Suraj Rajendran

420 E 70th St – New York, NY 10021

224-587-0371 • sur4002@med.cornell.edu • surajraj99

Education

Weill Cornell Medical College — **New York, NY**
Tri-Institutional PhD Program in Computational Biology — *Exp. May 2026*

Georgia Institute of Technology — **Atlanta, GA**
BS in Biomedical Engineering (GPA: 3.96/4.0) — *May 2021*
Minor in Computing and Intelligence (GPA: 4.0/4.0)

Research Experience

Weill Cornell Medical College — **New York, NY**
Graduate Researcher — *Aug 2021 – Present*

AI Agents and Foundation Models for Predictive Embryology in IVF Procedures
- Engineered a foundation model, FEMI, with a masked autoencoder architecture, on 20 million embryo images.
- Advanced prediction on multiple clinically relevant tasks by employing FEMI, including embryo quality and ploidy prediction.
- Developed AI agents to use FEMI as a tool to perform real-time 24/7 embryo monitoring and grading in the clinic.
- Integrated AI agents into the clinical workflow at Weill Cornell's Reproductive Center to boost clinician efficiency by 40%.

Generative Design of Clinical-Grade HLA-G5 Binders
- Targeted disordered HLA-G5 tails via RFdiffusion and partial diffusion to resolve $> 80\%$ MHC-I homology issues.
- Engineered soluble binders via ProteinMPNN "acidic supercharging" to mimic albumin and prevent aggregation.
- Validated designs using consensus of AlphaFold2, Chai-1, Boltz-1 ensuring $> 60\%$ folding success.
- Identified a lead candidate for IVF monitoring with zero predicted cross-reactivity against HLA-A/B/C alleles.

Agentic AI Framework for Clinical Trial Design Automation
- Developed TrialGenie, an agentic AI framework that automates computational clinical trial design and implementation.
- Engineered a 5-agent system via LangChain and LangGraph that autonomously analyzes real-world data.
- Validated the framework on the MIMIC-IV dataset across 4 acute diseases, generating trial reports in under 2 hours.

Structural Variant Detection in Short Read Cancer Data Using Deep Learning
- Engineered Cue-somatic, a deep learning pipeline using an hourglass network to detect subclonal structural variants.
- Achieved high recall, surpassing benchmark callers and maintaining $> 98.5\%$ recall on low-frequency variants.
- Validated on 3 real cancer cell lines, discovering unique, low allele frequency somatic SVs in key cancer genes.

Determining Steroid Efficacy in Sepsis through Target Trial Emulations
- Created a ML method to stratify subphenotypes based on trajectory, distinguishing responses to corticosteroids.
- Analyzed a large ICU dataset from three sources to enhance findings on corticosteroid efficacy across subphenotypes.
- Employed causal inference tools and trial emulation methods to achieve pseudo-randomization and proper outcome metrics.
- Utilized a Cox model to study 3 outcomes, deriving hazard ratios for tailored treatment strategies.

Investigating Impact of Data Heterogeneity in Federated Learning
- Delineated a protocol to test the impact of data heterogeneity in local, pooled, and federated model settings.
- Collected Acute Kidney Injury (AKI) and Sepsis data from 7 different hospitals within the New York medical system.
- Adapted the SecureBoost algorithm to investigate the effect of data discrepancies within federated XGBoost.
- Ranked 354 medications, lab test measurements, and vital signs, uniquely important at each New York hospital.

Predicting Ploidy Status of Embryos Using Deep Video Classification
- Developed a pipeline to train models on in vitro fertilization time-lapse videos to determine ploidy status for 2000 embryos.
- Collaborated with embryologists to determine ways to standardize videos of embryos to ensure non-biased predictions.
- Created a pre-trained CNN-LSTM architecture based on ImageNet to process and predict ploidy status.
- Explored state-of-the-art video classification models such as 3DConvNet and I3D to determine efficacy of application.

Wake Forest School of Medicine — **Winston-Salem, NC**
Bioinformatics Researcher, PI: Dr. Umit Topaloglu — *May 2019 – Sep 2021*

Privacy Preserving Methods Through Holographic Transformations
- Performed frequency domain transformations on multiple standard ML datasets to mask information more efficiently.
- Created pipeline for frequency domain masking of datasets, allowing for information control by active parties.
- Validated transformed dataset security through simulating a Generative Adversarial Network (GAN) attack.

Identification of Immunotherapy Related Adverse Events (irAEs)
- Used machine learning to develop prediction models that will aid providers in identifying patients at high risk.
- Implemented novel word embedding techniques like BioBERT to clinical notes in order to extract greater meaning.
- Developed an LSTM model with over 80% accuracy at multiclass prediction of different irAEs.

Predicting COVID-19 Diagnosis Using Televisit and Progress Notes
- Investigated the application of 3 deep neural network models on patient notes to extract COVID-19 symptoms.
- Preprocessed and cleaned patient notes via Python's NLTK NLP, regex, and autocorrect libraries.
- Trained DNN models on notebook instance provided by Google Cloud's HPC services.
- Classified over 40,000 presumptive patients as COVID positive or negative using best performing DNN models.

Federated Learning Using Cloud Computing
- Developed environment to test the efficiency and performance of three transfer learning methods.
- Created a protocol for sharing models via GitHub from one institution to another without sharing private data.
- Delineated the results of the developed federated learning mechanisms in a manuscript submitted to JCO CCI.

Detecting Smoking Status Using Natural Language Processing
- Created an NLP Pipeline that cleans EHR data using Python libraries such as gensim, spaCy, and Keras.
- Developed 6 machine learning models that classified patients based on their smoking status using progress notes.
- Determined optimal parameters for learning rate, batch size, and loss functions using different search techniques.

Biophotonics Lab **Atlanta, GA**
Undergraduate Researcher, PI: Dr. Shu Jia *Aug 2019 – May 2021*

Enhancing Temporal Resolution of Optical Microscopy
- Created a novel and practical method to increase temporal resolution two-fold in biomedical imaging.
- Developed a mathematical model to determine the intensity of light at different temporal subframes in an sCMOS.
- Drafted and parameterized the attributes of a custom fiber array that transforms images from 2D to 1D.
- Created a simulation that utilized the developed mathematical model to output a super-resolution image.
- Designed and implemented theoretical setup to validate efficiency and functionality of method.

Fast and Accurate sCMOS Noise Correction
- Transcribed code for Automatic Correction of sCMOS-related Noise (ACsN) from MATLAB to Python.
- Combined camera physics and layered sparse filtering to reduce most relevant noise sources in a sCMOS sensor.
- Improved the camera performance, enabling fast and quantitative optical microscopy with video-rate denoising.

Yunker Lab for Evolutionary Mechanics **Atlanta, GA**
Undergraduate Researcher *Jan 2019 – Jun 2019*

Investigating the Rise of Multicellularity Among Yeast Cells in Varying Flow Regimes
- Performed daily selection speed tests on snowflake yeast for 8 weeks in order to track multicellular growth.
- Implemented a protocol for experiments in a glycerol solution to test evolutionary mechanics in different flows.
- Developed a quantitative model to compare the theoretical flow of cells in a turbulent media to experimental values.

Professional Experience

Regeneron **Rensselaer, NY**
Data Scientist Intern *May 2024 – Aug 2024*

Streamlining Business Processes using AI and LLMs
- Reduced text processing time by 95% using NLP and LLMs for analyzing unstructured data for categorizing various texts.
- Led contradiction detection project for regulatory documents, enhancing integrity by 80% with comparison model.
- Implemented analytics and data visualization tools using GPT-4 and PandasAI to transform data into actionable insights.

Colgate-Palmolive **Topeka, KS**
R&D Bioinformatics Intern *Jun 2021 – Aug 2021*

Diagnosis Mapping and Searching within Veterinary Notes
- Developed two pipelines for standardizing new diagnoses entered by clinicians by using NLP and SNOMED relations.
- Created an algorithm to find diagnoses within medical notes supplementing veterinarian diagnosis.
- Added thresholds and features to the pipelines to allow for user modulation based on use application.
- Validated results of the pipelines with clinicians and the ground truth to verify high performance and accuracy.
- Presented pipelines and algorithms to professionals and created documentation to facilitate future use of the built programs.
- Constructed a multimodal Long Short-Term Memory (LSTM) model for early prediction of renal disease.

Mayo Clinic **Jacksonville, FL**

Convergence Science Intern *Dec 2020 – Jun 2021*

Lateral Access Torso Model for Surgical Simulation

- Developed an anatomical simulator for lateral access lumbar spine surgery to provide practical training of surgeons.
- Implemented a novel Python-based neurostimulation algorithm that provides random auditory and visual feedback.
- Integrated radiopaque vertebrae with synthetic tissue to provide realistic feedback within anatomical simulator.
- Conducted prior art research to determine patentability and freedom to operate within the surgery simulator field.
- Marketed and presented simulator to over 100 medical professionals, investors, and surgeons.

Traits AI **San Diego, CA**

Software Engineering Intern *Aug 2018 – Sep 2019*

Voice Assistants for Google Assistant and Amazon Alexa

- Coded script and UI which users would be interacting with on the Google Assistant Agents or Amazon Alexa.
- Created 3 Voice Activated Conversational AI using Google SDK, Node.js, and the Jovo Database.
- Monitored user interaction with Google Agents and Alexa Skills using Dialogflow Analytics and DynamoDB.
- Managed and regulated the AWS Lambda servers on which Actions and Skills were hosted.
- Designed seamless UI/UX for Traits AI website, Google Actions, and Alexa Skills.

Publications

Published and Accepted Manuscripts

[1] **Rajendran S**, Rehani E, Phu W, Zhan Q, Malmsten J, Meseguer M, Miller K, Rosenwaks Z, Elemento O, Zaninovic N, Hajirasouliha I. *A foundational model for in vitro fertilization trained on 18 million time-lapse images*. Nature Communications. (2025)

[2] Li H, Zang C, Xu Z, Pan W, Zang C, **Rajendran S**, Chen Y, Wang F. *Federated target trial emulation using distributed observational data for treatment effect estimation*. npj Digital Medicine. (2025)

[3] **Rajendran S**, Xu Z, Pan W, Zang C, Siempos I, Torres L, Xu J, Bian J, Schenck E, Wang F. *Multicenter target trial emulation to evaluate corticosteroids for sepsis stratified by predicted organ dysfunction trajectory*. Nature Communications. (2025)

[4] **Rajendran S**, Brendel M, Barnes J, Zhan Q, Malmsten J, Zisimopoulos P, Sigaras A, Ofori-Atta K, Meseguer M, Miller KA, Hoffman D, Rosenwaks Z, Zaninovic N, Elemento O, Hajirasouliha I. *Automatic Ploidy Prediction and Quality Assessment of Human Blastocyst Using Time-Lapse Imaging*. Nature Communications. (2024)

[5] Mandracchia B, Zheng C, **Rajendran S**, Liu W, Forghani P, Xu C, Jia S. *High-speed optical imaging with sCMOS pixel reassignment*. Nature Communications. (2024)

[6] **Rajendran S**, Pan W, Sabuncu MR, Zhou J, Wang F. *Learning Across Diverse Biomedical Data Modalities and Cohorts: Challenges and Opportunities for Innovation*. Cell Patterns. (2024)

[7] Pan W, Xu Z, **Rajendran S**, Wang F. *An adaptive federated learning framework for clinical risk prediction with electronic health records from multiple hospitals*. Cell Patterns. (2024)

[8] **Rajendran S**, Brendel M, Barnes J, Zhan Q, Malmsten J, Rosenwaks Z, Meseguer M, Zaninovic N, Elemento O, Hajirasouliha I. *Predicting Embryo Ploidy Status Using Time-lapse Images*. Human Reproduction 38. (2023)

[9] Cao Y, **Rajendran S**, Sundararajan P, Law R, Bacon S, Sumner S, Masuda N. *Web-Based Social Networks of Individuals With Adverse Childhood Experiences: Quantitative Study*. Journal of Medical Internet Research. (2023)

[10] Su C, Hou Y, Zhou M, **Rajendran S**, Maasch J, Abedi Z, Zhang H, Bai Z, Cuturrufo A, Guo W, Chaudhry F, Ghahramani G, Tang J, Cheng F, Li Y, Zhang R, DeKosky S, Bian J, Wang F et al. *Biomedical Discovery through the integrative Biomedical Knowledge Hub (iBKH)*. iScience. (2023)

[11] **Rajendran S**, Xu Z, Pan W, Ghosh A, Wang F. *Data Heterogeneity in Federated Learning with Electronic Health Records: Case Studies of Risk Prediction for Acute Kidney Injury and Sepsis Diseases in Critical Care*. PLOS Digital Health. (2023)

[12] Barnes J, Brendel M, Gao V, **Rajendran S**, Jim K, Li Q, Malmsten J, Sierra J, Zisimopoulos P, Sigaras A, Khosravi P, Meseguer M, Zhan Q, Rosenwaks Z, Elemento O, Zaninovic N, Hajirasouliha I. *Development of non-invasive artificial intelligence models for the prediction of human blastocyst ploidy*. Lancet Digital Health. (2023)

[13] Pullen M, Valero-Moreno F, **Rajendran S**, Shah V, Bruneau B, Martinez J, Ramos-Fresnedo A, Quinones-Hinojosa A, Fox C. *Creation of a Proof-of-Concept 3D-Printed Spinal Lateral Access Simulator*. Cureus 14. (2022)

[14] **Rajendran S**, Ong T, Zameza P, Wolfe S, Topaloglu U, Duncan P, Anwar M, Samuel R, Budigi B, Lack C, Sarwal A. *Including social determinants of health in prognostic models for intracerebral hemorrhage*. Critical Care Medicine 50 (1). (2022)

[15] Topaloglu M, Morrell E, **Rajendran S**, Topaloglu U. *In the Pursuit of Privacy: The Promises and Predicaments of*

Federated Learning in Healthcare. Frontiers in Artificial Intelligence. (2021)

[16] **Rajendran S**, Obeid J, Binol H, D'Agostino R, Foley K, Zhang W, Austin P, Brakefield J, Gurcan M, Topaloglu U. *A Cloud Based Federated Learning Implementation Across Medical Centers*. JCO Clinical Cancer Informatics. (2021)

[17] Margalski D, Lycan T, **Rajendran S**, Topaloglu U. *Machine learning for prospective identification of immunotherapy related adverse events (irAEs)*. Journal of Clinical Oncology (2020) 38.

[18] **Rajendran S**, Topaloglu U. *Extracting Smoking Status from Electronic Health Records Using NLP and Deep Learning*. AMIA Jt Summits Translational Sci Proc. (2020) 507-516.

Presentations

Poster Presentations

- **Rajendran S**, Holt J. *Effect of Flow Regime on Snowflake Yeast Geometry*. South Eastern Regional Yeast Conference (SERYM) (2019)

Oral Presentations

- **Rajendran S**, Brendel M, et al. *Automatic Ploidy Prediction and Quality Assessment of Human Blastocysts Using Time-Lapse Imaging*. Nature Communications. (2024)
- **Rajendran S**, Sundararajan P. *Methods and Results for Models Developed to Score Student Responses Using BERT*. National Assessment of Educational Progress Colloquium. (2022)
- **Rajendran S**, Sundararajan P. *Predicting criminal recidivism using specialized feature engineering and XGBoost*. National Institute of Justice Symposium. (2021)
- **Rajendran S**. *Extracting Smoking Status from Electronic Health Records Using NLP and Deep Learning*. AMIA Jt Summits Translational Sci Proc. (2020)

Fellowships and Accolades

Fellowships

PhD Student
National Science Foundation Graduate Research Fellowship (NSF GRFP) — *Aug 2023 – Aug 2026*

Honors & Awards

- **2023 NIH BEAMS Challenge Winner:** $5,000 Cash Prize — *Dec 2023*
- **2023 ARPA-H Dash to Accelerate Health Outcomes:** Quarterfinalist, $10,000 Cash Prize — *May 2023*
- **2023 UTSW Annual Healthcare Case Competition:** Top 5 Finalist — *Mar 2023*
- **US Dept. of HHS Giving=Living Innovation Challenge:** Winner, $10,000 Cash Prize — *Feb 2023*
- **2022 UPenn Healthcare Case Competition:** Top 5 Finalist — *Nov 2022*
- **Georgia Tech Student Innovation Competition:** Promoting Equity and Access, $1000 Prize — *Apr 2022*
- **US Dept. of Commerce Smart Tracking Challenge:** Phase I Winner, $5000 Prize — *Apr 2022*
- **Georgia Tech Hacklytics 2022:** 2nd Place Healthcare, 2nd Place Overall — *Feb 2022*
- **CDC & NASA Detecting Emerging Threats Challenge:** 1st Place, $7500 Prize — *Jan 2022*
- **DOE Automated Scoring Challenge:** 4th Place, $1250 Prize — *Jan 2022*
- **Mayo Clinic Healthcare Hackathon:** 3rd Place, $1000 Prize — *Oct 2021*
- **DOJ Recidivism Forecasting Challenge:** Won 4 Awards totalling $23,000 — *Aug 2021*
- **Georgia Tech $1B+ Startup Hackathon:** Finalist (17/193) — *Apr 2021*
- **Pueblo Data Mine Analytics Challenge:** 2nd Place (2/100), $800 Prize — *Mar 2021*
- **CarMax ML/AI Data Analytics Showcase:** 1st Place (1/200 Teams), $3000 Prize — *Feb 2021*
- **AAMI Foundation's Michael J. Miller Scholarship:** $3000 Prize — *Jan 2021*

Service & Outreach

Lumiere Education — **New York, NY**
Research Mentor — *Mar 2023 – Present*

- Mentored a student in ML for healthcare; supervised project on diagnosis prediction (boosted accuracy 2-3%).
- Supervised a project on diagnosis prediction using medical imaging data, boosting diagnostic accuracy by 2-3
- Co-authored a conference-accepted research paper; contributed to student's recognition at a conference in India.
- Fostered the student's growth, contributing to their recognition at a prestigious conference in India

iMentor NYC **New York, NY**

Mentor *Nov 2023 – Present*

- Coached a mentee through the college application cycle, from building a college list to submitting 5+ applications and completing the FAFSA.
- Dedicated 40+ hours of one-on-one mentorship, assisting a student in developing key college-readiness skills, including essay writing, resume building, and financial aid literacy

NIH BEAMS Challenge **New York, NY**

Team Leader *Oct 2023 – Dec 2023*

- Spearheaded a lesson plan around "Telephone" focusing on making concepts around DNA accessible to middle school students
- Created an interactive and story-driven curriculum that simplified the principles of genetic science and CRISPR technology
- Collaborated with a team of educators to develop educational materials that will be shared with schools in 2024

HHS Blood Donation Campaign **New York, NY**

Team Leader *Oct 2022 – Feb 2023*

- Developed a winning proposal for the "Giving=Living" campaign, aimed at promoting blood donations to address shortages
- Conducted formative research and identified 4 barriers to blood donation, leading to a reduction in negative experiences
- Designed a community-driven approach that increases short-term blood supply and ensures long-term sustainability
- Collaborated with stakeholders to ensure donations among diverse populations, specifically Black and Latino communities

Addressing Inequities in Academic Recognition **Atlanta, GA**

Team Leader *Dec 2021 – Apr 2022*

- Proposed a policy to ensure that students with disabilities get academic recognition for completed courses at Georgia Tech
- Interviewed 5 Georgia Tech individuals, including staff at Office of Disability Services and Diversity and Inclusion Fellows
- Presented to Georgia Tech College of Engineering Diversity & Inclusion Council, along with the Dean and Associate Deans
- Analyzed consequences of instituting policy and determined that 50%+ of students with disabilities will be positively affected

Alpha Chi Omega **Atlanta, GA**

Science Outreach Chair *Jan 2020 – Dec 2020*

- Collaborated with Georgia Tech and other STEM organizations in the Atlanta area to host events that celebrate chemistry
- Arranged food drive with local community center to provide underprivileged residents with aid during COVID-19
- Organized tutoring sessions through which fraternity members could support local K-8 students in a myriad of subjects

Save the Water **Surfside, FL**

Project Leader - Research Branch *Apr 2015 – May 2018*

- Guided R&D projects including DILOS Program; coordinated 50+ articles; reduced turnover 25% via training.
- Interviewed and trained newly hired associates on proper process of research leading to a 25% decrease in turnover rate
- Maintained the organization's website by optimizing SEO score for different articles and webpages

Projects

Investigating Actionable Molecules for Biosynthesis *Feb 2023 – Mar 2023*

- Performed extensive market research to identify a feasible molecule the client can produce through synthetic methods
- Conducted market analysis of pterostilbene, the identified molecule, and calculated its growth in the supplement market
- Generated a detailed competitive landscape of the pterostilbene space and actionable strategies to market the supplement
- Delivered actional business strategies to the client and was selected as top 5 teams out of an initial pool of 40 applicants

Marketing Strategies for Integral Molecular *Oct 2022 – Nov 2022*

- Worked with a team of graduate students to create plans for marketing a bispecific antibody therapy for multiple myeloma
- Conducted market research and analyzed competitors in the multiple myeloma space, accounting for clinical development
- Recommended actionable strategies to Wharton professors, industry professionals, and Integral Molecular's CEO

Analyzing ACEs Using NLP Prediction Networks *Dec 2021 – Present*

- Trained an LSTM network on textual input from social media to predict Adverse Childhood Experiences (ACEs)
- Performed cluster analysis to determine similarities between Reddit posts focused on traumatic childhood events
- Generated a graph network with 50+ nodes depicting Twitter users and their likelihood of having experienced an ACE
- Conducted social network analysis (SNA) to determine communities in which external intervention was necessary

Aether Analytics - Anonymous Job Search Service *Apr 2022 – Sep 2022*

- Worked in a team of 4 engineering graduates to pursue a startup venture for improving the job search experience
- Identified a problem of interest based on unmet needs >100 customer interviews including students and companies
- Developed a web platform for hosting users using DJANGO and MongoDB which we iteratively improved via beta testing
- Presented final product to investors and academics at Georgia Tech CREATE-X Demo Day

Department of Education Automated Scoring Challenge *Nov 2021 – Jan 2022*
- Constructed 10+ machine learning models to score constructed response items for the NAEP's reading assessments
- Utilized BERT language models through Pytorch in combination with natural language cleaning processes for classification
- Fulfilled the performance requirements to use automated models in a real-world setting within 5% margin of error
- Accounted for racial and gender disparities within student data using data augmentation to minimize model bias

Spike.io - Diabetes Prevention Application *Oct 2021 – Dec 2021*
- Ideated a solution to aid people with pre-diabetes in mending their lifestyle to prevent onset of diabetes
- Designed a sample user interface for the Spike.io application to present potential use cases
- Developed a business plan for attaining data and marketing Spike.io to a wider audience of health conscious individuals

Predicting Criminal Recidivism Using XGBoost *Jun 2021 – Oct 2021*
- Utilized state of the art machine learning techniques to assist in predicting recidivism to aid in evaluating prison efficiency
- Preprocessed and standardized large datasets to ready them for a XGBoost model with fine-tuned parameters
- Added features to the dataset to ensure that models were not biased against certain demographics

Deriving Actionable Strategies Using Machine Learning *Jan 2021 – Feb 2021*
- Analyzed historical CarMax data to identify trends in customer purchases and preferences across many demographics
- Developed neural network and random forest models to predict customer decisions based on various attributes
- Determined marketing & inventory strategies for CarMax to utilize to draw in distinct segments of customers
- Created an interface which allows personalized experience custom fit for each customer to maximizes their satisfaction

Plaza: Your Local Business Recommender *May 2020 – Sep 2020*
- Crafted a Google Assistant chatbot allows users to local businesses in a specific market and provide support to them
- Uses natural language processing and parts-of-speech identification to recognize and process different user inputs
- Integrated the Google Maps API into the assistant in order to find target businesses and pertinent information about them

Device to Detect Parkinson's Symptoms *Jan 2019 – May 2019*
- Created a device that can measure wrist movements using an IMU to capture the frequency and strength of rest tremors
- Constructed an algorithm which could deduce the presence of a Parkinsonian off-period based on the frequency of tremors
- Presented development process and experimental prototyping to a panel of judges and professors

Skills

- **Software & Cloud:** Google Cloud Platform (GCP) | TensorFlow | PyTorch | SQL | SolidWorks
- **Bioinformatics:** NGS | Structural Variant Calling | Genomic Analysis | Flow Analysis | Microscopy
- **Programming:** Python | R | Git/GitHub | Bash
- **Data Science:** Natural Language Processing (NLP) | Inferential Statistics | Technical Writing
- **Engineering Tools:** 3D Printers | Laser Cutters | Soldering | Workshop Tools (Band saw, Router, Planer)
- **Languages:** English (Native) | French (Intermediate)
- **Certifications:** CITI Training | Certified SOLIDWORKS Professional