

# Derry Tanti Wijaya

Associate Professor of Data Science · Computational Linguistics & Natural Language Processing

Email: [Derry.Wijaya@monash.edu](mailto:Derry.Wijaya@monash.edu) · [Google Scholar](#) · [Website](#)

<b>4,388+</b> Citations (Google Scholar)	<b>\$2.3M+</b> Research Funding Raised (PI/Co-PI)	<b>8</b> PhD Students Supervised	<b>90+</b> Peer-Reviewed Publications
---	---	-------------------------------------	--

Publication venues include ICML, ICLR, and \*CL conferences. h-index: 28 (Google Scholar, 2026).

## PROFILE

I work on the **interpretability, evaluation, and controllability of language models**, and I apply these methods to the settings where models are least understood and most consequential: **low-resource, multilingual, and culturally grounded languages**. My thesis is that under-served languages are not just an equity concern but a scientific frontier—the place where the gaps in how models represent meaning, follow instructions, and can be trusted become visible and testable. My research spans mechanistic interpretability and model editing, evaluation frameworks and LLM-as-a-judge, and data-efficient adaptation for low-resource languages, alongside culturally grounded resources, and benchmarks (indigenous scripts, honorifics, code-switching) and the use of NLP to study bias, polarization, and misinformation in the world's emerging democracies. Ph.D. from Carnegie Mellon University (advisor: Tom Mitchell); previously Assistant Professor of Computer Science at Boston University and postdoctoral researcher at the University of Pennsylvania (advisor: Chris Callison-Burch). Author of work at **ICML, ICLR, ACL, NAACL, and EMNLP**; principal or co-investigator on grants from NSF, DARPA, Google, the Open Society Foundation, and Danida.

## RESEARCH INTERESTS

- **Interpretable & steerable LMs:** mechanistic interpretability of multilingual/multimodal models, safe model editing, controllability, hallucination reduction.
- **Evaluation of generative models:** meta-metrics, LLM-as-a-judge, rubric-agnostic reward reasoning, alignment with human preferences.
- **Low-resource & multilingual NLP:** data-efficient transfer (circuit-targeted fine-tuning), machine translation from dictionaries and grammars, code-switching generation.
- **Multicultural AI & cultural heritage:** culturally aware large multimodal models, preservation of indigenous scripts, honorific systems.
- **NLP for society:** bias, polarization, toxicity, hate speech, and misinformation; computational framing analysis; AI for health.

## ACADEMIC APPOINTMENTS

**Associate Professor, Data Science — Monash University, Indonesia** 2022 – present  
Program Coordinator, Data Science (2024–present). Co-Director, Monash Data & Democracy Research Hub (2023–present).

**Assistant Professor, Computer Science — Boston University** 2018 – 2022  
Now Adjunct. Co-Director, AI and Education Initiative, Hariri Institute (2021–2022).

**Postdoctoral Researcher, Computer & Information Science — University of Pennsylvania** 2016 – 2018  
Advisor: Chris Callison-Burch.

## EDUCATION

**Ph.D., Language Technologies Institute — Carnegie Mellon University** 2016  
Advisor: Tom Mitchell. Thesis: VerbKB — A Knowledge Base of Verbs for Natural Language Understanding.

**M.Sc., School of Computing — National University of Singapore** 2010

## AWARDS & FELLOWSHIPS

---

- **Best Data Researcher, Data Science Indonesia Award** (2021)
- **University Provost's Career Development Professorship, Boston University** (2019)
- **Anita Borg Scholarship, Google** (2013)
- **International Science & Technology Fellowship, Fulbright** (2010–2016)

## SELECTED PUBLICATIONS

---

*Name in bold. See Google Scholar for the complete list of 60+ works (4,388+ citations, h-index 28).*

### Interpretability, Model Editing & Evaluation

- Do Language Models Track Entities Across State Changes?  
Z. Tang, Q. Zhao, G. Franco, **D. Wijaya**, A. Mueller, S. Schuster, N. Kim  
*ICML 2026* [\[link\]](#)
- Beyond Transfer Accuracy: Faithful Circuits for Controlled Low-Resource Adaptation  
K. Nur'aini, A. Purwarianti, A. F. Aji, **D. Wijaya**  
*Preprint (in submission)* [\[link\]](#)
- mR3: Multilingual Rubric-Agnostic Reward Reasoning Models  
D. Anugraha, S-Y. Hung, Z. Tang, A. E-S. Lee, **D. Wijaya**, G. I. Winata  
*ICLR 2026* [\[link\]](#)
- MetaMetrics: Calibrating Metrics for Generation Tasks Using Human Preferences  
G. I. Winata, D. Anugraha, L. Susanto, G. Kuwanto, **D. Wijaya**  
*ICLR 2025* [\[link\]](#)
- Deductive Closure Training of Language Models for Coherence, Accuracy, and Updatability  
A. F. Akyürek, E. Akyürek, L. Choshen, **D. Wijaya**, J. Andreas  
*ACL Findings 2024* [\[link\]](#)
- DUnE: Dataset for Unified Editing  
A. F. Akyürek, E. L. Pan, G. Kuwanto, **D. Wijaya**  
*EMNLP 2023* [\[link\]](#)
- RL4F: Generating Natural Language Feedback with Reinforcement Learning for Repairing Model Outputs  
A. F. Akyürek, E. Akyürek, A. Kalyan, P. Clark, **D. Wijaya**, N. Tandon  
*ACL 2023* [\[link\]](#)

### Multilingual, Multimodal & Multicultural NLP

- NusaAksara: A Multimodal and Multilingual Benchmark for Preserving Indonesian Indigenous Scripts  
M. F. Adilazuarda, M. Wijanarko, L. Susanto, K. Nur'aini, **D. Wijaya**, A. F. Aji  
*ACL 2025* [\[link\]](#)
- Do Language Models Understand Honorific Systems in Javanese?  
M. R. Farhansyah, I. Darmawan, A. Kusumawardhana, G. I. Winata, A. F. Aji, **D. Wijaya**  
*ACL 2025* [\[link\]](#)
- WorldCuisines: A Massive-Scale Benchmark for Multilingual and Multicultural Visual Question Answering on Global Cuisines  
G. I. Winata et al., **D. Wijaya**  
*NAACL 2025* [\[link\]](#)
- What Do Indonesians Really Need from Language Technology? A Nationwide Survey  
M. D. A. Kautsar, L. Susanto, **D. Wijaya**, F. Koto  
*EMNLP 2025* [\[link\]](#)
- Linguistics Theory Meets LLM: Code-Switched Text Generation via Equivalence Constrained Large Language Models  
G. Kuwanto, C. Agarwal, G. I. Winata, **D. Wijaya**  
*Workshop on Computational Developmental Linguistics 2026* [\[link\]](#)
- Mitigating Translationese in Low-resource Languages: The Storyboard Approach

G. Kuwanto et al., **D. Wijaya**

*LREC-COLING 2024* [\[link\]](#)

- Explain-then-Translate: An Analysis on Improving Program Translation with Self-Generated Explanations  
Z. Tang, M. Agarwal, A. G. Shypula, B. Wang, **D. Wijaya**, J. Chen, Y. Kim  
*EMNLP Findings 2023* [\[link\]](#)

### NLP for Society: Bias, Polarization, Misinformation & Health

- A Multi-Labeled Dataset for Indonesian Discourse: Examining Toxicity, Polarization, and Demographics  
L. Susanto, M. Wijanarko, P. Pratama, Z. Tang, F. Akyas, T. Hong, I. Idris, A. F. Aji, **D. Wijaya**  
*ACL Findings 2025* [\[link\]](#)
- COVID-19 Vaccine Misinformation in Middle Income Countries  
J. Kim, B. R. Bak, A. Agrawal, J. Wu, V. J. Wirtz, T. Hong, **D. Wijaya**  
*EMNLP 2023* [\[link\]](#)
- Effects of #coronavirus Content Moderation on Misinformation and Anti-Asian Hate on Instagram  
T. Hong, Z. Tang, M. Lu, Y. Wang, J. Wu, **D. Wijaya**  
*New Media & Society 2023* [\[link\]](#)
- Agenda Setting, Cross-Cutting Effects, and Political Expression on Social Media: The Gun Violence Case  
L. Guo, Y. Zhang, K. Mays, A. F. Akyürek, **D. Wijaya**, M. Betke  
*Communication Research 2023* [\[link\]](#)
- Proposing an Open-Sourced Tool for Computational Framing Analysis of Multilingual Data  
L. Guo, C. Su, S. Paik, V. Bhatia, V. P. Akavoor, G. Gao, M. Betke, **D. Wijaya**  
*Digital Journalism 2023* [\[link\]](#)

## COMPETITIVE RESEARCH GRANTS

---

Over \$2.3M USD raised as Principal Investigator or Co-Investigator. Selected:

- **IndigiSEA: Indigenous Digital Futures for Southeast Asia** — Monash Incubator Grant, \$53k, 2026
- **Digitalising Waste Management for a Just Green Transition in Indonesian Smart Cities** — Danida (Denmark), \$300k, 2025-2029
- **Sea-Tech: Sustainable Seaweed Cultivation through Smart Technology** — PAIR (Australia-Indonesia), \$61k, 2025-2027
- **Contextualizing LLMs for Women's Sexual & Reproductive Health in Indonesia** — Google Research Award, \$60k (PI), 2024
- **ASEAN Disinformation Index** — Monash Data Futures Institute, \$33k, 2024
- **Boosting Monash's Global Reputation: Data & Democracy Hub** — Monash Incubator Grant, \$53k, 2024
- **Monitoring Hate Speech toward Vulnerable Communities, 2024 Indonesia Election** — Open Society Foundation, \$37k, 2023
- **Exploring Generative AI for Moderating Online Polarisation** — Monash Data Futures Institute, \$35k (PI), 2023
- **Bhinneka: Multilingual Indonesian Resources** — Indonesian Ministry of Education (US Collaboration), \$45k, 2022-2023
- **Exploring the Evolution of Racial Biases through Framing Analysis** — Google Research Scholar Award, \$60k (PI), 2021
- **LEARN: Label-Efficient Active Resilient Network** — DARPA (Learning with Less Labels), \$462k (Co-PI), 2019-2022
- **Bridging Linguistic and Visual Knowledge through Visual Genome** — Hariri Institute Research Incubation Award, \$28k (PI), 2019-2021
- **AIDA: Semi-supervised Learning of Multimodal Representations** — DARPA, \$100k (with C. Callison-Burch), 2019-2020
- **ExploreCSR: Supporting and Promoting Undergraduate Research for Women** — Google, \$18k, 2019
- **BIGDATA: Multiplatform, Multilingual, and Multimodal Tools for Analyzing Public Communication in 100+ Languages** — NSF, \$1,000,000 (with M. Betke, L. Guo, P. Ishwar), 2018-2022

## RESEARCH SUPERVISION

---

### Doctoral Students (8 total: 4 ongoing, 4 completed, all with top-venue output)

- **Ongoing:** Khumaisa Nur'aini (Monash), Adryan Kusumawardhana (Monash), Zilu Tang (BU), Ge Gao (BU).

- **Tahsina Hashem** (Monash, 2026) — Faithful and Fair Text Generation from Structured and Visual Data.
- **Muhammed Yusuf Kocyigit** (BU, 2025) — Improving Evaluation Practices in NLP: Contamination and Metric Limitations.
- **Afra Feyza Akyürek** (BU, 2025) — Advancing Neural Networks Beyond Initial Training.
- **Isidora Tourni** (BU, 2023) — Neural Machine Translation for Low-Resource Conditions.

### Master's Students (selected; many with first-author papers at ACL/ICLR/Q1 Journals)

- **Garry Kuwanto** (Master's BU, 2026) — MetaMetrics: a Bayesian optimization framework for calibrating LLM evaluation metrics (ICLR 2025).
- **Hairurahman** (Master's Monash, 2025) — Hybrid knowledge graph for Indonesian sentencing analysis (AI and Law 2026).
- **Iwan Darmawan** (Master's Monash, 2024) — Javanese Honorific System corpus (ACL 2025).
- **Fariz Akyas** (Master's, Monash, 2024) — Annotator Identity in Indonesian Political Polarization Corpus (ACL Findings 2025).

*Full supervision record (15+ Master's and Undergraduate students) available on request.*

## TEACHING

---

Unit Coordinator, Master of Data Science, Monash University Indonesia. Selected courses with most recent student evaluation (SETU) scores:

- **Machine Learning (IT15201)** — 82/100 (2026). Grounds ML in real-world societal challenges.
- **Data Analysis for Semi-Structured Data (IT15212)** — 86/100 (2026). Focus on generative AI and LLMs.
- **Master Thesis supervision (IT15126–5229)** — 88/100 (2025). Students published and presented at top conferences.
- **Data Wrangling (IT15196)** — 86/100 (2024)
- **Introduction to Data Science (IT15145)** — 82/100 (2022)

## LEADERSHIP & PROFESSIONAL SERVICE

---

### Academic Leadership

- **Program Coordinator, Data Science**, Monash University Indonesia (2024–present).
- **Co-Director, Monash Data & Democracy Research Hub** (2023–present) — partnerships across government, academia, and civil society; convened the Data & Democracy Dialogue (Jakarta 2024; Kuala Lumpur 2025).
- **Faculty of IT Course Governance, Quality & Resourcing Sub-committee** (2026–present); Education Committee (2024–present); Research Committee (2024–2025).
- **Co-Director, AI & Education Initiative**, Hariri Institute, Boston University (2021–2022).

### Community Service to the Field

- **Area Chair / Senior Area Chair / Session Chair** for ACL conferences, ICLR, and NeurIPS (ongoing).
- **Organizer:** Demo Chair AACL 2026 (Hengqin); 2nd Workshop on Southeast Asian Language Processing (COLING 2025); 7th Workshop on Computational Approaches to Linguistic Code-Switching (NAACL 2025); Local Chair AACL 2023 (Bali).

---

*References available upon request.*