National University of Engineering (UNI)
School of Computer Science
Syllabus 2027-I

1. COURSE

AI362. AI for edge computing (Elective)

2. GENERAL INFORMATION

2.1 Course	:	AI362. AI for edge computing
2.2 Cycle	:	CS300 Cycle.
2.3 Credits	:	3
2.4 Hours	:	2 HT; 4 HP;
2.5 Duration of the period	:	16 weeks
2.6 Type of course	:	Elective
2.7 Learning modality	:	Face to face
2.8 Prerrequisites	:	AI264. Deep Learning. (7^{th} Sem)

3. PROFESSORS

Meetings after coordination with the professor

4. INTRODUCTION TO THE COURSE

This course covers AI model deployment on edge devices (IoT, mobile, embedded), addressing optimization techniques, specialized frameworks, and privacy/security considerations in distributed systems.

5. GOALS

- Implement AI models optimized for edge hardware
- Design distributed inference pipelines
- Evaluate accuracy/latency/energy trade-offs

6. COMPETENCES

AG-C02) Ethics: Applies ethical principles and commits to professional ethics and the standards of professional computing practice. (Usage)

AG-C09) Design and Development of Solutions: Designs, implements, and evaluates solutions for complex computing problems. (Assessment)

AG-C12) Uses systemic approaches to select, develop, apply, integrate, and manage secure computing technologies to achieve user objectives. (Assessment)

7. TOPICS

Unit 1: Edge AI Fundamentals (12 hours)	
Competences Expected: AG-C09	
Topics	**Learning Outcomes**
• Hardware architectures for edge computing • Edge-friendly model taxonomies (MobileNets, TinyML) • Deployment frameworks (TF Lite, ONNX Runtime)	• Select models matching hardware constraints [Evaluar] • Profile energy consumption on real devices [Usar]
Readings : [HL22]	

Unit 2: Optimization Techniques (24 hours)	
Competences Expected: AG-C09	
Topics	**Learning Outcomes**
• Quantization and pruning • Knowledge distillation • Specialized compilation (TVM, Apache MXNet)	• Shrink models for microcontrollers [Evaluar] • Implement hybrid (edge+cloud) pipelines [Usar]
Readings : [How+17], [SW21]	

Unit 3: Security & Privacy (8 hours)	
Competences Expected: AG-C09	
Topics	**Learning Outcomes**
• Federated learning at edge • Homomorphic encryption • Edge model attacks (adversarial examples)	• Analyze edge system vulnerabilities [Familiarizarse] • Implement basic defenses [Usar]
Readings : [HL22]	

8. WORKPLAN

8.1 Methodology

Individual and team participation is encouraged to present their ideas, motivating them with additional points in the different stages of the course evaluation.

8.2 Theory Sessions

The theory sessions are held in master classes with activities including active learning and roleplay to allow students to internalize the concepts.

8.3 Practical Sessions

The practical sessions are held in class where a series of exercises and/or practical concepts are developed through problem solving, problem solving, specific exercises and/or in application contexts.

9. EVALUATION SYSTEM

Theory Sessions:
The theory sessions are held in master classes with activities including active learning and roleplay to allow students to internalize the concepts.

Lab Sessions:
In order to verify their competences, several activities including active learning and roleplay will be developed during lab sessions.

Oral Presentations:
Individual and team participation is encouraged to present their ideas, motivating them with additional points in the different stages of the course evaluation.

Reading:
Throughout the course different readings are provided, which are evaluated. The average of the notes in the readings is considered as the mark of a qualified practice. The use of our virtual campus allows each student to access the course information, and interact outside the classroom with the teacher and with the other students.

The final grade NF depends on the following grades:

- PE: Partial Exam.
- FE: Final Exam.
- P: Permanent Evaluation (Average of practical work and graded assignments).

The final grade NF is calculated as follows:

$$NF = \frac{PE + FE + P}{3}$$

10. BASIC BIBLIOGRAPHY

[How+17] Andrew Howard et al. "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications". In: *arXiv:1704.04861* (2017).

[SW21] Vikram Saraph and Pete Warden. "TinyML: Enabling Always-On ML on the Edge". In: *ACM HotNets*. 2021.

[HL22] Song Han and Yujun Lin. *Edge AI: Systems Design for Embedded Machine Learning*. Springer, 2022.