

Vijay Keswani

Durham, USA 27701
☎ +1-475-655-8485
✉ vijay.keswani@duke.edu
📄 vijaykeswani.github.io

Research Interests

I study problems related to algorithmic fairness, data and tech ethics, and community-focused AI development. Broadly, my work explores the interactions between society and automation and addresses various questions around the potentially harmful impacts of AI applications.

Experience

- 2023-now **Postdoctoral Associate**, *Duke University, USA*.
- 2022-2023 **Resident Fellow**, *Information Society Project (ISP), Yale Law School, USA*.
- 2022 **Policy Fellow**, *Yale Institute of Social and Policy Studies (ISPS), USA*.
- 2021 & 2020 **Research Intern**, *Amazon AWS AI, Palo Alto, USA*.
- 2016-2017 **Product Engineer**, *Sprinklr, Gurgaon, India*.
- 2014 **Research Intern**, *INRIA Paris-Rocquencourt and ENS, Paris, France*.
- 2013 **Research Intern**, *CSA Department, Indian Institute of Science, Bangalore, India*.

Education

- 2019-2023 **Ph.D., Statistics and Data Science**, *Yale University, U.S.*
- 2017-2019 **Ph.D. Candidate, Computer Science**, *EPFL, Switzerland*, (Moved to Yale in 2019).
- 2015-2016 **Master in Technology**, *Computer Science and Engineering, IIT Kanpur, India*.
- 2011-2016 **Bachelor in Technology**, *Computer Science and Engineering, IIT Kanpur, India*.

Research

Publications

- ICML 2024 **Fair Classification with Partial Feedback: An Exploration-Based Data-Collection Approach**, Vijay Keswani, Anay Mehrotra, L. Elisa Celis.
In partial feedback settings, outcomes are observed only for samples classified positively. We propose methods to train classifiers in this regime, focusing on outcome exploration of minority groups.
- EAAMO 2023 **Addressing Strategic Manipulation Disparities in Fair Classification**, Vijay Keswani, L. Elisa Celis.
In strategic settings, we show that fair classifiers do not address disparity in strategic manipulation cost across demographic groups. To address this, we propose a constrained optimization framework that constructs classifiers that lower the strategic manipulation cost for disadvantaged groups.
- EcAI **Social Media Platform Structures and Their Implications**, Vijay Keswani.
Workshop Using philosophical notions of “structures”, I analyze the structures imposed by social media platforms and how the use of these structures imply that these platforms do not qualify as “neutral publishers”.
- HHAI 2023 **Designing Closed-Loop Models for Task Allocation**, Vijay Keswani, L. Elisa Celis, Matthew Lease, Krishnam Kenthapadi.
To train an online allocation model, we exploit weak prior information on human-task similarity to bootstrap accurate model training, even when human decision-makers are biased.

- EAAMO 2022 **The Representation Pact - A Case Study in Computational Participatory Elections**, Florian Evéquo, Johan Rochel, Vijay Keswani, L. Elisa Celis.
A novel participatory electoral process to select representative committees. Complemented by a case study of our implementation of this process in primary elections in Valais, Switzerland.
- ICML 2022 **A Convergent and Dimension-Independent Min-Max Optimization Algorithm**, Vijay Keswani, Oren Mangoubi, Sushant Sachdeva, Nisheeth K. Vishnoi.
A feasible algorithm to find local equilibrium points for min-max optimization problems with applications for training Generative Adversarial Networks (GANs).
- BHCC 2021 **Designing human-in-the-loop approaches for closed deferral pipelines**, Vijay Keswani, Non-archival *Matthew Lease, Krishnaram Kenthapadi*.
A closed pipeline design to combine label elicitation and learning components of a decision-making framework with an option of deferring to human experts for contentious input data.
- KDD 2021 **Auditing for Diversity using Representative Examples**, Vijay Keswani, L. Elisa Celis.
A cost-effective approach to approximate the disparity of any given unlabeled dataset, with respect to a protected attribute, using a small set of labeled representative examples.
- AIES 2021 **Towards Unbiased and Accurate Deferral to Multiple Experts**, Vijay Keswani, Matthew Lease, Krishnaram Kenthapadi.
A framework to learn a classifier and a deferral model that defers to a domain expert in cases where the classifier has low confidence in its inference.
- ICML 2021 **Fair Classification with Noisy Protected Attributes: A Framework with Provable Guarantees**, L. Elisa Celis, Lingxiao Huang, Vijay Keswani, Nisheeth K. Vishnoi.
An optimization framework for learning a fair classifier in the presence of noisy perturbations in the protected attributes that comes with provable guarantees on both accuracy and fairness.
- The Web Conf. 2021 **Dialect Diversity in Text Summarization on Twitter**, Vijay Keswani, L. Elisa Celis.
Analysis of how standard text summarizations can under-represent certain dialects and application of a post-processing algorithm to generate dialect-diverse summaries for Twitter datasets.
- CSCW 2020 **Implicit Diversity in Image Summarization**, L. Elisa Celis, Vijay Keswani.
A post-processing algorithm for fair image search and summarization that uses a small set of diverse examples to induce diversity in the generated image summary.
- ICML 2020 **Data preprocessing to mitigate bias: A maximum entropy based approach**, L. Elisa Celis, Vijay Keswani, Nisheeth K. Vishnoi.
A pre-processing framework to mitigate biases that leverages the principle of maximum entropy.
- FAT* 2019 **Classification with Fairness Constraints: A Meta-Algorithm with Provable Guarantees**, L. Elisa Celis, Lingxiao Huang, Vijay Keswani and Nisheeth K. Vishnoi.
A meta-algorithm for fair classification takes the fairness type and constraint as input and returns a classifier which satisfies the fairness constraint at minimal cost to accuracy.
- ICML 2018 **Fair and Diverse DPP-based Data Summarization**, L. Elisa Celis, Vijay Keswani, Damian Straszak, Amit Deshpande, Tarun Kathuria, Nisheeth K. Vishnoi.
A simple linear-time approximate algorithm for fair summarization that samples from DPP (Determinantal Point Process) distributions with fairness constraints.
- IJCAI-ECAI 2018 **Balanced News Using Constrained Bandit-based Personalization**, Sayash Kapoor, Vijay Keswani, Nisheeth K. Vishnoi, L. Elisa Celis.
A news-search prototype that de-polarizes the news feed by presenting balanced viewpoints across liberal and conservative articles.

Working Papers

- 2024 **Algorithmic Fairness From the Perspective of Legal Anti-discrimination Principles**, *Vijay Keswani, L. Elisa Celis*.
We use the legal framework of anti-subordination to study the motivations of a fair classifier and its applications. Using this principle, we propose guidelines that a fair machine learning algorithm could follow to ensure an equitable and progressive impact on the affected population.
- 2024 **On the Pros and Cons of Active Learning for Moral Preference Elicitation**, *Vijay Keswani, Vincent Conitzer, Hoda Heidari, Jana Schaich Borg, Walter Sinnott-Armstrong*.
Using prior literature from moral psychology, we highlight possible issues with using active learning to extract moral preferences and the downstream implications in training ethical AI systems.

Ph.D. Dissertation

- 2023 **Algorithmic Decision-Making with Stakeholder Participation**, *with L. Elisa Celis*, Statistics and Data Science, Yale University.
This dissertation advances methods to build trustworthy decision-making systems by focusing on stakeholder participation, in the forms of user feedback incorporation or domain expert participation.

Master's Dissertation

- 2016 **Laplacian Solvers and Graph Sparsification**, *with Rajat Mittal*, CSE, IIT Kanpur.
This thesis explored the scope of spectral sparsification algorithms used in Laplacian solvers and the relations between different state-of-the-art Laplacian solvers.

Demos

- 2021-22 **Imaginaries**.
Using the algorithm from our KDD 2021 paper on auditing using representative samples, we developed a Firefox plugin, called Imaginaries, that can efficiently and accurately quantify the extent of representational biases in Google Image Search results.
- 2019 **Application of the Fair Multiwinner Elections framework**.
Helped employ the framework for Fair Multiwinner Elections for the primary elections of Appel Citoyen in the region of Valais, Switzerland.
- 2018 **Balanced News demo**.
We develop a demo to demonstrate what a balanced content delivery engine would look like, comparing our news content delivery (balanced news) with existing approaches (unfiltered news).

Talks

- May 2024 **AI in the Public Domain: Applications and Ethical Challenges**, *Durham Center for Senior Living, USA*.
- April 2024 **On the Effectiveness of Computational Methods for Modeling Moral Judgments**, *Moral Psychology Research Group, Duke University, USA*.
- March 2024 **Guest Lecture on Affirmative Action**, *Ethics Capstone Seminar, Duke University, USA*.
- August 2023 **Participation and Fairness in Algorithmic Decision-Making**, *Moral Attitudes and Decision-making Group, Duke University, USA*.
- July 2023 **Social Media Platform Structures and Their Implications**, *Workshop on Ecology of AI (EcAI), Virtual*.
- July 2023 **Designing Closed-Loop Models for Task Allocation**, *International Conference on Hybrid Human-Artificial Intelligence (HHAI 2023), Virtual*.

- October 2022 **Diverse Representation via Computational Participatory Elections - Lessons from a Case Study**, *Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO 2022)*, Arlington, USA.
- October 2022 **Strategic Manipulation Disparities in Fair Classification**, *INFORMS 2022*, Indianapolis, USA.
- October 2021 **Designing human-in-the-loop approaches for closed deferral pipelines**, *Biases in Human Computation and Crowdsourcing (BHCC 2022)*, Virtual.
- May 2021 **Towards Unbiased and Accurate Deferral to Multiple Experts**, *Artificial Intelligence, Ethics, and Society (AIES 2021)*, Virtual.