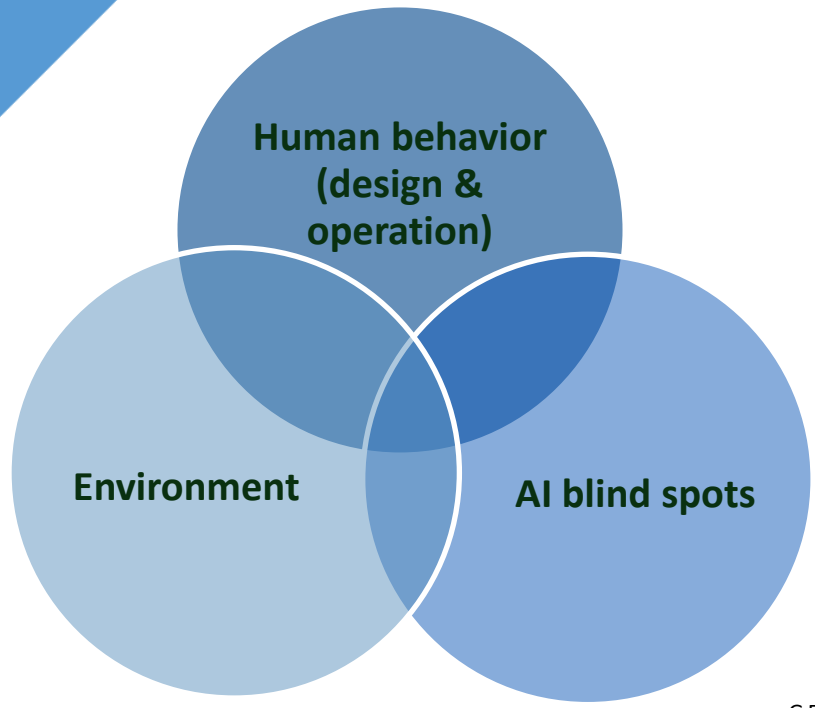


"ility"	Needs
Traditional	
Usability	• AI operational limitations and competency boundaries should be made transparent to users. • In appropriate settings, user should be able to conduct sensitivity analyses to explore a decision space, as well as the limitations. • Routine feedback about usability should be elicited from users, including post-software updates.
Suitability	• A process should be implemented that maps any operational dependencies created in the implementation of AI systems in order to determine what downstream processes could be negatively affected if an AI system is degraded or fails.
Sustainability	• A process for identifying changes in operations or environmental conditions that affect model outcomes should be implemented, including when retraining should occur for connectionist AI systems. • An incident repository should be created and routinely analyzed for all AI systems where users and supervisors can document erroneous, unusual and unexpected system behaviors. • A process for tracking software changes and possible unintended impact on either operations or human activity should be developed. • A process for tracking and documenting issues with concept drift as well as operator disuse, misuse or abuse of AI should be implemented.
New	
Auditability	• Data and resulting models should be periodically audited to uncover issues with suitability and sustainability, as well as possible issues with bias. • Automated tools will be needed to support humans conducting auditing tasks.
Passive vulnerabilities	• Adversarial machine learning vulnerabilities need to be identified and mitigated.

Design & Testing Implications







Where to go from here

- Formalizing SE in AI systems, especially safety-critical ones
- Testing & certification needs significantly more effort
  - Drawing a line for simulation validation
- Workforce development
  - Responsible AI
  - Designate Chief AI Risk/Safety/Teaming Officer
  - AI maintenance and risk management are key
- Generative AI
  - Never in safety-critical systems

Responsible AI

- AI Fundamentals
- Engineering AI Systems
- Risk
- AI Ethics

Questions?

https://nap.nationalacademies.org/catalog/26355/human-ai-teaming-state-of-the-art-and-research-needs

https://www.frontiersin.org/journals/neuroergonomics/articles/10.3389/fnrgo.2023.1102165/full