Weihan Xu

Seattle, WA | wx35@uw.edu | +1 734-548-7537 | https://wx83.github.io/

Education

University of Washington **Seattle, WA**
Doctor of Philosophy in Electrical and Computer Engineering May 2031
Research Directions: Multimodal Sensing and Perception, Generative Models and Embodied Intelligence

Duke University **Durham, NC**
Master of Science in Computer Science (AI/ML track) May 2025
GPA: 3.83/4.0
Thesis: *Alleviating Style Mimicry in Collaborative Human-AI Music Composition*
Coursework: Deep Learning(A+), Theory and Algorithm for Machine Learning(A+), Introduction to Medical Robotics, Natural Language Processing, User Security and Privacy, AI Security and Privacy

University of Michigan-Ann Arbor **Ann Arbor, MI**
Bachelor of Science in Computer Science (with Honors) and Data Science April 2023
GPA: 3.87/4.0
Honors Thesis: *Recurrent Neural Network on Predicting Intensive Care Transfers and Other Unforeseen Events*
Coursework: Deep Learning for Computer Vision, Machine Learning, Applied Regression Analysis, Database, Data Structures and Algorithms, Theoretical Statistics, Probability and Statistics, Linear Algebra, Introduction to Computer Organization, Introduction to Natural Language Processing

Publications

(* denotes Equal Contribution)

Preprints:

2027:

- **ACTS: Learning Visuo-Tactile World Models from Human Physical Interaction**
  Yuxiang Ma*, ***Weihan Xu****, Pengfei Ye, Anna Mei, Ray Yuxin Song, Tingle Li, Haonan Chen, Yilun Du, Edward H Adelson
  Do Robots Need World Models? @ CoRL 2026; Under Review at ICLR 2027
- **AVIO: Learning to Add and Remove Sounding Objects in Audiovisual Scenes**
  Weihan Xu, Kan Jen Cheng, Koichi Saito, Jingyu Shi, Tingle Li, Yisi Liu, Liming Wang, Masato Ishii, Takashi Shibuya, Gopala Anumanchipalli, Paul Pu Liang
  Under Review at ICLR 2027
- **Tac-Bench: Benchmarking Diverse Tactile Manipulation and Multimodal World Models**
  Peilin Wu, Zhiyi Li, ***Weihan Xu***, Haonan Chen, Yilun Du
  Tactile for FM @ RSS 2026; Under Review at ICRA 2027

2026:

- **Schrödinger Audio-Visual Editor: Object-Level Audiovisual Removal**
  Weihan Xu*, Kan Jen Cheng*, Koichi Saito, Muhammad Jehanzeb Mirza, Tingle Li, Yisi Liu, Alexander H. Liu, Liming Wang, Masato Ishii, Takashi Shibuya, Yuki Mitsufuji, Gopala Anumanchipalli, Paul Pu Liang
  Preprint 2026
- **CAVE: Coherent Audio-Visual Emphasis via Schrödinger Bridge**
  Kan Jen Cheng*, ***Weihan Xu****, Koichi Saito, Nicholas Lee, Yisi Liu, Jiachen Lian, Tingle Li, Alexander H. Liu, Fangneng Zhan, Masato Ishii, Takashi Shibuya, Paul Pu Liang, Gopala Anumanchipalli
  Preprint 2026
- **Clips2Story: Training-Free Video Storyboarding and Editing Using Multimodal Retrieval-Embedded Generation**
  Junnuo Wang, ***Weihan Xu***, Hao-Wen Dong
  Preprint 2026

Published:

2025:

- **REGen: Multimodal Retrieval-Embedded Generation for Long-to-Short Video Editing**
  Weihan Xu, Yimeng Ma, Jingyue Huang, Yang Li, Wenye Ma, Taylor Berg-Kirkpatrick, Julian McAuley, Paul Pu Liang, Hao-Wen Dong
  The Thirty-Ninth Annual Conference on Neural Information Processing Systems (NeurIPS)
- **Generating Symbolic Music from Natural Language Prompts using an LLM-Enhanced Dataset**
  Weihan Xu, Julian McAuley, Taylor Berg-Kirkpatrick, Shlomo Dubnov, Hao-Wen Dong
  The 26th International Society for Music Information Retrieval(ISMIR)
- **TeaserGen: Generating Teasers for Long Documentaries**
  Weihan Xu, Paul Pu Liang, Haven Kim, Julian McAuley, Taylor Berg-Kirkpatrick, Hao-Wen Dong
  The Thirteenth International Conference on Learning Representations(ICLR)
- **Video-Guided Text-to-Music Generation Using Public Domain Movie Collections**
  Haven Kim, Zachary Novack, ***Weihan Xu***, Julian McAuley, Hao-Wen Dong
  The 26th International Society for Music Information Retrieval(ISMIR)

2024:

- **A New Dataset for Tag- and Text-Based Controllable Symbolic Music Generation**
  Weihan Xu, Julian McAuley, Taylor Berg-Kirkpatrick, Shlomo Dubnov, Hao-Wen Dong
  The 25th International Society for Music Information Retrieval - Late Breaking Demo
- **A New Dataset, Notation Software, and Representation for Computational Schenkerian Analysis**
  Stephen Hahn, ***Weihan Xu***, Jerry Yin, Rico Zhu, Yue Jiang, Simon Mak, Cynthia Rudin
  The 25th International Society for Music Information Retrieval (ISMIR)
- **Smart Tools, Smarter Concerns: Navigating Privacy Perceptions in Academic Settings**
  Yimeng Ma, ***Weihan Xu***, Hongyi Yin, Yuxuan Zhang, Pardis Emami-Naeini
  The Twentieth Symposium on Usable Privacy and Security (SOUPS Poster)
- **SentHYMNent: An Interpretable and Sentiment-Driven Model for Algorithmic Melody Harmonization**
  Stephen Hahn, Jerry Yin, Rico Zhu, ***Weihan Xu***, Yue Jiang, Simon Mak, Cynthia Rudin
  The 30th SIGKDD Conference on Knowledge Discovery and Data Mining - Applied Data Science Track

2023:

- **Equipping Pretrained Unconditional Music Transformers with Instrument and Genre Controls**
  Weihan Xu, Julian McAuley, Shlomo Dubnov, Hao-Wen Dong
  IEEE Big Data 1st Workshop on AI Music Generation with AI Music Competition

Research Experience

NeuroAI Lab **Seattle, WA**
Advisor: Dr. Eli Shlizerman Jul 2026 – Present

- Designed and implemented an adaptive benchmark for diagnosing multimodal models across perception, event understanding, cross-event reasoning, and goal-conditioned planning using egocentric household videos.
- Built an annotation-grounded question generation pipeline and an examiner–solver–judge framework spanning 49 HD-EPIC videos, with per-skill difficulty adaptation and interactive evaluation tools.
- Conducted controlled history ablations across 10 tasks, identifying planning as a persistent bottleneck for Gemini-class models and Qwen3-Omni even when supplied with correct perceptual context.

MIT Media Lab and MIT CSAIL **Cambridge, MA**
Advisor: Dr. Paul Liang and Dr. Ted Adelson Apr 2025 – Apr 2026

- Built a scalable data pipeline pairing visual observations with object-generated environmental sounds, and fine-tuned a pretrained audiovisual generative model for controllable editing of sounding objects in physical scenes.
- Developed a human-inspired audiovisual highlighting framework that jointly enhances object motion and associated environmental sounds to emphasize cross-modal cues in physical environments. Implemented a flow-based generative model using a Schrödinger bridge, resulting in a co-first-authored paper.

- Developed AVIO, an object-centric audiovisual model for jointly adding and removing objects and their associated sounds, and AVIOBench, a 37.9-hour paired dataset spanning 1,878 target-object names. Enabled optional visual guidance for object appearance and placement control, resulting in a first-authored paper.
- Designed, implemented, and trained an action-conditioned visuo-tactile world model that integrates a top-down camera view, left and right wrist-camera views, and tactile observations from bilateral GelSight sensors to jointly predict visual and tactile dynamics conditioned on motion-and-force actions. Co-developed a 4.9-hour human interaction dataset and evaluated human-to-robot transfer, resulting in a co-first-authored paper.

Professor Julian McAuley's Research Lab **San Diego, CA**
Advisor: Dr. Julian McAuley and Dr. Hao-Wen Dong May 2023 - May 2025

- Developed a system that generates symbolic music aligned with user-specified genre and instrument inputs by fine-tuning a pretrained music transformer with prepended tokens, resulting in a first-authored workshop paper accepted at the 1st Workshop on AI Music Generation at IEEE Big Data 2023 and winning the **AI Music Innovation Award**
- Collected and constructed a large-scale dataset of nearly one million entries by extracting metadata tags from the MuseScore platform and generating captions with a large language model, and successfully built two conditional music generation systems—one that generates symbolic music based on multiple tags and another that generates music using free-form text inputs—resulting in a first-authored paper accepted at ISMIR 2025
- Constructed a dataset of 1,000+ multimodal documentary videos (avg. 30 minutes each) by scraping raw data from YouTube, annotating paired teaser and main videos, and separating audio-visual components. Extracted text via transcription and separated audio tracks into speech, music, and sound effects
- Proposed and implemented an interval-based model and a learning-based model for generating automatic teasers under 3 minutes for long documentaries, and designed new evaluation metrics to complement subjective testing, resulting in a first-author paper accepted at ICLR 2025
- Proposed and implemented a new multimodal retrieval-embedded generation framework that allows an LLM to quote multimodal resources while maintaining a coherent narrative, and designed new evaluation metrics to complement subjective testing, resulting in a first-author paper accepted at NeurIPS 2025

Interpretable Machine Learning Research Lab **Durham, NC**
Advisor: Dr. Cynthia Rudin Dec 2023 - May 2024

- Contributed to the development of a Schenkerian Analysis dataset by annotating music depth and integrating the music data into software for computational recognition
- Proposed a novel clustering approach using a graphical neural network (GNN) that eliminates the need for predefined cluster numbers in each layer, specifically designed for music hierarchical structures
- Developed a codebase from scratch using PyTorch Geometric (PyG) and co-authored an accepted ISMIR paper
- Reimplemented a baseline model with updated libraries, obtaining results for an accepted KDD paper

Professor Sardar Ansari's Research Lab **Ann Arbor, MI**
Advisor: Dr. Sardar Ansari Nov 2021 - May 2023

- Preprocessed raw medical data to develop a Recurrent Neural Network and Transformer-based model to improve accuracy in predicting patient deterioration rates
- Explored replacing an existing XGBoost model with a deep learning approach for time series data, contributing valuable insights into model performance
- Completed an Honor Thesis under the supervision of Professor Sardar Ansari and Professor Kayvan Najarian

Work Experience

OfSpectrum. Inc **Los Angeles, CA**
Machine Learning Engineer Jun. 2025 - Feb. 2026

- Conducting literature reviews and co-brainstorming innovative approaches for audio watermarking
- Designing model architectures for more robust watermarking generators and detectors

AIML. Inc **San Francisco, CA**
Machine Learning Intern Jun. 2025 - Aug. 2025

- Trained generative models to compose short music pieces in specific genres and styles
- Built systems that answer natural language questions about images

- Developed interactive codebooks and articles for demonstration purposes

Teaching Experience

Electrical and Computer Engineering Department, Duke University **Durham, NC**
Course: Introduction to Deep Learning (Taught by Prof. Vahid Tarokh) Sep 2024 - Dec 2024

- Assisted in teaching a graduate-level deep learning course
- Held weekly office hours to guide students on mathematical proofs and coding assignments
- Mentored nine project teams (four members each) through the design, implementation, and presentation of their final projects
- Developed exam questions

Projects

Alleviating Style Mimicry in Collaborative Human-AI Music Composition **Durham, NC**
Master Thesis Project (Advised by Pardis Emami-Naeini) Sep 2024 - May 2025

- Developed a comprehensive survey to investigate human concerns and attitudes toward AI-generated music among individuals from diverse musical backgrounds
- Proposed a perturbation-attack-and-fine-tuning paradigm aimed at mitigating style replication in text-to-music generation models

Diagnostic Classification in Educational Measurement **Ann Arbor, MI**
Data Science Major Capstone (Advised by Prof. Gongjun Xu) Jan 2022 - May 2022

- Proposed a new algorithm to predict psychological attributes with educational assessment data with Cognitive Diagnosis Models package in R
- Overcame computational difficulty of estimating the correlation between answering specific exam questions correctly and psychological attributes with machine learning tools (Restricted Boltzmann Machines)
- Experimented with 4 different methods of determining the correlation matrix, resulting in a research report

Accessibility Tools for the Visually Impaired **Ann Arbor, MI**
Computer Science Major Capstone May 2022 - Jun 2022

- Collectively (in a group of four) built a website that allows visually impaired people to upload photos of surroundings and describes the photos for them
- Designed various features into the website, such as giving warnings if photo includes roadblocks or other obstacles
- Constructed the code and user interface for the website with one other team member
- Tested the website and made a demo describing the function of the website

Leadership & Activities

Shanghai Adream Charitable Foundation **Shanghai, China**
Project: Bridging Education Gaps in Underrepresented Areas Jun 2020 - Oct 2020

- Managed the materials group, coordinating the procurement and organization of 14 types of materials for an educational conference with over 150 expert participants
- Supported the Project Manager in preparing the project proposal and delivering reports to donors
- Analyzed the operational effectiveness of over 500 Adream classrooms in Shandong Province and formulated actionable project plans

Honors & Awards

Departmental Fellowship, University of Washington Sep 2026
E14 EPOCH Winner (Team Lead), E14 Fund Jan 2026
Selected for MIT The Engine Blueprint(Team Lead) Apr 2026

Professional Service

Reviewer: International Journal of Computer Vision(IJCV); NeurIPS; ISMIR; ICLR
Program Committee: NeurIPS 2025 AI for Music Workshop

Skills

Technical skills: C; C++; Python(PyTorch & Tensorflow); Java; SPSS; R; MongoDB; SQL; HTML; CSS; LaTeX

Language: English, Mandarin

Instruments: Piano(Advanced), French horn(Beginner), Violin(Beginner)