The Hong Kong University of Science and Technology

UG Course Syllabus

AI Processor Architecture

ELEC 4350

3 Credits

Pre-requisites: ELEC 2350 or COMP 2611

Name: Fengbin TU

Email: fengbintu@ust.hk

Office Hours: By email appointments

Course Description

Artificial Intelligence (AI) techniques have achieved great success in a wide range of applications like computer vision, natural language processing, and scientific computing. Traditional processors are not optimized for AI tasks, which can result in slow performance and high energy consumption. To unlock the full potential of AI, both academia and industry have developed many AI processors for efficient AI computing from edge to cloud, with specialized architecture for the complex computations of AI applications.

This is an introductory course to advanced processor architecture for AI computing. The topics covered include AI algorithm basics, processing element, dataflow, memory system, software-hardware co-design for AI processors. This course will also introduce benchmarking and recent advances of AI processors.

Intended Learning Outcomes (ILOs)

By the end of this course, students should be able to:

ILO1: Understand the AI computing basics and challenges.

ILO2: Understand the programming and dataflow of AI processors.

ILO3: Understand the hardware components of AI processors.

ILO4: Survey and present at least one of the major AI processor technology trends.

ILO5: Acquire the typical methods to design and evaluate AI processor performance.

Assessment and Grading

This course will be assessed using criterion-referencing and grades will not be assigned using a curve. Detailed rubrics for each assignment are provided below, outlining the criteria used for evaluation.

Assessments:

Assessment Task	Contribution to Overall Course grade (%)	Due date
Written assignment	20%	09/04/2026
In-class quiz	20%	23/04/2026
Lab experiments	20%	14/05/2026
Project report	20%	29/05/2026
Presentation	20%	07/05/2026

* Assessment marks for individual assessed tasks will be released within two weeks of the due date.

Mapping of Course ILOs to Assessment Tasks

Assessed Task	Mapped ILOs	Explanation
Written assignment	ILO1, ILO2, ILO3	This task assesses students' ability to explain and apply AI computing basics and challenges (ILO1), understand and describe programming and dataflow in AI processors (ILO2), and identify and analyze hardware components of AI processors (ILO3).
In-class quiz	ILO1, ILO2, ILO3	This task assesses students' ability to catch up with the lecture materials that cover the basics of AI processors (ILO1, ILO2, ILO3).
Lab experiments	ILO2, ILO3	This task enriches and assesses students' knowledge in programming and dataflow of AI processors (ILO2), and hardware components of AI processors (ILO3).
Project report	ILO1, ILO2, ILO3, ILO4, ILO5	This task assesses students' ability to demonstrate understanding of AI computing basics and challenges (ILO1), programming and dataflow (ILO2), hardware components (ILO3), survey and integrate major AI processor technology trends (ILO4), and apply methods to design and evaluate AI processor performance (ILO5), through a comprehensive written analysis.
Presentation	ILO1, ILO2, ILO3, ILO4, ILO5	This task assesses students' ability to articulate AI computing basics and challenges (ILO1), explain programming and dataflow (ILO2), describe hardware components (ILO3), present a survey of major AI processor technology trends (ILO4), and demonstrate methods for designing and evaluating AI processor performance (ILO5), showcasing communication and synthesis skills.

Grading Rubrics

Written Assignment Grading Rubric

Marks	**4**	**3**	**2**	**1**	**0**
Understanding of Concepts Max: 8	Demonstrates comprehensive understanding of AI computing basics, challenges, programming/ dataflow, or hardware components, with insightful analysis and accurate application. (8)	Shows solid understanding of key concepts, with clear explanations and minor gaps in application. (6)	Displays basic understanding of concepts, but with some inaccuracies or incomplete explanations. (4)	Shows limited understanding of concepts, with frequent errors or superficial coverage. (2)	Lacks understanding of concept, with major errors or irrelevant content. (0)
Analysis and Justification Max: 8	Provides thorough analysis with strong justification, using evidence from course materials to support claims effectively. (8)	Offers good analysis with reasonable justification and some supporting evidence. (6)	Includes basic analysis with partial justification, but lacks depth or sufficient evidence. (4)	Presents minimal analysis with weak justification and little evidence. (2)	No analysis or justification provided. (0)
Clarity and Organization Max: 4			Writing is clear, well-organized, and logically structured within the word limit. (4)	Writing is mostly clear and organized, with minor issues in structure. (2)	Writing is unclear and poorly organized. (0)

In-class Quiz Grading Rubric

Marks	**4**	**3**	**2**	**1**	**0**
Understanding of Concepts Max: 8	Demonstrates comprehensive understanding of AI computing basics, challenges, programming/dataflow, or hardware components, with insightful analysis and accurate application. (8)	Shows solid understanding of key concepts, with clear explanations and minor gaps in application. (6)	Displays basic understanding of concepts, but with some inaccuracies or incomplete explanations. (4)	Shows limited understanding of concepts, with frequent errors or superficial coverage. (2)	Lacks understanding of concept, with major errors or irrelevant content. (0)

Analysis and Justification Max: 8	Provides thorough analysis with strong justification, using evidence from course materials to support claims effectively. (8)	Offers good analysis with reasonable justification and some supporting evidence. (6)	Includes basic analysis with partial justification, but lacks depth or sufficient evidence. (4)	Presents minimal analysis with weak justification and little evidence. (2)	No analysis or justification provided. (0)
Clarity and Organization Max: 4			Writing is clear, well-organized, and logically structured within the word limit. (4)	Writing is mostly clear and organized, with minor issues in structure. (2)	Writing is unclear and poorly organized. (0)

Lab Experiments Grading Rubric

Marks	**4**	**3**	**2**	**1**	**0**
Result Quality and Functionality Max: 8	Lab results demonstrate comprehensive and fully functional outcomes with accurate and insightful performance relevant to AI processors. (8)	Lab results show solid functionality with mostly accurate outcomes and minor issues. (6)	Lab results demonstrate basic functionality but with partial accuracy and noticeable errors. (4)	Lab results repeat basic course examples with minimal functionality and significant errors. (2)	No functional results or outcomes are demonstrated. (0)
Code Implementation and Justification Max: 6	Code is well-structured, efficient, and fully justified with clear comments and no errors. (6)	Code is adequately structured with reasonable efficiency and partial justification, minor errors in implementation. (4)	Code provides basic implementation but lacks structure, efficiency, or justification, with noticeable errors. (2)	Code shows minimal implementation with poor structure and limited justification, frequent errors. (1)	No code provided or it is non-functional and unjustified. (0)
Documentation and Submission Max: 6	Report provides comprehensive, well-organized documentation including detailed explanations, code snippets,	Report documents the lab clearly with good organization, including most key elements but minor	Report includes adequate descriptions with some organization, but lacks depth in analysis and	Report provides basic or superficial documentation with significant disorganizati	No report or documentation provided, or it is incoherent with no evidence submitted. (0)

	diagrams, results analysis, and complete evidence submission. (6)	omissions in analysis or evidence. (4)	partial evidence submission. (2)	on and incomplete evidence. (1)	

Project Report Grading Rubric

Marks	**4**	**3**	**2**	**1**	**0**
Critical Analysis and Justification Max: 8	Provides deep critical analysis with strong justification. Synthesizes literature effectively. Arguments are strictly evidence-based. (8)	Offers good analysis based on literature. Clearly explains key concepts and evaluates trends, though some critical evaluation may lack depth. (6)	Analysis is present but largely descriptive rather than critical and lacks depth in evaluating the "why"and "how". (4)	Presents weak analysis with limited critical engagement. Descriptions are superficial or contain inaccuracies. (2)	No meaningful analysis, justification, or critical engagement with the literature. (0)
Original Insights and Synthesis Max: 6		Extends literature with original synthesis, novel insights on AI processor trends, and creative connections across sources. (6)	Integrates trends with some original extensions or thoughtful synthesis beyond basic summary. (4)	Basic summary of one or more trends with minimal original insight. (2)	Lacks originality, synthesis, or meaningful integration of trends. (0)
Report Quality and Completeness Max: 6		Report is comprehensive, error-free, well-structured, with strong synthesis, clear writing, illustrative visuals, and proper referencing. (6)	Report is mostly complete and coherent with effective logic flow. Writing is clear with only minor errors and correct reference. (4)	Covers main points but has noticeable gaps in structure or logic. Writing is readable but may contain grammatical errors. (2)	Report is incomplete, poorly written, or fails to address key elements. (0)

Presentation Grading Rubric

Marks	**4**	**3**	**2**	**1**	**0**
Content Coverage Max: 8	Presentation fully covers AI basics, dataflow, hardware, trends, and evaluation	Covers key elements with good depth and	Includes basic coverage but lacks depth in	Limited coverage with superficial content. (2)	Fails to cover essential content. (0)

	with insightful depth. (8)	minor omissions. (6)	some areas. (4)		
Clarity and Organization Max: 8	Clear, well-organized delivery with effective visuals and logical flow within time limit. (8)	Mostly clear and organized with some effective elements. (6)	Somewhat organized but with unclear sections. (4)	Poorly organized and unclear. (2)	Disorganized and incoherent. (0)
Engagement and Justification Max: 4			Engages audience with strong justification and original insights on trends. (4)	Basic engagement with partial justification. (2)	Minimal engagement or justification. (0)

Final Grade Descriptors:

Grades	Short Description	Elaboration on subject grading description
A	Excellent Performance	Demonstrates a comprehensive grasp of AI computing basics, challenges, programming, dataflow, hardware components, technology trends, and design/evaluation methods (ILO1-ILO5). Exhibits expertise in analyzing and optimizing AI processor architectures, significant creativity in applying concepts to real-world scenarios, and goes beyond core requirements through innovative projects and presentations.
B	Good Performance	Shows good knowledge and understanding of AI computing fundamentals, programming/dataflow, hardware, trends, and performance evaluation (ILO1-ILO5). Displays competence in problem-solving for AI processor designs, the ability to analyze and evaluate architectural choices, and high motivation to collaborate effectively in course activities.
C	Satisfactory Performance	Possesses adequate knowledge of core AI processor concepts, including basics, dataflow, hardware, trends, and evaluation methods (ILO1-ILO5). Demonstrates competence in handling familiar AI computing problems, some capacity for analysis and critical thinking in processor design, and persistence in achieving broadly defined learning goals.
D	Marginal Pass	Has threshold knowledge of AI computing basics, programming/dataflow, hardware components, trends, and design/evaluation (ILO1-ILO5). Shows potential to develop skills in surveying AI processor technologies and making basic judgments on performance, benefiting from the course with room for growth in the discipline.
F	Fail	Demonstrates insufficient understanding of AI processor architecture fundamentals, including computing challenges, dataflow, hardware, trends, and evaluation methods (ILO1-ILO5). Lacks necessary problem-solving skills for AI designs, shows limited critical or analytical thinking, and exhibits minimal effort

		toward achieving learning goals, failing to meet threshold requirements for the field.

Course AI Policy

The use of Generative AI in project is permitted with proper acknowledgement and will NOT be contributed to the students' work.

Communication and Feedback

Assessment marks for individual assessed tasks will be communicated via Canvas within two weeks of submission. Feedback on assignments will include [specific details, e.g., strengths, areas for improvement]. Students who have further questions about the feedback including marks should consult the instructor within five working days after the feedback is received.

Resubmission Policy

No resubmission policy.

Required Texts and Materials

No textbook.

Here are recommended/supplementary learning materials:

1. V. Sze, Y.-H. Chen, T.-J. Yang, J. S. Emer, Efficient Processing of Deep Neural Networks, Synthesis Lectures on Computer Architecture, Morgan & Claypool Publishers, 2020

2. B. Reagen, R. Adolf, P. Whatmough, G.-Y. Wei, D. Brooks, Deep Learning for Computer Architects, Synthesis Lectures on Computer Architecture, Morgan & Claypool Publishers, 2017

3. B. Moons, D. Bankman, M. Verhelst, Embedded Deep Learning: Algorithms, Architectures and Circuits for Always-on Neural Network Processing, Springer, 2018

Academic Integrity

Students are expected to adhere to the university's academic integrity policy. Students are expected to uphold HKUST's Academic Honor Code and to maintain the highest standards of academic integrity. The University has zero tolerance of academic misconduct. Please refer to Academic Integrity | HKUST – Academic Registry for the University's definition of plagiarism and ways to avoid cheating and plagiarism.