Teaching Statement

Yun-Yun Tsai

Teaching Philosophy

As AI tools become increasingly integrated into education and research, I believe computer science education must evolve to help students use these technologies effectively without becoming overly dependent on them. Large language models can now generate code, explain algorithms, and solve technical problems within seconds. While these capabilities can accelerate learning, obtaining a correct answer does not necessarily mean understanding the underlying problem. I believe students should learn to leverage AI as a tool for exploration and feedback while retaining the ability to think critically, ask meaningful questions, and solve problems independently. My teaching philosophy centers on three principles. First, **cultivating curiosity and critical thinking**. I encourage students to ask not only *how* an algorithm works, but also *why* it works, what assumptions it makes, and under what conditions it might fail. Rather than accepting existing solutions at face value, students should learn to question their limitations and investigate alternative approaches. I believe this habit of questioning is essential for developing independent thinkers who can contribute new ideas rather than simply reproduce existing methods. Second, **connecting theoretical understanding with hands-on implementation**. I believe students develop a deeper understanding when they translate mathematical concepts into working code. Through implementation-based assignments, I encourage students to examine individual algorithmic components, debug unexpected behavior, and understand how design choices affect results. When students encounter difficulties, I guide them to break down problems into smaller steps, formulate hypotheses, and test their reasoning through experiments rather than immediately providing solutions. Third, **leveraging AI to enhance learning while fostering independent thinking**. I encourage students to use AI tools to explore unfamiliar concepts, examine alternative approaches, and receive feedback on their work. However, I emphasize that AI-generated solutions should serve as starting points for deeper investigation rather than substitutes for understanding. Students should be able to critically evaluate these solutions, identify underlying assumptions and limitations, and independently verify their correctness. I also encourage students to explore how AI can support personalized and self-directed learning, such as identifying knowledge gaps and generating targeted exercises. Ultimately, my goal is to cultivate students who can **think beyond existing AI capabilities**, formulate original research questions, and develop innovative solutions through their own reasoning and creativity.

Teaching Experience

During my Ph.D. studies at Columbia University, I gained teaching experience in two courses: Engineering Software as a Service (ESaaS) and Computer Vision, working with students from diverse technical backgrounds. In ESaaS, I helped students understand software development principles, service-oriented application design, and programming with Ruby. In Computer Vision, I taught foundational methods in visual perception and introduced modern deep learning approaches, while supporting hands-on assignments in which students implemented and experimented with generative models.

One particularly meaningful experience was designing assignments that required students to implement Generative Adversarial Networks (GANs) and diffusion models from scratch. For the diffusion model assignment, I provided a dataset and Python starter code with structured TODO blocks, guiding students to implement the forward diffusion process, which progressively adds Gaussian noise to images, and the reverse denoising process, in which a neural network learns to predict noise and reconstruct images. A common challenge was understanding how the reverse process could recover meaningful images from random noise and how the training objective relates to the denoising steps. To address this, I explained the mathematical intuition and encouraged students to visualize intermediate noisy and denoised images, helping them identify implementation errors and understand the model's behavior. This experience reinforced my belief that effective teaching should connect theoretical foundations with hands-on experimentation.

- Future Teaching and Course Development

I am prepared to teach fundamental undergraduate courses in computer science, including *Introduction to Algorithms*, *Linear Algebra*, *Artificial Intelligence*, and *Machine Learning*, as well as advanced courses in *Deep Learning*, *Unsupervised Learning*, and *Computer Vision*. I am also interested in developing a new graduate-level course on **Trustworthy and Secure AI Systems**, designed to equip students with the *theoretical foundations and practical skills* to understand emerging *AI safety threats and develop effective safeguards*. As illustrated in Figure 1, this course will progress through three stages. First, students will study the **foundations of adversarial robustness**, including white-box and black-box threat models, adversarial attack methods, and robustness evaluation. Second, they will explore **defense mechanisms** across training and test time, including adversarial training, test-time adaptation, and uncertainty-based detection, while examining how supervised, self-supervised, few-shot, and reinforcement learning can support defense design. Finally, the course will address **LLM and agent safety**, covering jailbreak attacks and defenses, model watermarking, alignment, and safety benchmarking. To bridge theory and practice, students will critically review recent research, implement representative attacks and defenses through hands-on assignments, and conduct red-teaming and safety evaluations. The course will culminate in a research-oriented project in which students investigate real-world LLM or agent vulnerabilities, develop and evaluate potential defenses, and communicate their findings. My goal is to prepare students not only to understand existing AI safety challenges but also to independently identify emerging threats and develop more reliable and secure AI systems.

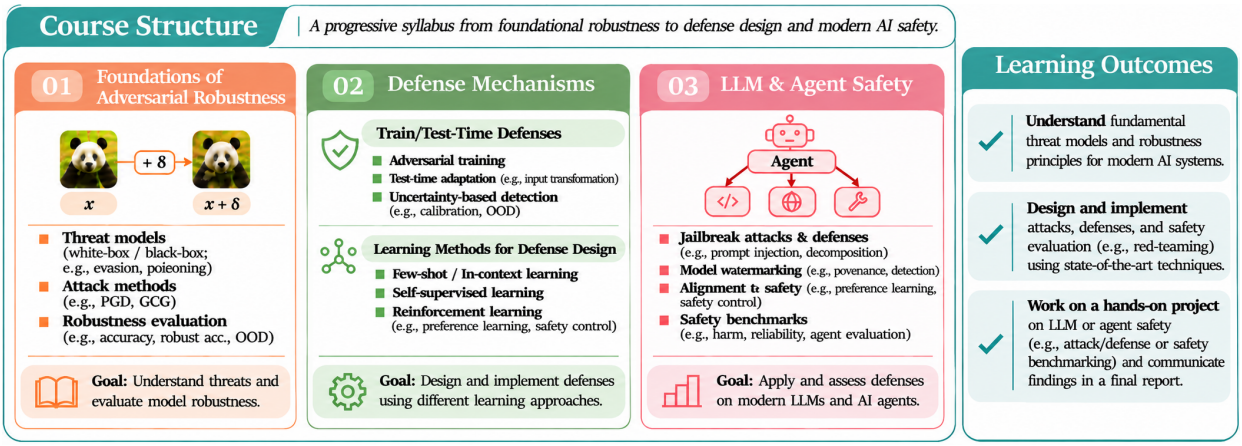


Figure 1: **Proposed graduate course: Trustworthy and Secure AI Systems.**

Research Mentoring Experience

Over the past six years of my Ph.D., I have mentored more than 20 undergraduate and master's students at Columbia University, guiding them from research problem formulation to publication. My mentoring emphasizes **developing research independence through hands-on guidance and gradually increasing responsibility**. Through regular one-on-one meetings, I help students identify research questions, break down technical challenges, design experiments, and interpret results. I also guide them in scientific writing, from articulating technical contributions to presenting experimental evidence and refining manuscripts, while encouraging them to develop their own ideas and take ownership of their research.

My mentoring spans both empirical and theoretical research. For example, I guided **Lei Hsiung** on *compositional adversarial attacks* throughout the entire research process, from initial idea development and implementing algorithms from scratch to experimental evaluation and paper writing, resulting in a first-author publication at **CVPR 2023**. I worked with **Ruize Xu** on *adversarial attacks against AI-generated image detectors*, investigating the robustness of diffusion-based detection methods under black-box and white-box attacks. I also mentored **Qingyuan Liu** on *REVEAL*, focusing on AI-generated video detection and experimental design, and **Jincen Song** on *CAPTCHA reasoning*, emphasizing evaluation methodology and failure analysis. In addition, I worked with **Jiahao Liang** on the theoretical foundations of *RAFP model fingerprinting*, guiding proof development and mathematical analysis. These collaborations have contributed to student-led publications, including work at **CVPR 2023**, the **CVPR 2024 GenAI Workshop**, and **AACL 2026 Findings**. I am particularly proud that Lei and Qingyuan subsequently pursued Ph.D. studies at Dartmouth College and Nanyang Technological University, respectively.

My research internships at **Meta GenAI and Meta Superintelligence Labs** have further shaped my mentoring approach. At Meta, I worked on *robust LLM fingerprinting (RAFP)* for model provenance verification and *WebUIProof and VisRL*, which use execution-based evaluation and reinforcement learning to assess and improve LLM-generated web applications. Together with my experience at **Amazon Lab126** on diffusion-based test-time adaptation, these experiences exposed me to large-scale AI research, interdisciplinary collaboration, and the challenges of translating research ideas into practical systems. I aim to bring these perspectives into my future research group, helping students pursue rigorous and impactful research while developing the technical, communication, and collaboration skills needed for careers in both academia and industry.

Promoting Diversity and Inclusive Research Communities

As a woman in computer science, I recognize the challenges that women and students from underrepresented backgrounds may face in accessing research opportunities, finding mentors, and building professional networks. I am particularly committed to supporting women in computing by creating opportunities for research participation, mentorship, and career development. I believe that fostering diversity requires not only encouraging students to enter the field, but also providing sustained support that helps them build confidence, develop research independence, and pursue their long-term goals.

In my future research group, I aim to create an environment where students feel comfortable asking questions, sharing ideas, and learning from mistakes. Recognizing that students enter research with different levels of preparation, I will provide structured onboarding, individualized guidance, and accessible learning resources to help them develop essential research skills. I also plan to encourage peer mentoring, connecting junior students with more experienced researchers to foster collaboration and provide additional support as they navigate their early research experiences.

Beyond my research group, I plan to organize **mentoring workshops, invited talks by women researchers, and informal networking lunches**, connecting students with researchers from academia and industry to exchange experiences, discuss career paths, and explore research opportunities. I also hope to establish an **AI Security reading group** open to undergraduate and graduate students, featuring regular paper discussions, guest speakers, and opportunities to develop collaborative research projects. Through these initiatives, I aim to strengthen women's participation in AI research, foster supportive mentoring networks, and build an inclusive academic community where students from diverse backgrounds can thrive.