BAYU ARYOYUDANTA (YUDAN)

baa181@pitt.edu — linkedin.com/in/yudanta — Google Scholar

EDUCATION

University of Pittsburgh, Pittsburgh, PA 2024 – Present
PhD in Rehabilitation Science
Advisor: Bambang Parmanto, PhD. Research focus: AI-enabled mHealth and AI-mediated behavioral health interventions for precision rehabilitation.

Gadjah Mada University, Yogyakarta, Indonesia Graduated: September 2016
Master of Engineering in Information Technology
Master Thesis: Semi-supervised Learning Approach for Indonesian Named Entity Recognition using Co-Training Algorithm.

Yogyakarta State University, Yogyakarta, Indonesia Graduated: July 2012
B.S in Information Technology for Education
Thesis: Route and Trip Planning for Yogyakarta's Rapid Transportation Bus Riders with Djikstra Algorithm

RESEARCH INTERESTS

- **Clinical & Biomedical Natural Language Processing**
- **Large Language Models & Generative AI**
- **AI-Enabled mHealth & Behavioral Health Interventions**
- **Health Informatics**
- **Rehabilitation Informatics**
- **Computational Social Science**

RESEARCH EXPERIENCE

School of Health and Rehabilitation Science, University of Pittsburgh Pittsburgh, PA
Graduate Student Researcher August 2024 – Present

- PhD student in Rehabilitation Science supervised by Prof. Bambang Parmanto, PhD. Research area includes application of Generative AI, AI Agents, and Mobile Health as ecological momentary intervention assisting individuals with disabilities and chronic conditions.

Health and Rehabilitation Informatics (HARI) Lab, University of Pittsburgh Pittsburgh, PA
Research Scientist June 2023 – August 2024

- Conducted research on the application of Large Language Models and Generative AI in rehabilitation and caregiving. Co-developed the Caregiving Language Model (CaLM), a retrieval-augmented, fine-tuned language model to support family caregivers (published in *JMIR Formative Research*, 2024).

IISE Diplomacy Lab, University of Pittsburgh Pittsburgh, PA
Technical Support, Volunteer July 2022 – Present

- Developed automatic keyword extraction using statistical and unsupervised algorithmic approaches for extracting frequent and important keywords from a domain-specific corpus, in collaboration with the U.S. State Department, to support foreign and consular counselor training materials. Reduced both processing time and cost by up to 80%.

PROFESSIONAL EXPERIENCE

Stemly Singapore
Data Scientist August 2021 – July 2022

- Designed and implemented an active learning classification system for large-scale transaction logs as a company SAAS module, achieving up to 95% accuracy and reducing processing time by more than 90% compared to the previous approach.

Citra Multimedia Indonesia Jakarta, Indonesia
Lead Machine Learning Engineer June 2019 – July 2020

- Designed and orchestrated an Indonesian NLP model built with deep learning architecture. The model and data pipeline scaled from a few hundred to several thousand texts processed daily with limited human intervention, reducing cost and increasing throughput up to fourfold.

Netalase Yogyakarta, Indonesia
Machine Learning Engineer August 2018 – June 2019

- Developed an e-commerce item classification system and content-based recommendation system for a hybrid online media commerce platform, reducing item placement error by 35–40% with minimal human intervention.

Dattabot Jakarta, Indonesia
Software Engineer February 2018 – August 2018

- Developed middleware connecting wearable devices to an in-house fatigue detection system, capturing and transmitting human vital information.

Nuwira Integrasi Mandiri Yogyakarta, Indonesia
Application Developer July 2012 – July 2015

- Developed a job-order management and CRM application, converting a paper-based workflow into a fully paperless system with near real-time monitoring, improving job distribution and response SLA by up to 90%.

PROJECTS

Night Shift Nerds – MEDIQA-SYNUR 2026 Shared Task Pittsburgh, PA
github.com/yudanta/nsn-mediqa-synur 2026

- Fine-tuned a sub-2B-parameter LLM (Qwen3-1.7B) with SFT + GRPO-based RLVR for clinical observation extraction and normalization from nurse dictation against a 193-concept ontology. Achieved 0.740 F1 on the held-out shared-task test set. Published as a first-author workshop paper (see Publications).

Caregiving Language Model (CaLM), HARI Lab Pittsburgh, PA
July 2023

- Research on developing a domain-specific knowledge AI assistant using a retrieval-augmented generation (RAG) approach for the caregiving domain.

Automatic Keywords Extraction, IISE Diplomacy Lab & State Department Pittsburgh, PA
July 2022 – Present

- Developed automatic keyword extraction for State Department-related corpora, producing important and highly technical keywords for foreign and consular learning materials.

Automatic Named Entity Recognition, Binocular Yogyakarta, Indonesia
November 2016

- Developed NER, sentiment classification, and topic modeling as part of a data enrichment program for a media analytics company.

Indonesian Text Pre-processing Utilities
github.com/yudanta/id-sentence-segmenter July 2016

- Text pre-processing utilities for the Indonesian language, including sentence segmentation, tokenization, stemming, and n-gram features.

SecSurf, DTETI UGM Yogyakarta, Indonesia
March 2016

- Provided secure, content-safe internet access by maintaining a database of unsafe web pages and a pipeline for multi-modal restricted web-page classification.

PUBLICATIONS

[1] Bayu Aryoyudanta, Maria Yuliana, Mikie Rachman, and I Made Agus Setiawan. Night Shift Nerds at MEDIQA-SYNUR 2026: Pushing Small Large Language Model Capability for Clinical Observation Extraction and Normalization from Nurse Dictation using RLVR. In *Proceedings of the 8th Workshop on Clinical Natural Language Processing (Clinical NLP) @ LREC 2026*, pages 163–173. ELRA Language Resources Association (ELRA), May 2026.

[2] Bambang Parmanto, Bayu Aryoyudanta, Timothius Wilbert Soekinto, I Made Agus Setiawan, Yuhan Wang, Haomin Hu, Andi Saptono, and Yong Kyung Choi. A Reliable and Accessible Caregiving Language Model (CaLM) to Support Tools for Caregivers: Development and Evaluation Study. *JMIR Formative Research*, 8:e54633, July 2024.

[3] Bayu Aryoyudanta, Teguh Bharata Adji, and Indriana Hidayah. Semi-supervised learning approach for indonesian named entity recognition (ner) using co-training algorithm. *2016 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, pages 7–12, 2016.

MANUSCRIPTS UNDER REVIEW

- Setiawan, I.M.A., Yuliana, M., **Aryoyudanta, B.**, Indradhirmaya, F.A., Patel, D., Beck, D.C., Saptono, A., Hu, H., & Parmanto, B. *Expert validation of AI-generated clinical assessment items for physician assistant certification: a multi-dimensional rubric and dual-reliability method.* Under review, *npj Digital Medicine*. Preprint: `https://doi.org/10.21203/rs.3.rs-10475937/v1`

TEACHING & TAs

Department of Health Information Management, School of Health and Rehabilitation Sciences, University of Pittsburgh — Pittsburgh, PA
Teaching Assistant — Fall 2025 – Present

- HI 2453: Machine Learning in Health Science — Fall 2025
- HI 2021: Practical Statistics & Programming Using R — Spring 2026
- HI 1454: Data Science in Health Informatics — Spring 2026
- HI 2460: Data Science & Machine Learning — Spring 2026, Fall 2026
- HI 2022: Introduction to Python for Health Informatics — Fall 2026

Communication Science, FISHIPOL, YSU — Yogyakarta, Indonesia
Part-time University Lecturer — March 2017 – Jan 2019

- Taught Communication Technology and Media Analytics classes.

Electrical Engineering Department, Engineering Faculty, YSU — Yogyakarta, Indonesia
Teaching Assistant — Undergrad

- TA for Programming, Data Structures, Digital Media, and Computer Networking classes.

PRESENTATIONS

- **A Self-reflection Rewriting Method for Readability of Answers to Questions about Spinal Cord Injury**, AMIA Annual Symposium, 2025
- **Productionizing Machine Learning**, AllData Indonesia, 2023
- **Person's Name Gender Prediction using Recurrent Neural Network**, Python Indonesia Chapter Yogyakarta, 2016

SKILLS

- **Relevant Coursework:** Advanced Machine Learning, Computer Modelling and Simulation, Cloud Computing, Programming.
- **Programming:** Python, Bash, JS.
- **Machine Learning Algorithms**: Bayesian, SVM, KNN, K-Means, Markov Model
- **Deep Learning Frameworks**: PyTorch, TensorFlow, HuggingFace TRL
- **Generative AI**: Autoregressive Models, RAG, RLVR/GRPO Fine-Tuning, Vision Language Models (VLM)