JUNSEOK (JUNE) OH

+1-206-538-8452 | juneoh@seas.upenn.edu | linkedin.com/in/joh1020

EDUCATION

University of Pennsylvania — Philadelphia, PA
MSE in Data Science — *(Expected) Aug 2026*

Purdue University — West Lafayette, IN
BSc in Computer Science, Minor in Mathematics — *May 2024*

JOURNALS & PUBLICATIONS

[1] H.C. Cho*, **J. Oh**, S. Jeong (In Preparation)

[2] **J. Oh***, D. Lee*, L.-P. Morency, C. Breazeal, H. Park "Social Egocentric Head Gaze Prediction with Vision Embeddings Fused with Speaker Audio Language" (In Preparation)

[3] H. Min*, H. A. Mina, **J. Oh**, A. J Deering, J. P Robinson, B. Rajwa, E. Bae "Smartphone-integrated optomechanical dual-mode instrument for Salmonella Typhimurium detection" (IEEE Sensors, 2025)

RESEARCH EXPERIENCE

Purdue University, Human-Agent Interaction (HAI) Lab — Dec 2025 – Present
Advisors: Hyung Chan Cho, Prof. Sooyeon Jeong — *(Remote) West Lafayette, IN*

- Project involves curating and validating high-fidelity multimodal datasets to ensure robust performance of Large Language Models (LLMs) in human-agent interaction tasks.
- Will focus on establishing data quality standards and filtering protocols for visual and auditory data streams to minimize noise and improve model grounding.

Massachusetts Institute of Technology, Personal Robots Group — Feb 2025 – Present
Advisors: Dong Won Lee, Dr. Hae Won Park, Prof. Cynthia Breazeal — *(Remote) Cambridge, MA*

- Developed a multimodal transformer model utilizing a spatially grounded cross-attention mechanism to fuse audio and language embeddings (DeepSeek-R1) into visual streams for real-time gaze forecasting.
- Constructed a comprehensive 40+ hour conversation-centric egocentric benchmark by aggregating and curating datasets from Aria, Ego4D, and EgoCom.
- Engineered a scalable preprocessing pipeline utilizing CoTracker to generate proxy head-gaze labels (validated 0.47° MAE against IMU) and integrated WhisperX for automated speaker diarization.
- Built a PyTorch-based framework for distributed training and evaluation of large-scale models across multi-GPU systems, processing over 5 million frames of video data.
- Defined novel heuristic behavioral classes (e.g., Mutual Gaze, Joint Attention) based on speaker-ROI alignment to quantitatively evaluate social interaction dynamics.

Stanford University, Stanford Cardiovascular Institute — Jan 2022 – Aug 2022, Jan 2024 – May 2024
Advisor: Dr. Siyeon Rhee, Prof. Joseph C. Wu — *(Remote) Stanford, CA*

- Engineered an automated scRNA-seq analysis pipeline to process public datasets (GEO), identifying Differentially Expressed Genes (DEGs) between adult and p12 developmental stages.
- Conducted advanced statistical analysis including Gene Set Enrichment Analysis (GSEA) and cell-to-cell interaction profiling to discover key markers in adipogenesis and angiogenesis.
- Expanded research scope by incorporating Epicardial Adipose Tissue (EAT) datasets, successfully integrating multi-source data to uncover latent biological patterns.
- Applied dimensionality reduction algorithms (PCA, UMAP) to visualize high-dimensional single-cell data, facilitating the interpretation of complex tissue development contexts.

Purdue University, Department of Mechanical Engineering — Sept 2022 – Oct 2023
Advisors: Dr. Hyun Jung Min, Prof. Eui-won Bae — *West Lafayette, IN*

- Developed a real-time Android application for bacterial detection, integrating OpenCV for image recognition and a quartz crystal microbalance (QCM) sensor.
- Designed a data pipeline to capture and store experimental data in a local database, enabling real-time chart visualization of frequency and temperature changes using GraphView.

- Enhanced detection accuracy by implementing an automated image capture system triggered by sensor thresholds.

Wittgen Biotechnology, UC Berkeley Skydeck May 2022 – Aug 2022
Research Intern *(Remote) Berkeley, CA*

- Worked on a research project leveraging machine learning techniques for high-resolution tumor classification and tailed drug recommendations and discoveries.
- Analyzed and processed extensive RNA sequence data from over 40 cancer patients to develop cutting-edge ML algorithms.
- Contributed to the creation of an AI-driven platform for cancer heterogeneity profiling, collecting and curating patient data.
- Employed the Seurat library to effectively differentiate cell types in single-cell RNA analysis, enhancing our understanding of cancer at a cellular level.

PROFESSIONAL

Cloud Software Engineer July 2024 – Present
Hewlett Packard Enterprise, Aruba AIOps *San Jose, CA*

- Designed and built an agentic AI system with a Retrieval-Augmented Generation (RAG) framework to automate root cause analysis from internal documentation.
- Deployed generative AI into production network troubleshooting workflows, improving anomaly detection and recommendation accuracy.
- Optimized AIOps data recall through targeted algorithmic improvements, achieving an **83%** reduction for DNS and **91%** for DHCP.

Data Science Intern May 2023 – Oct 2023
Hewlett Packard Enterprise, Aruba AIOps *San Jose, CA*

- Developed and deployed a production ML system integrating Llama-2-7b within a scalable pipeline to enhance a critical business workflow.
- Increased root cause prediction accuracy from **50% to 78%** by designing and implementing an optimized tree-based model with extensive hyperparameter tuning.
- Engineered 38 features from over 2,500 TAC cases and analyzed 1M+ firewall and DNS logs with PySpark to identify root cause patterns.

Completed military service with the Republic of Korea Army (2019 – 2021), honorably discharged as a Sergeant.

TEACHING

Computer Science Department, Purdue University Jan 2023 – May 2023
CS 25200 - Systems Programming (Prof. Gustavo Rodriguez-rivera) *West Lafayette, IN*

- Mentored over 100 students weekly on complex systems concepts including process control, concurrency (pthreads), and memory management (malloc/free).
- Guided students in debugging C and C++ code using GDB and Valgrind to resolve segmentation faults, memory leaks, and race conditions.
- Facilitated lab sessions focused on shell scripting and network programming (TCP/IP, Socket), improving student assignment completion rates.

TECHNICAL SKILLS

Languages: Python, SQL, Java, C/C++
Frameworks: PyTorch, TensorFlow, Spark
Developer Tools: AWS, GCP, Kubernetes, Kafka, Git
Libraries: Scikit-learn, NumPy, Pandas