Ashwin Kumar

ashwinkumar@wustl.edu (314) 629-3447 ashwinkumar97

CAREER SUMMARY

- Computer Science Ph.D. candidate, working in algorithmic fairness and Explainable AI Planning (XAIP). My focus is on improving fairness and transparency of AI systems.
- Skilled in modeling complex problems and building cross-functional teams and relationships.

EDUCATION

- **Washington University in St. Louis, School of Applied Science and Engineering** — **St. Louis, MO, USA**
  Ph.D., Computer Science (Current GPA: 4.0) — **Aug 19 - Ongoing**
- **Washington University in St. Louis, School of Applied Science and Engineering** — **St. Louis, MO, USA**
  Master of Science, Computer Science (GPA: 4.0) — **Aug 19 - May 22**
- **Shiv Nadar University, School of Engineering** — **Dadri, India**
  Bachelor of Engineering, Mechanical Engineering (CGPA 9.81, Gold Medalist) — **Aug 15 - May 19**

TECHNICAL SKILLS

- **Programming Languages/ Frameworks**: Python, C/C++, JavaScript, MATLAB, SQL, PyTorch, TensorFlow, Transformers, Gurobi.
- **Skills**: Deep Reinforcement Learning, Machine Learning, Natural Language Processing, Constraint Optimization
- **Relevant courses:** Data Structures and Algorithms, Advanced Algorithms, Machine learning, Bayesian methods in Machine Learning, Artificial Intelligence, Adversarial Methods in Machine Learning, Human-in-the-Loop Computation, Advanced Visualization

PROFESSIONAL AND ACADEMIC WORK EXPERIENCE

Research Scientist Intern, Meta Platforms, Inc., Menlo Park, CA — **Aug 23 –Jan 24**

- Engineered techniques to detect and mitigate demographic bias in preference models used in RLHF finetuning.
- Designed an attack to reveal prefix bias in preference models based on various LLM architectures (llama, vicuna, OPT, flan, GPTJ) trained on popular preference datasets. Developed metrics to detect and quantify this bias.
- Designed a data augmentation technique to mitigate prefix bias, significantly improving model fairness.

Research Scientist Intern, Meta Platforms, Inc., New York, NY — **May 22 –Aug 22**

- Designed methods for detection of bias feedback loops in content classifiers which use Model Assisted Sampling to send data points for human review.
- Quantified the propagation of bias to future models and devised sampling strategies to mitigate bias bootstrapping and accelerate recovery.
- Worked with multiple product teams to identify issues with different sampling techniques and generalize experiments.

Assistant in Instruction, Department of Computer Science, Washington University in St. Louis — **Jan 22 –May 22**

- Assisted over 130 students with coursework and homework relating to AI concepts and algorithms.
- Graded assignments and exams and delivered guest lectures on Reinforcement Learning and conducted problem-solving sessions on Logic and MDPs.

Research Intern, Indian Institute of Science, Bangalore, India — **May 18 - Jul 18**

- Applied dimensionality reduction techniques to strain energy expressions in hyper-elastic strip geometry, improving theoretical models for helicopter rotor blades to reduce vibration.

ACADEMIC PROJECTS

Using Machine Learning Models for Robot Control Prediction — **Jan 22 – May 22**

- Predicted control feature values using techniques like Gradient-Boosted Trees, Neural Network and Kernel Regression on a supervised learning problem.
- Performed feature engineering using correlation features, outlier removal, K-Means Clustering and Principle Component Analysis.

Adversarial Training Against Semantic Attacks Using SP Layers — **Jan 20 - Apr 20**

- Adversarially trained an MNIST digit classifier against Semantic attacks (translation, occlusion, brightness and contrast) by using semantic layers, showing higher bounds for certified robustness.

Ashwin Kumar

ashwinkumar@wustl.edu (314) 629-3447 ashwinkumar97

Hostile Takeovers in EM-based Crowdsourcing Systems **Aug 19 - Dec 19**

- Explored the possibility of hostile takeovers in crowd-sourced systems where a small group of influential users can ensure wrong decisions in a bi-partisan voting scenario, where algorithms like Expectation-Maximization are used to aggregate votes.

Smart Hexapod (Undergraduate Project) **Aug 18 – May 19**

- Designed and fabricated a smart hexapod robot with automated gait selection and terrain sensing capabilities for search and rescue operations in disaster areas. Created an algorithm for gait-based control of any general legged robot.

PUBLICATIONS

- **DECAF: Learning to be Fair in Multi Agent Resource Allocation**.
  Ashwin Kumar, and William Yeoh. *Under review.*
- **Using Simple Incentives to Improve Two-Sided Fairness in Ridesharing Systems**.
  Ashwin Kumar, Yevgeniy Vorobeychik, and William Yeoh. In *Proceedings of the International Conference on Automated Planning and Scheduling (ICAPS)*, 2023.
- **Fairness in Scarce Societal Resource Allocation: A Case Study in Homelessness Applications**.
  Ashwin Kumar, and William Yeoh. In *the 4th International Workshop on Autonomous Agents for Social Good (AASG)*, 2023.
- **Improving Zonal Fairness While Maintaining Efficiency in Rideshare Matching**.
  Ashwin Kumar, Yevgeniy Vorobeychik, and William Yeoh. In *Proceedings of the Workshop on Agents in Traffic and Transportation (ATT-22),* 2022.
- **VizXP: A Visualization Framework for Conveying Explanations to Users in Model Reconciliation Problems**.
  Ashwin Kumar, Stylianos Loukas Vasileiou, Melanie Bancilhon, Alvitta Ottley, and William Yeoh. In *Proceedings of the International Conference on Automated Planning and Scheduling (ICAPS)*, 2022.
  And In *Proceedings of the Workshop on Explainable Planning (XAIP)*, 2021.
- **A Logic-based Explanation Generation Framework for Classical and Hybrid Planning Problems**.
  Stylianos Loukas Vasileiou, William Yeoh, Tran Cao Son, **Ashwin Kumar,** Michael Cashmore, and Daniele Magazzeni. *In Journal of Artificial Intelligence Research (JAIR) Vol 73,* 2022.
- **FairVizARD: A Visualization System for Assessing Fairness of Ride-Sharing Matching Algorithms.**
  Ashwin Kumar, Sanket Shah, Meghna Lowalekar, Pradeep Varakantham, Alvitta Ottley, and William Yeoh. In *Proceedings of the International Conference on Automated Planning and Scheduling (ICAPS)* (System Demonstration), 2021