

Kefei Yan

kyan66@gatech.edu

kefeiyan.com | Google Scholar | ORCID | github.com/cliff003 | linkedin.com/in/kefeiyan

Atlanta, Georgia, United States

Research Interests

Evaluation of machine learning and multi-agent large language model systems that support consequential decisions in environmental and other high-stakes settings, where ground truth is contested and errors fall unevenly across the people affected; how sampling design and detection limits shape what a monitoring record can support; and decision support that makes its own reasoning, and its own failures, legible.

Education

Georgia Institute of Technology

Atlanta, Georgia, United States

Master of Science in Computer Science, Artificial Intelligence Specialization *Expected December 2026*

- GPA: 3.88/4.00. Selected graduate coursework: Artificial Intelligence, Machine Learning, Graduate Algorithms, Sustainability and Computing, Game AI, and Deep Learning.

University of Colorado Boulder

Boulder, Colorado, United States

Master of Science in Data Science

June 2024

- GPA: 3.99/4.00. Graduate training in machine learning, statistical inference, data mining, algorithms, scalable analytics, and big-data systems.

University of Toronto

Toronto, Ontario, Canada

Bachelor of Science in Biological Chemistry

December 2015

Publications

Peer-Reviewed Conference Proceedings

1. **Kefei Yan***, Frank F. Yang*, Yiyang Wang, Ziyi Yang, Kewen Liu, Yuzhuo Chen, Alex Cabral, and Josiah Hester. “FireFair: Equity-Adjusted Multi-Agent Triage for Wildfire Ignition Forecasting.” In *Proceedings of the ACM International Conference on Information Technology for Social Good (GoodIT ’26)*, Pisa, Italy, 2026. Work-in-Progress paper, presented September 2026. DOI: 10.1145/3794786.3830751.

Preprints

1. Yiyang Wang, Moeiini Reilly, Britney Johnson, **Kefei Yan**, Alex Cabral, and Josiah Hester. “Culti-vAgents: Cultivating Relationship-Centered Multi-Agent Systems for Personalized Gardening.” *arXiv preprint arXiv:2605.23193*, 2026. arxiv.org/abs/2605.23193.

*Equal contribution.

Presentations

1. “FireFair: Equity-Adjusted Multi-Agent Triage for Wildfire Ignition Forecasting.” Conference talk, speaker. ACM International Conference on Information Technology for Social Good, Pisa, Italy, September 2026.
2. **Kefei Yan**, Neelima Pandey, John Hiedo, and Marissa Douglas. “Bridging Empirical and Modeled Biodiversity: Tracking Agreement between Camera Traps and Species Range Maps.” Poster, Online Master of Science in Computer Science Conference, Georgia Institute of Technology, May 2026.

Research Experience

FireFair: Equity-Adjusted Wildfire Ignition Triage

Atlanta, Georgia, United States

Co-first Author and Graduate Researcher, Georgia Institute of Technology

January 2026–Present

Advisors: [Alex Cabral](#) and [Josiah Hester](#). Project site: fire-fair.github.io

- Audited a California wildfire ignition forecasting model across quintiles of the Social Vulnerability Index from the Centers for Disease Control and Prevention and the Agency for Toxic Substances and Disease Registry, and found an equal-opportunity recall gap of -0.155 between the most- and least-vulnerable communities.
- Cut that disparity by 75%, from -0.155 to -0.038 , with an auditable equity-aware score adjustment that cost 1.6 percentage points of aggregate F_1 .
- Designed a three-layer decision-support framework that combines equity-adjusted scoring, community-specific routing gates, and multi-agent verification over wildfire forecasts, satellite imagery, active-fire detections, and social-vulnerability data, then built an interactive prototype whose visible tool-call traces let users follow the evidence behind each recommendation.
- Validated the system on a held-out California test fold and on two 2025 wildfire case studies that fell after the model’s training window.

CultivAgents: Relationship-Centered Gardening Support

Atlanta, Georgia, United States

Graduate Researcher and Co-author, Georgia Institute of Technology

January 2026–May 2026

Advisors: [Alex Cabral](#) and [Josiah Hester](#). Project site: hello-diana.github.io/CultivAgents

- Co-designed a multi-agent large language model system in which experiential, environmental, and ethnobotanical agents work together to offer personalized, culturally grounded gardening guidance.
- Carried out the quantitative analysis for a three-phase mixed-methods evaluation, run under an approved institutional review board protocol with gardeners at the Georgia Tech community garden, covering participant confidence, motivation, and trust alongside perceptions of multi-agent collaboration, and prepared the workshop materials the sessions were run from.

Spatial Camera-Trap Biodiversity Project

Atlanta, Georgia, United States

Graduate Researcher, Georgia Institute of Technology

May 2025–Present

Human-Augmented Analytics Group. Advisors: [Jenny McGuire](#) and [Steve Mussmann](#). Repository: github.com/cliff003/HAAG_Spatial_Camera_Trapping_Fall2025

- Built a reproducible continental-scale pipeline that merges 987,979 Snapshot USA wildlife records with International Union for Conservation of Nature range maps and COMBINE species-

trait data, producing 713,319 analysis-ready records across 109 mammal species, 7,340 cameras, and 262 arrays.

- Showed how much sampling design shapes what a sensing network appears to capture: pooling individual cameras into arrays more than doubled mean agreement between observed and expected communities, from 0.254 to 0.575, so the agreement a network reports depends heavily on the spatial scale at which it is read.
- Traced most of the remaining error to detection sensitivity, since limiting the analysis to species of at least 500 grams closed about 96% of the national richness gap.
- Trained a Random Forest model on habitat, land-cover, development, and road-distance covariates to predict agreement at each camera, and found that local site conditions, led by distance to the nearest road, then cropland and forest cover, mattered most at the scale of individual cameras.
- Found that longer monitoring helps as well, with five years of continuous survey raising mean similarity from 0.518 to 0.660 across 41 repeatedly surveyed arrays.
- Built validated one-kilometer camera footprints and array-level spatial unions in EPSG:5070, then compared communities using spatial joins, Jaccard similarity, species-accumulation curves, and Chao–Jost sample coverage.

Teaching Experience

Georgia Institute of Technology
Graduate Teaching Assistant

Atlanta, Georgia, United States
August 2025–Present

CS 7632: Game AI

Fall 2025 and Fall 2026

- Led weekly office hours, supporting students with Unity, C#, artificial intelligence algorithms, debugging, and project development.
- Guided students through behavior trees, pathfinding, decision-making systems, and game-agent design, then reviewed six programming assignments and two course projects, writing reference solutions and grading guidance along the way.

CS 6239: Enterprise Cybersecurity Management

Spring 2026

- Assessed graduate-level case studies involving enterprise risk, security governance, and regulatory requirements, providing written feedback on security-management frameworks, risk analysis, compliance, and responsible use of generative artificial intelligence.

Academic Service

ACM International Conference on Information Technology for Social Good
Conference Reviewer

2026

- Reviewed a submission on artificial intelligence governance and multi-agent optimization, weighing its methodology, empirical support, reproducibility, and reference accuracy, then wrote detailed comments for the authors and confidential recommendations for the program

committee.

Awards and Honors

Digital Green Crop Yield Prediction Challenge

Gold-Level Recognition, ranked 20th among 647 teams

2023

Research-Relevant Industry Experience

Dematic

Artificial Intelligence Engineer Intern

Atlanta, Georgia, United States

June–August 2026

- Prototyped enterprise agentic workflows over project records for knowledge retrieval, risk registries, and schedule forecasting.
- Developed criteria for evaluating those agents in a setting with no labeled ground truth, where expert judgment served as the reference standard and task usefulness, reliability, governance, and deployment constraints had to be weighed together.

Zhongke HZ Technology Co., Ltd.

Artificial Intelligence Algorithm Engineer

Guangdong, China

March 2022–April 2024

- Developed an end-to-end environmental surveillance pipeline integrating a Java backend with Python-based machine-learning inference for real-time scene analysis.
- Developed a runway foreign-object-debris detection module.
- Evaluated YOLO, DETR, and reinforcement-learning components for object detection and automated decision support, then tuned inference latency, frame rate, and predictive accuracy.

Technical Skills

Programming and Systems

Python, R, Java, C#, C/C++, SQL, PySpark, Spark, Databricks, Docker, Linux, Amazon Web Services, Microsoft Azure, and Google Cloud Platform

Machine Learning and Generative Models

PyTorch, TensorFlow, scikit-learn, pandas, NumPy, computer vision, deep learning, Random Forests, algorithmic fairness, held-out and mixed-methods evaluation, large language models, multi-agent systems, retrieval-augmented generation, and tool-using agents

Geospatial and Ecological

GeoPandas, Shapely, QGIS, spatial joins, coordinate-reference systems, species-accumulation curves, Jaccard similarity, Chao–Jost coverage, occupancy modeling, rpy2, and vegan

Last updated: September 16, 2026