You (Neil) Zhang | Teaching Statement

Education shapes the next generation of researchers and practitioners. Throughout my career, I have prioritized fostering a collaborative learning environment and the free exchange of ideas. My approach emphasizes sparking curiosity, building foundational knowledge, and encouraging interdisciplinary thinking. Growing up in a family of educators—both my parents are high school teachers—I developed a deep passion for teaching and was fortunate to learn from exceptional educators. This background has shaped my approach to education and reinforced my commitment to helping students thrive.

Teaching Philosophy

Sparking Interest. From my experiences as both a student and mentor, I have learned that the most effective learning occurs when the subject matter resonates with a student's interests. I aim to foster curiosity by incorporating practical applications and hands-on experiences—such as enabling students to experience their own implemented spatial audio renderings. Pure interest drives active learning through questions and collaboration, deepening engagement and sustaining enthusiasm.

Building a Strong Foundation. A solid grasp of foundational concepts is essential for understanding advanced topics and sustaining intellectual engagement. As a TA for the *Computer Audition* course, I emphasize key foundations in mathematics, signal processing, and machine learning through lectures and office hours. By reinforcing these foundations, I empower students to build knowledge confidently, approach problem-solving creatively, and witness their academic growth.

Encouraging a Broader Perspective . Innovation thrives at the intersection of disciplines. I encourage students to connect ideas from physics, neuroscience, and computer science to advance audio technology. Inspired by Richard P. Feynman's advice — "*Study hard what interests you the most in the most undisciplined, irreverent, and original manner possible*" — I promote interdisciplinary thinking to help students approach challenges critically and from multiple perspectives.

Teaching Experience

Classroom Teaching. Over 9 semesters in the Department of Electrical and Computer Engineering at University of Rochester, **I have served as a graduate teaching assistant for 6 courses**, including *Audio Signal Processing*, *The Art of Machine Learning*, and *Computer Audition*. I received an average teaching assistant rating of 4.63 out of 5 in the AEFIS evaluation system.

Beyond regular responsibilities such as grading assignments and holding office hours, I have **delivered 9 guest lectures** on topics including Python programming for audio, speech technology, neural network training, and machine learning concepts such as support vector machines (SVM), generative adversarial networks (GANs), and my research on audio deepfake detection. I have also **designed 9 homework assignments** that reinforce learning through real-world applications. For instance, one assignment tasks students with using convolutional neural networks (CNNs) to detect COVID-19 from lung ultrasound videos, while another involves AI-based conversational analysis, requiring students to build models that identify who spoke when in recordings. By incorporating practical applications into assignments, I aim to help students develop technical skills and confidence in solving real-world problems while deepening their interest in the subject.

In the *Selected Topics in AR/VR* course, I **worked with students from diverse academic backgrounds**—including schools of education, nursing, and engineering—many of whom had limited programming experience. Teaching advanced concepts like spatial audio rendering to such a varied cohort required a flexible, supportive approach. To lower barriers, I provided structured skeleton code in Jupyter Notebooks as a starting point for assignments and held dedicated office hours tailored to students with minimal coding experience. During these sessions, I guided students step-by-step through key coding logic, clarified concepts, and addressed their individual challenges. This personalized support helped students gain confidence as they successfully completed their implementations and saw their code come to life.

Research Community Teaching (Tutorials). I am dedicated to creating engaging and accessible learning experiences for the research community through tutorials that cover state-of-the-art methods, open challenges, and hands-on activities. For example, **I led a 3-hour tutorial on multimedia deepfake detection** at *IEEE International Conference on Multimedia & Expo (ICME) 2024*, co-organized with field experts, where I presented cutting-edge techniques and key research challenges. At the *AES 6th International Conference on Audio for Games*, **I delivered a 1-hour lecture on personalized spatial audio**, covering HRTF personalization fundamentals and machine learning approaches. Additionally, at the *ASA 187th Meeting*, I

co-presented a 2-hour tutorial on machine learning for acoustics, contributing a 30-minute lecture and interactive Jupyter Notebooks for hands-on exploration. My goal is to equip young researchers or cross-discipline researchers with both foundational knowledge and practical skills, empowering them to drive advancements in audio and machine learning.

Public Education and Outreach. I have delivered invited talks on machine learning and audio research across labs in the USA, Japan, and China, engaging diverse audiences. Through platforms like *ConnectEd* and *SpeechHome* WeChat channels, I shared accessible insights on computer audition and my research. I also presented remote lectures on audio deepfake detection at the Generative AI Spring School and the Global AI Bootcamp in Ukraine. As a mentor in the Lumiere Research Scholar Program, I guided two high school students from Turkey and Egypt on machine learning projects—one focused on detecting temporal inconsistencies in deepfakes, and the other surveyed diagnostic applications of AI in medicine. These experiences reflect my commitment to fostering inclusive, global learning and inspiring students to engage with research.

Mentoring Experience

As I progressed through my PhD, I had the opportunity to **mentor 5 undergraduate and 4 graduate students from diverse backgrounds**, including Audio and Music Engineering, Computer Science, and Data Science. Together, we worked on a range of research projects, including *Audio(-Visual) Deepfake Detection*, *HRTF Personalization for Spatial Audio*, and *Speech Emotion Representation Learning*. These collaborative efforts have resulted in 12 publications [1-12] at prestigious conferences, including IEEE ICASSP, WASPAA, and ISCA Interspeech. Of these, 10 papers feature my students as first authors [1-10].

My advising approach centers on fostering independence while providing targeted support. I involve mentees in all research stages, from experiment design to publication, equipping them with skills for independent research. For example, Siwen Ding, who co-authored a paper with me and presented it at ICASSP [4], initially approached me with a general interest in audio but limited AI research experience. I introduced her to my idea of addressing the generalization issue in audio deepfake detection, and we established clear milestones. I provided her with the necessary research tools, technical guidance, and regular feedback through weekly meetings where we reviewed progress, resolved challenges, and planned the next steps. Siwen reflected that this experience boosted her research confidence, and she later started pursuing PhD at NYU. Among my 6 graduated students, 4 of them joined PhD programs at Lehigh University, UL Lafayette, NYU, and UIUC, respectively.

Several of my mentees have gone on to achieve notable successes, reflecting the impact of my mentoring approach. For example, Yongyi Zang published his first paper with me at Interspeech [5] during his junior year and quickly grew into a semi-independent researcher. He has since led multiple projects and collaborated with other lab members. He co-led the Singing Voice Deepfake Detection Challenge [12] with me and made key contributions as the dataset developers [7, 10].

Future Plan for University Teaching

I plan to develop and teach courses centered around the theme of "Audio and AI", spanning foundational topics in audio signal processing and machine learning to advanced research-oriented applications in audio. My goal is to inspire students to build expertise, embrace diverse perspectives, and contribute to the evolving field of audio and AI research.

Audio AI (Graduate-Level Course). This course will closely align with my research expertise in speech and audio signal processing. Students will explore the design, implementation, and evaluation of AI models for audio applications, such as sound event detection, speech enhancement, generation, and deepfake detection. Through hands-on projects, they will gain practical experience with state-of-the-art tools and engage with recent advancements in the audio research community.

Foundational Courses (Undergraduate- and Graduate-Level Courses). Courses such as *Probability and Random Process*, *Signals and Systems*, *Audio Signal Processing*, and *Machine Learning* will help students build a strong mathematical and technical foundation. These courses align with my teaching philosophy of reinforcing core concepts to empower students to approach complex problems with confidence and creativity.

History and Perspectives in AI (Undergraduate-Level Course). This multidisciplinary course will provide an overview of key milestones in AI, including the Nobel-winning work on neural networks, ImageNet, AlphaGo, ChatGPT, and AlphaFold. By demystifying these advancements and examining their underlying principles, the course aims to foster AI literacy and inspire curiosity. Designed for a broad audience, it will be offered to engineering students and even senior high school students through summer and winter programs, encouraging learners to appreciate how ideas in AI emerge and thrive.

References

[1] Xinhui Chen*, **You Zhang***, Ge Zhu*, and Zhiyao Duan. UR channel-robust synthetic speech detection system for ASVspoof 2021. In *Proc. Automatic Speaker Verification and Spoofing Countermeasures Challenge*, pages 75--82, 2021 (*** equal contribution**).

[2] Abudukelimu Wuerkaixi, Kunda Yan, **You Zhang**, Zhiyao Duan, and Changshui Zhang. DyViSE: Dynamic vision-guided speaker embedding for audio-visual speaker diarization. In *Proc. IEEE 24th International Workshop on Multimedia Signal Processing (MMSP)*, pages 1--6, 2022.

[3] Abudukelimu Wuerkaixi, **You Zhang**, Zhiyao Duan, and Changshui Zhang. Rethinking audio-visual synchronization for active speaker detection. In *Proc. IEEE 32nd International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 01--06, 2022.

[4] Siwen Ding, **You Zhang**, and Zhiyao Duan. SAMO: Speaker attractor multi-center one-class learning for voice anti-spoofing. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.

[5] Yongyi Zang, **You Zhang**, and Zhiyao Duan. Phase perturbation improves channel robustness for speech spoofing countermeasures. In *Proc. Interspeech*, pages 3162--3166, 2023.

[6] Yutong Wen, **You Zhang**, and Zhiyao Duan. Mitigating cross-database differences for learning unified HRTF representation. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 1--5, 2023.

[7] Yongyi Zang*, **You Zhang***, Mojtaba Heydari, and Zhiyao Duan. SingFake: Singing voice deepfake detection. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 12156--12160, 2024 (*** equal contribution**).

[8] Enting Zhou, **You Zhang**, and Zhiyao Duan. Learning arousal-valence representation from categorical emotion labels of speech. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 12126--12130, 2024.

[9] Kyungbok Lee, **You Zhang**, and Zhiyao Duan. A multi-stream fusion approach with one-class learning for audio-visual deepfake detection. In *Proc. IEEE 26th International Workshop on Multimedia Signal Processing (MMSP)*, pages 1--6, 2024.

[10] Yongyi Zang, Jiatong Shi, **You Zhang**, Ryuichi Yamamoto, Jionghao Han, Yuxun Tang, Shengyuan Xu, Wenxiao Zhao, Jing Guo, Tomoki Toda, and Zhiyao Duan. CtrSVDD: A benchmark dataset and baseline analysis for controlled singing voice deepfake detection. In *Proc. Interspeech*, pages 4783--4787, 2024.

[11] **You Zhang**, Fei Jiang, Ge Zhu, Xinhui Chen, and Zhiyao Duan. Generalizing voice presentation attack detection to unseen synthetic attacks and channel variation. In Sébastien Marcel, Julian Fierrez, and Nicholas Evans, editors, *Handbook of Biometric Anti-Spoofing: Presentation Attack Detection and Vulnerability Assessment*, pages 421--443. Springer, 2023.

[12] **You Zhang**, Yongyi Zang, Jiatong Shi, Ryuichi Yamamoto, Tomoki Toda, and Zhiyao Duan. SVDD 2024: The inaugural singing voice deepfake detection challenge. In *Proc. IEEE Spoken Language Technology Workshop (SLT)*, 2024.