Form AR 140

The Hong Kong Polytechnic University

Subject Description Form

Subject Code	CMS6001
Subject Title	Advanced AI Programming
Credit Value	3
Level	6
Pre-requisite/ Co-requisite/ Exclusion	- Background in algorithms, data structures, linear algebra, and probability and statistics. - Prior exposure to artificial intelligence or machine learning. - Proficiency in Python programming and development tools.
Objectives	This subject is designed for doctoral students seeking to transition from being consumers of AI models to producers of novel AI research and high-level systems. Moving beyond foundational concepts, the subject provides a rigorous, hands-on exploration of the architectural design and expert-level implementation of state-of-the-art artificial intelligence. Students will deconstruct and program complex models such as Transformers and Diffusion Models using modern deep learning frameworks. A significant emphasis is placed on the engineering challenges of large-scale AI. Crucially, the curriculum intertwines this technical depth with a critical examination of trustworthy AI, focusing on the practical implementation of explainability techniques and the algorithmic approaches to ensuring fairness and mitigating bias. The subject culminates in a significant research project where students will formulate a novel problem, develop a sophisticated AI-driven solution, and justify their work, preparing them for leadership roles in academic research and advanced industrial R&D. The objectives of the subjects are as follows. 1. To cultivate expertise in designing, programming, and deploying complex, state-of-the-art AI systems using cutting-edge frameworks and methodologies. 2. To enable students to critically evaluate current AI research and develop the skills to formulate, conduct, and contribute novel work to the field. 3. To instil a deep understanding of the ethical, interpretability, and societal challenges of AI, ensuring graduates can lead the development of trustworthy and fair systems.

<table>
<tr>
<td>Intended Learning Outcomes</td>
<td>Upon completion of the subject, students will be able to:

a. Design, program, and rigorously evaluate sophisticated AI models (e.g., advanced deep learning architectures, generative models) to solve complex, real-world problems.
b. Formulate a unique research question, conduct a literature review, and develop a proposal for a novel AI methodology or application, communicating the findings effectively.
c. Critically assess the ethical implications and potential biases of an AI system and apply advanced techniques to improve its transparency and interpretability.</td>
</tr>
<tr>
<td>Subject Synopsis/ Indicative Syllabus</td>
<td>The syllabus is structured into four core modules, each building upon the last. The approach combines theoretical understanding with intensive, practical implementation.

<u>Module 1: Foundations of Advanced AI Engineering</u>
This module sets the stage by focusing on the professional tools and large-scale engineering practices required to build robust and scalable AI systems, including

• Advanced Framework Programming. Deep dive into PyTorch or TensorFlow APIs beyond the basics. Custom layers, hooks, computation graphs, and just-in-time (JIT) compilation. Principles of writing modular, reusable, and extensible AI code.

• High-Performance and Distributed Training. Fundamentals of GPU architecture (CUDA basics). Data parallelism vs. model parallelism. Introduction to distributed training frameworks (e.g., PyTorch's DistributedDataParallel, Horovod). Strategies for massive datasets and models that don't fit into single-GPU memory.

• MLOps and Reproducible Research. Principles of Machine Learning Operations (MLOps). Tools for experiment tracking (e.g., Weights & Biases), data and model versioning (e.g., DVC), and containerization (Docker) for reproducible research environments.

<u>Module 2: State-of-the-Art Model Architectures</u>
This module explores the theory and implementation of the dominant AI paradigms that have defined the current state of the art, including:

• Transformers and Attention Mechanisms. The original Transformer architecture (including Self-attention, multi-head attention, and positional encodings). Architectural variants like BERT, GPT, and the Vision Transformer (ViT).

• Advanced Generative AI. The theory and common failure modes of Generative Adversarial Networks (GANs). Advanced architectures like StyleGAN. Introduction to Variational Autoencoders (VAEs) and Diffusion Models.

• Advanced Reinforcement Learning. Moving beyond basic Q-learning to policy gradient methods (e.g., REINFORCE, A2C). Introduction to modern algorithms like PPO, DPO, and GRPO. The role of RL in training large language models (RLHF).

<u>Module 3: Trustworthy and Explainable AI (XAI)</u></td>
</tr>
</table>

<table>
<tr>
<td></td>
<td>
This module addresses the critical need for transparency, fairness, and ethical considerations in AI systems, including:

Foundations of Explainable AI (XAI). The "black box" problem. Taxonomy of XAI methods: local vs. global, model-specific vs. model-agnostic. Introduction to key techniques like LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations).
Implementing and Evaluating Interpretability. Gradient-based attribution methods for neural networks (e.g., Grad-CAM). Critically evaluating the fidelity and reliability of explanations. The trade-off between model performance and interpretability.
Algorithmic Fairness, Bias, and Ethics. Defining and quantifying fairness (e.g., demographic parity, equalized odds). Sources of bias in the AI pipeline (data, algorithm, human interpretation). Algorithmic debiasing techniques (pre-processing, in-processing, post-processing). Ethical frameworks for AI development.

<u>Module 4: Research Formulation and Application</u>
This final module transitions students toward their final project, focusing on the skills needed to conduct original research with Research Proposal and Peer Review. It includes: How to conduct a systematic literature review and identify a research gap. Formulating a clear problem statement, hypothesis, and methodology. Structuring a doctoral-level research proposal.
</td>
</tr>
<tr>
<td>Teaching/Learning Methodology</td>
<td>
Lectures and Tutorials provide foundational knowledge teaching. The instructor presents the core theories, mathematical underpinnings, and architectural concepts of advanced topics.

Labs provide hands-on guidance to enable students translate theory into practice by implementing complex AI models.

Group Discussions will focus on developing critical research skills through the peer review of code and research proposals, training students to articulate and defend their work effectively.
</td>
</tr>
<tr>
<td>Assessment Methods in Alignment with Intended Learning Outcomes</td>
<td>
<table>
<tr><td rowspan="2">Specific assessment methods/tasks</td><td rowspan="2">% weighting</td><td colspan="3">Intended subject learning outcomes to be assessed (Please tick as appropriate)</td></tr>
<tr><td>a</td><td>b</td><td>c</td></tr>
<tr><td>1. Assignment and Quizzes</td><td>50%</td><td>√</td><td></td><td>√</td></tr>
<tr><td>2. Project and Presentation</td><td>50%</td><td>√</td><td>√</td><td>√</td></tr>
<tr><td>Total</td><td>100 %</td><td colspan="3"></td></tr>
</table>

Explanation of the appropriateness of the assessment methods in assessing the intended learning outcomes:
</td>
</tr>
</table>

	The assessment methods are appropriate because they are strategically aligned to measure distinct outcomes in a scaffolded manner. Assignments and quizzes directly target ILO (a) by testing the hands-on programming of models and ILO (c) by requiring the application of specific ethical analysis tools. The final project is then uniquely appropriate for assessing the synthesis of all three ILOs, as it requires students to demonstrate mastery of the complete research cycle in ILO (b) (formulating a novel question and communicating findings) while applying the advanced implementation and critical evaluation skills from ILOs (a) and (c).	
Student Study Effort Expected	Class contact:	
	▪ Lecture/Tutorials/Labs	39 Hrs.
	Other student study effort:	
	▪ Self-Study and After-class Reading, Assignments and Projects, Preparations for Quizzes and Presentations.	83 Hrs.
	Total student study effort	122 Hrs.
Reading List and References	1. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. 2. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30. 3. Russell, S. J., & Norvig, P. (2021). Artificial Intelligence: A Modern Approach (4th ed.). Pearson. 4. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10684–10695). IEEE. 5. O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown. 6. Bishop, C. M. (2024). Deep Learning: Foundations and Concepts. Springer. 7. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30. 8. Barocas, S., Hardt, M., & Narayanan, A. (2023). Fairness and Machine Learning: Limitations and Opportunities. MIT Press.	