

Muhammad Imam Luthfi Balaka

luthfibalaka@uchicago.edu | luthfibalaka.github.io | github.com/luthfibalaka | Chicago, IL 60637

Research Interests

Data Management, Data Discovery, Data Integration, and AI-assisted Data Systems.

Education

The University of Chicago, Chicago, IL 2025–Present
Ph.D. in Computer Science
Advisor: Prof. Raul Castro Fernandez

University of Indonesia, Depok, Indonesia 2020–2024
Bachelor of Computer Science, *cum laude*

Publications

Danni Liu, **Muhammad Imam Luthfi Balaka**, Raul Castro Fernandez. *StepLens: Supporting Inspection and Evaluation in Agentic Tabular Data Work to Narrow Down the Gulf of Evaluation*. Under review at UIST 2026.

Muhammad Imam Luthfi Balaka, John Hillesland, Kemal Badur, Raul Castro Fernandez. *Pneuma-Seeker: Relational Reification of Information Needs for Agentic Data Discovery and Preparation*. Under review at VLDB 2026.

Muhammad Imam Luthfi Balaka, Raul Castro Fernandez. *Demonstration of Pneuma-Seeker: Agentic System for Reifying and Fulfilling Information Needs on Tabular Data*. CAIS 2026.

Muhammad Imam Luthfi Balaka, Raul Castro Fernandez. *The Pneuma Project: Reifying Information Needs as Relational Schemas to Automate Discovery, Guide Preparation, and Align Data with Intent*. CIDR 2026.

Muhammad Imam Luthfi Balaka, David Alexander, Qiming Wang, Yue Gong, Adila Krisnadhi, Raul Castro Fernandez. *Pneuma: Leveraging Large Language Models for Tabular Data Representation and Retrieval in an End-to-End System*. SIGMOD 2025.

Research Experience

The University of Chicago, Chicago, IL Jun 2025–Present
Research Assistant, Advisor: Prof. Raul Castro Fernandez

- Architected PNEUMA-SEEKER, an end-to-end agentic system for tabular data discovery and preparation
- Deployed PNEUMA-SEEKER on Azure to support real-world procurement workflows at the University of Chicago
- Achieved up to 1.7× improvement in answer quality over baseline agentic frameworks such as SMOLAGENTS

The University of Chicago (Remote) Feb 2024–Mar 2025
Research Collaborator, Advisor: Prof. Raul Castro Fernandez

- Designed PNEUMA, an LLM-based data discovery system for tabular data
- Developed a binary-search-based algorithm for dynamically adjusting batch size for local LLMs

- Reduced index size by up to 99.8% while improving retrieval accuracy by up to 55% compared to baselines such as LLAMAINDEX and SOLO (SIGMOD '24)

Teaching Experience

University of Indonesia, Depok, Indonesia
Teaching Assistant

Jan 2023–Mar 2025

- Led tutorials, designed laboratory assignments, and mentored 150+ students across courses in Natural Language Processing, Information Retrieval, Computer Vision, Statistics, and Data Science.

Industry Experience

Polyrific

Feb 2025 – Apr 2025

AI Software Engineer Intern

- Contributed to GENIVERSE.AI, an AI assistant for data interaction, focusing on backend development
- Improved question-answering accuracy by up to 20% using multimodal models (text, audio, image)
- Built tabular QA pipelines integrating APIs, databases, and structured data sources (CSV, Excel)
- Reduced memory usage of PII masking system from 6GB to 200MB (30× reduction)

Honors and Awards

Eckhardt Graduate Scholarship, The University of Chicago

2025

Scholarship awarded to outstanding doctoral students in the Physical Sciences Division.

Garuda ACE Fellowship, Ministry of Education, Indonesia

2024

National fellowship supporting top Indonesian undergraduates to collaborate with U.S. faculty.

Technical Skills

Languages: Python, C/C++, Java

Machine Learning: PyTorch, TensorFlow, smolagents

Databases: PostgreSQL, DuckDB

Cloud Platforms: Google Cloud Platform (GCP), Microsoft Azure