國立台灣科技大學 114學年 第2學期　課程大綱

Spring 2026 NTUST Course Outline

授課教師：鍾昕燁

Instructor:Jhong, Sin-Ye

課程名稱：多媒體系統

Course Title：Multimedia Information Systems

2026/11/3

課程代號：SI5011701 Course Code 學分數：3 Credits	必選修：選修/半學年 Required/Electve:Elective/Half Yr. 先修課程： Prerequisites
節次教室：W6(華夏聚鈺樓H205) W7(華夏聚鈺樓H205) W8(華夏聚鈺樓H205) Time/Location	
專業核心能力： Core Professional Competencies	■ 具備解決工程與管理問題之能力 ■ 專業知識(Comprehensive management knowledge) ■ 研究議題之能力(Research capability in practical discipline)
課程網址： Course Website	
課程宗旨： Course Objectives	This course diverges from traditional multimedia systems curricula by focusing exclusively on the rapidly evolving field of Multimodal Machine Learning (MMML). Adopting a problem-based learning (PBL) framework, the course is built around reading, implementing, and analyzing the most recent state-of-the-art AI methodologies published between 2024 and 2026. Students will deeply explore the mathematical foundations and computational models required to integrate linguistic, acoustic, and visual data streams. Rather than surveying basic multimedia formats, scholars will tackle core MMML challenges, including heterogeneous representation, cross-modal alignment, reasoning, and multimodal generation. As an advanced-level course, it assumes that students already possess a strong theoretical background in deep learning alongside practical coding capabilities in Python and PyTorch. The curriculum requires participants to design and optimize complex multimodal systems, bridging the gap between recent academic breakthroughs and practical applications. Enrolled students must have access to a high-end GPU with at least 24GB of VRAM to handle the intensive training, fine-tuning, and deployment demands of modern architectures. By engaging directly with recent literature and conducting hands-on implementations, participants will develop the specialized engineering and research skills necessary to innovate within advanced artificial intelligence.
課程大綱： Outline of Lectures	

Week 1: Course Introduction & Logistics
Week 2: Datasets & Benchmarks: Exploring MMML tasks, baseline guidelines, and establishing project scopes.
Week 3: Unimodal Representations: Mathematical foundations for linguistic, acoustic, and visual data pipelines.
Week 4: Multimodal Representations I: Modeling cross-modal interactions utilizing early, late, and hybrid fusion.
Week 5: Multimodal Representations II: Designing coordinated representations and contrastive learning paradigms.
Week 6: Multimodal Alignment: Explicit alignment algorithms, dynamic time warping, and precise cross-modal grounding.
Week 7: Self-Attention & Transformers: Implementing state-of-the-art masked multimodal architectures for learning.
Week 8: Midterm Project Presentations & Demo (I)
Week 9: Midterm Project Presentations & Demo (II)
Week 10: Multimodal Reasoning I: Structured and hierarchical inference paradigms for interconnected data streams.
Week 11: Multimodal Reasoning II: Integrating external knowledge graphs and modeling logical causal relationships.
Week 12: Multimodal Generation I: Generative processes for modality translation and cross-modal summarization.
Week 13: Multimodal Generation II: Implementing advanced generative models, including state-of-the-art diffusion.
Week 14: Transference & Quantification: Managing cross-modal co-learning, dataset biases, and model robustness.
Week 15: Final Project Presentations & Demo (I)
Week 16: Final Project Presentations & Demo (II)

授課方式： Method of Instruction	講授 Lecture：50% 分組討論 Group discussion：0% 案例研討 Case study：25% 操做練習 Practical exercises：25% 講授 Lecture：This course uses a Problem-Based Learning (PBL) model centered on project work. The format includes lectures, regular student presentations on MMML paper research, and major midterm/final projects. You must independently train and fine-tune multimodal AI models; evaluations heavily weigh model performance and efficiency. CRITICAL: A high-spec PC equipped with a GPU containing at least 24GB of VRAM is strictly mandatory and is essential for your final grade.%
教科書： Textbooks	Please be explicitly aware that as an advanced, research-oriented course, there is no single textbook that covers this curriculum. The primary pedagogical materials will consist exclusively of state-of-the-art (SOTA) Multimodal Machine Learning (MMML) papers published between 2024 and 2026 in top-tier venues (e.g., CVPR, ACL, NeurIPS, ICML, and IEEE Transactions).
參考書目： References	[1] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal Machine Learning: A Survey and Taxonomy," IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 2, pp. 423-443, Feb. 2019. [2] P. Liang, A. Zadeh, and L.-P. Morency, "Foundations & trends in multimodal machine learning: Principles, challenges, and open questions," ACM Comput. Surv., vol. 56, no. 10, Art. no. 264, pp. 1–42, Oct. 2024.
修課須知： Notice	Note for Week 1 (February 25th): Attendance is strictly mandatory for all enrolled students. HW0 will be assigned during this session, directly accounting for 15% of your overall course grade. This preliminary assignment is specifically designed to rigorously verify your practical engineering skills in Python, ROS, Git, Docker, and PyTorch. Furthermore, it will conclusively confirm that you possess the required high-end GPU hardware resources to succeed in this intensive multimodal course.
評量方式： Grading	

	1. Participation & Paper Reports (35%): Includes class engagement and four MMML paper implementation reports. HW0 accounts for 15% of the overall course grade. 2. Midterm Project (30%): Strictly evaluated on a Live Demo (20%), Performance Metrics (20%), a 20-min Oral Presentation (30%), and a Written Report (30%). 3. Final Project (35%): Strictly evaluated on a Live Demo (20%), Performance Metrics (20%), a 20-min Oral Presentation (30%), and a detailed Written Report (30%).
備註說明： Notes	This advanced course strictly requires prior knowledge of deep learning and digital image processing. High proficiency in Python, Git, Docker, and the PyTorch framework is expected. Students must possess a high-end GPU with a minimum of 24GB VRAM for model training. Each scholar will present and implement four top-tier MMML conference or journal papers. Evaluations emphasize performance, efficiency, and optimization. Expect to allocate a minimum of 20 hours per week for this heavy workload.