

# PHORK NORAK

Data Science and Engineering Graduate | Prospective Integrated MS-PhD Student, Fall 2027

Phone: +855 66 362 958 | Email: phorknorak2020@gmail.com | Phnom Penh, Cambodia

Portfolio: [phorknorak.com](https://phorknorak.com) | LinkedIn: [linkedin.com/in/phork-norak-410592282](https://linkedin.com/in/phork-norak-410592282)

GitHub: [github.com/PhorkNorak](https://github.com/PhorkNorak) | Dataset: [Hugging Face](#) | ResearchGate: [report](#)

## EDUCATION

**Bachelor of Engineering, Data Science and Engineering**

2022 - 2026

Royal University of Phnom Penh | Graduated 2026

**CGPA: 3.38/4.00 | Year 4, Semester 2 GPA: 4.00/4.00**

Selected coursework: Applied Machine Learning; Natural Language Processing; LLMs and Prompt Engineering; Ethics and Responsible Data Science.

## RESEARCH INTERESTS

Big Data Integration, LLMs, trustworthy retrieval-augmented generation, and Khmer NLP.

## RESEARCH OUTPUTS

**Undergraduate research report:** *Explainable AI for Automated Grading of Khmer Language Short-Answer Questions in Education* (RUPP, 2026); dataset v1.0.1 (CC BY-NC 4.0); code v1.1.0 (MIT).

[Hugging Face dataset](#) | [Kaggle dataset](#) | [ResearchGate report](#) | [GitHub code](#) | [Model lineage](#)

## SELECTED RESEARCH AND EXPERIENCE

**Explainable AI for Automated Grading of Khmer Language Short-Answer Questions in Education**

2026

*Undergraduate research | RUPP | Supervisor: Dr. Khim Chamroeun*

- **Problem:** manual grading of Khmer short answers is slow and inconsistent; Khmer lacks graded-answer data.
- **My contribution:** created the **Khmer ASAG Benchmark** of **1,184 teacher-graded answers from 203 students at two schools** (41 questions, four subjects); fine-tuned and evaluated every model family; built SHAP explanations and a Gradio teacher-assist prototype.
- **Methods:** SVR, BiLSTM, and multilingual encoders; QLoRA-adapted **Qwen-KhmerGrader-4B**, **Gemma-KhmerGrader-4B**, and **SEA-LION-KhmerGrader-E2B**; report-listed GPT, Claude, Gemini, and DeepSeek baselines; QWK and exact/within-one agreement.
- **Outputs:** Report-stated test QWK 0.850 for Qwen-KhmerGrader-4B (n = 137; agreement with one teacher's marks; 67% exact, 80% within one point), above classical 0.802, BiLSTM 0.789, encoder 0.777 (n = 178), and the best zero-shot frontier model, Claude Opus 4.8 (0.734); public dataset, code, and report.
- **Limitations:** one grader; student-derived references; answer-level split; metrics predate corrected v1.0.1 and some saved values were not reproduced ([evaluation evidence](#)); no classroom study.

**COPAI knowledge assistant on MISTI infrastructure**

Mar 2025 - Oct 2026

*Lead AI Research and Innovation | Gen I and Gen II | Volunteer, part-time, hybrid*

Platform of [UN ESCAP and MISTI](#) | Supervision: Dr. Chen Sovann; H.E. Dr. Seingheng Hul; Mr. Rafael Torquato Cruz.

- **Problem:** make Cambodian AI policy documents queryable through a local assistant with cited sources.
- **My contribution:** led eight team members across Gen I and Gen II; led AI system development, built the assistant on MISTI infrastructure, and led data collection and knowledge-base preparation.
- **Methods:** document processing, BGE-M3 embeddings, Qdrant retrieval, BGE reranking, and vLLM inference on an NVIDIA A40 ([MISTI HPC](#)) through Open WebUI and Docker, with access controls and monitoring.
- **Outputs:** English-language RAG assistant with source citations and initial Khmer support.
- **Limitations:** Khmer answer quality not formally evaluated; no independently verified usage or retrieval metrics.

## RELEVANT EXPERIENCE

---

- AI Engineer**, CAMAI Software Solutions (project-based, hybrid) **Jan 2026 - Present**  
• Develop internal AI systems with multi-agent and human-in-the-loop workflows; work on ExamAI.
- AI Engineer**, Witamala, Brazil (part-time, remote) **Jul 2026 - Present**  
• Develop an internal AI-integrated operations system and learning studio.
- Teaching Assistant**, Royal University of Phnom Penh (part-time) **Apr 2025 - Jun 2026**  
• Supported SQL, Python, data science, and AI teaching through labs, assignments, and student projects.
- AI Tutor**, WEAVE Japan, ME-NEXT Program (part-time, remote) **Nov 2024 - Jan 2025**  
• Taught practical AI tools for research and digital work; guided students in building websites with AI assistance.

## TECHNICAL SKILLS

---

**ML, NLP, and evaluation:** Python, PyTorch, TensorFlow, scikit-learn, QLoRA, SHAP, pandas, NumPy, Gradio.

**Retrieval and local models:** RAG, vLLM, Open WebUI, Ollama, BGE embeddings, reranking, Qdrant.

**Data and systems:** SQL, PostgreSQL, FastAPI, Django, Docker, Git, Langfuse, Grafana.

## AWARDS AND RESEARCH COMMUNICATION

---

- **First Prize, 7th Engineering Day Awards 2024:** Facial Recognition for Secured Authentication in the Quantum Era, supervised by Dr. Khim Chamroeun.
- **Huawei ICT Competition:** selected for the Cambodia National Final 2024-2025, innovation track.
- **UN Women AI School, 22 July 2025:** session on small language models, RAG, and privacy-focused AI.
- **UN 2.0 Week, 9 June 2025:** presented Cambodian youth-led, community-focused open-source AI innovation.
- **COPAI Hub 9, 2025:** demonstrated the AI knowledge base supported by MISTI HPC infrastructure.

## AI SAFETY AND RESEARCH SERVICE

---

- **Singapore AI Safety Red Teaming Challenge 2026:** participant with Red Team Cambodia, probing AI risks and bias across languages.
- **Report review contribution:** acknowledged for comments and reviews in [Artificial Intelligence: Empowering Innovation, Ethics, and Inclusion in Cambodia](#), by MISTI and UN ESCAP (printed page vi).

## SELECTED ADDITIONAL STUDY

---

|                                                                       |                 |
|-----------------------------------------------------------------------|-----------------|
| <b>AWS Academy Machine Learning Foundations</b>                       | <b>Dec 2024</b> |
| <b>UNESCO UNITWIN Data Science Camp</b> , Handong Global University   | <b>Feb 2024</b> |
| <b>UNESCO UNITWIN Data Analytics Camp</b> , Handong Global University | <b>Feb 2023</b> |

## LANGUAGES

---

**Khmer:** native.

**English:** medium of instruction at RUPP; certificate (25 August 2026) available on request.

**Korean:** no qualification yet; starting study for TOPIK.

## REFERENCES

---

### Dr. Khim Chamroeun

Vice-Dean, Faculty of Engineering, RUPP

### H.E. Dr. Seingheng Hul

Under Secretary of State, MISTI

Contact details available on request.

### Dr. Chen Sovann

Director, STI Data Management Department, MISTI

### Mr. Rafael Torquato Cruz

Project Coordinator and Community Manager, UN ESCAP