Curriculum Vitae/Resume
Zahra Zahedi
https://zahrazahedi28.github.io/

Summary
A PhD in Computer Science, motivated and solution-oriented with a variety of professional and research experience. I am working on developing computational models and frameworks for trust and understanding human cognitive states to enhance human-AI interaction. I have extensive research experience in diverse areas such as human-AI interaction, human-robot interaction, reinforcement learning, machine learning, behavior modeling of humans using data, probabilistic modeling, user studies and user-centered research, automated planning, decision-making frameworks, game theory.

Education

Arizona State University, Tempe, AZ.
- **PhD** student in Computer Engineering. (GPA: 4/4)　　　　　　2018-2023
  - ○ **Supervisor:** Dr. Subbarao Kambhampati
  - ○ **Dissertation Title:** Computational Accounts of Trust in Human AI Interaction

Shiraz University, Shiraz, Iran.
- **Master** student in Electrical Engineering, Control. (GPA: 3.79/4) 2015-2017
  - ○ **Thesis Title:** Nash equilibrium seeking without steady-state oscillation
  - ○ **Supervisors:** Dr. Alireza Khayatian, Dr. Mohammad Mehdi Arefi
- **B.Sc.**, Electrical Engineering, Control.(GPA: 3.73/4)　　　　　2011-2015

Experience

Scientist
- Honda Research Institute, San Jose, CA.　　　　　　　　　　　Current

Research Scientist Intern
- Honda Research Institute, San Jose, CA.　　　　　　Sep 2022-May 2023
- ○ **Key Responsibilities**
  - − Developed models to enhance adaptive decision-making in autonomous vehicles (AVs).
  - − Applied a DBN model for inference and leveraged a Causal Inference Model (CIM) for informed decision-making.
  - − Contributed to a comprehensive framework for responsible AV decision-making.

Skills

- **Programming Language:** *Python, MATLAB, Clojure, C/C++.*
- **AL/ML:** *Reinforcement Learning, Generative AI, LLMs, Transformers, Probabilistic modeling (Probabilistic Graphical Models, Markov Models, Bayesian Networks, and Statistical Modeling).*
- **Framework and libraries:** *Pgmpy, Keras, Pytorch, Scikit-learn, Pymc3, Pandas, TensorFlow.*
- **Generative AI related skills:** *In context learning, ReAct, RLHF, Fine-tuning (LoRa, PEFT)*

Publications

1. *Z. Zahedi, Modeling, Engendering and Leveraging Trust in Human-Robot Interaction: A Mental Model based Framework," HRI Pioneer 2024.*

2. *Z. Zahedi, S. Sengupta, S. Kambhampati, "'Why didn't you allocate this task to them?' Negotiation-Aware Task Allocation and Contrastive Explanation Generation," AAAI 2024.*

3. *Z. Zahedi, S. Sreedharan, S. Kambhampati, "A Mental Model based Theory of Trust," XAI workshop, IJCAI 2023.*

4. *Z. Zahedi, S. Mehrotra, T. Misu, K. Akash, "Toward Informed AV Decision-Making: A Computational Model of Well-being and Trust in Mobility," Under review.*

5. *S. Mehrotra, Z. Zahedi, T. Misu, K. Akash, Wellbeing in Future Mobility: "Toward AV Policy Design to Increase Wellbeing through Interactions," ITSC 2023.*

6. *Z. Zahedi, M. Verma, S. Sreedharan, S. Kambhampati, "Trust-Aware Planning: Modeling Trust Evolution in Iterated Human-Robot Interaction," HRI 2023.*

7. *Z. Zahedi, S. Sreedharan, K. Valmeekam, S. Kambhampati, "Explicable or Optimal: Trust Aware Planning in Iterated Human Robot Interaction," Accepted as a stand-alone video, ICRA 2023.*

8. *Z. Zahedi, S. Sengupta, S. Kambhampati, "'Why didn't you allocate this task to them?' Negotiation-Aware Explicable Task Allocation and Contrastive Explanation Generation," In proceeding of AAMAS 2023.*

9. *Z. Zahedi, S. Sreedharan, S. Kambhampati, "A Mental-Model Centric Landscape of Human-AI Symbiosis," R2HCAI Workshop, AAAI 2023.*

10. *Z. Zahedi, S. Sreedharan, M. Verma, S. Kambhampati, "Modeling the Interplay between Human Trust and Monitoring," HRI (LBR), 2022.*

11. *S. Sengupta*, Z. Zahedi*, S. Kambhampati, "Game-theoretic Model of Trust to Infer Human's Observation Strategy of Robot Behavior," Robot for people workshop, R:SS 2021.*

12. *Z. Zahedi, M. Verma, S. Sreedharan, S. Kambhampati, "Trust-Aware Planning:Modeling Trust Evolution in Longitudinal Human-Robot Interaction," XAIP and PlabRob Workshop, ICAPS 2021.*

13. *Z. Zahedi, S. Kambhampati, "Human-AI Symbiosis: A Survey of Current Approaches," 2021.*

14. *Z. Zahedi*, S. Sengupta*, S. Kambhampati, "'Why didn't you allocate this task to them?' Negotiation-Aware Task Allocation and Contrastive Explanation Generation," Cooperative AI Workshop, NeurIPS2020.*

15. S. Sengupta*, Z. Zahedi*, S. Kambhampati, "To Monitor or Not: Observing Robot's Behavior based on a Game-Theoretic Model of Trust," 21st International Workshop on Trust in Agent Societies (co-located with AAMAS), 2019.

16. Z. Zahedi*, A. Olmo*, T. Chakraborti, S. Sreedharan, S. Kambhampati, "Towards Understanding User Preferences for Explanation Types in Model Reconciliation," HRI Late Breaking Report, 2019.

17. Z. Zahedi, M. M. Arefi, A. Khayatian, "Seeking Nash equilibrium in non-cooperative differential games," Journal of Vibration and Control, 2022.

18. Z. Zahedi, M. M. Arefi, A. Khayatian, "Fast convergence to Nash equilibria without steady-State oscillation," Systems and Control Letters, 2019.

19. Z. Zahedi, M. M. Arefi, A. Khayatian, and H. Modares, "Fast seeking of Nash equilibria without steady-state oscillation in games with non-quadratic payoffs," in proc. 2018 American Control Conference.

20. Z. Zahedi, M. M. Arefi, and A. Khayatian, "Convergence without oscillation to Nash equilibria in non-Cooperative games with quadratic payoffs," 25th Iranian Conference on Electrical Engineering (ICEE), May 2017.

21. F. Zahedi, Z. Zahedi, "Real-time, Simultaneous Multi-Channel Data Acquisition Systems with no time skews between input channels," International Journal of Signal Processing Systems (IJSPS), vol. 4, no. 1, pp. 17-21, 2016.

22. F. Zahedi, Z. Zahedi, "A review of Neuro-fuzzy Systems based on Intelligent Control," Journal of Electrical and Electronic Engineering, vol. 3, no. 2-1, pp. 58-61, 2015.

23. F. Zahedi, Z. Zahedi, "Real-time, Simultaneous Multi-Channel Data Acquisition Systems with no time skews between input channels," 6th International Conference on Signal Processing Systems (ICSPS), 2014.

24. F. Zahedi, Z. Zahedi, "Review of Neuro-fuzzy based on Intelligent Control," 8th Symposium on Advances in Science & Technology," 2014 (Persian).

Recent Research Projects

- **Mental-Model based Theory of Trust**
  - We contextualize the user's trust and their consequent choice on whether or not to rely on the AI agent in terms of their mental model of the agent, and their preexisting expectations about the optimal way of solving the task. The agent on the other hand can either use behavior generated through its model, or explanations about its model, to influence both the user's expectations about the agent and their expectations about the task.

- **Trust Inference and Calibration with LLMs**
  - We are trying to utilize Large Language Models (LLMs) as an auxiliary model to infer human trust on the AI agent and calibrate the trust to an appropriate level of trust by updating human's information bases of trust.

- **Trust-Aware Planning in longitudinal human robot interaction**
  - We propose a computational model for capturing and modulating trust in longitudinal human-robot interaction where the robot integrates human's trust and their expectations into planning to build and maintain trust over the interaction horizon. By establishing the required level of trust, the robot can focus on maximizing the team goal by eschewing explicit explanatory or explicable behavior. The human also with a high level of trust in the robot,

might choose not to monitor the robot, or not to intervene by stopping the robot and save their resources.

- ***Game-Theoretic Model of Trust***
  - *In this work, we modeled a game theoretic model of trust in Human robot interaction setting and introduced a solution as a trust boundary which is a proper human's monitoring strategy given the inferred trust. By choosing a monitoring strategy in that boundary, the human supervisor can assure that the robot or the artificial intelligent system will show a safe and interpretable behavior.*

- ***Negotiation-Aware Task Allocation***
  - *We considered a task-allocation problem where an AI Task Allocator (AITA) comes up with a fair allocation for a group of humans. And the AITA is able to provide explanation in the form of negotiation tree to convince humans who thinks a counterfactual allocation is more fair.*

Honors

- *Dean's Dissertation Award, Arizona State University*   *2024*
- *Human Robot Interaction Pioneer*   *2024*
- *ASU Iranian American Alumni Academic Scholarship,*   *2023*
- *Grace Hopper's Scholarship, AnitaB.org*   *2022*
- *Fulton Fellowship Award, Arizona State University*   *2019-2020*
- *Grad Fellowship Award, Arizona State University   Fall 2019 and Spring 2020*
- *Hassan Mahdi Award, Shiraz University*   *2018*
- *Honored as **Outstanding student**, Shiraz University*   *2017*
- *Granted merit-based admission to M.Sc. in Shiraz University*   *2015*