Kemal Maulana Kurniawan

4.129, Melbourne Connect, 700 Swanston St, Carlton VIC 3053, Australia

kemal@kkurniawan.com | kurniawan.k@unimelb.edu.au

Education

PhD, Engineering

University of Melbourne, Australia, 2023

- Natural Language Processing group, School of Computing and Information Systems.
- Thesis: *Source-Free Transductive Transfer Learning for Structured Prediction*. Supervised by Trevor Cohn and Lea Frermann.

MSc, Artificial Intelligence

University of Edinburgh, Scotland, United Kingdom, 2017

- Thesis: *Exploring Recurrent Neural Network Grammars for Parsing Low-Resource Languages*. Supervised by Frank Keller.
- Graduated with Distinction.

S.Kom., Computer Science

Universitas Indonesia, Depok, Indonesia, 2015

- Thesis: *Off-Topic Post Detection in Discussion Forums with Anomalous Text Detection Method*. Supervised by Hisar Maruli Manurung.
- Graduated *cum laude* with CGPA of 3.87/4.00.
- Participated in ACM-ICPC Asia-Jakarta 2010 and 2012, Asia-Kuala Lumpur 2011, and Asia-Hanoi 2012.

Young Scientist Exchange Program

Tokyo Institute of Technology, Tokyo, Japan, 2014

- Final report: *Predicting Within- and Inter-community Links in Networks*. Supervised by Tsuyoshi Murata.
- Presented the work in the 102nd conference of The Japanese Society for Artificial Intelligence.
- Wrote link prediction module for NetworkX, a Python library for manipulating and analyzing networks and graphs.

Publications

Kurniawan, K., Mistica, M., Baldwin, T., & Lau, J. H. (2025). Training and Evaluating with Human Label Variation: An Empirical Study. http://arxiv.org/abs/2502.01891 (*under review*)

Simonds, T., Kurniawan, K., & Lau, J. H. (2024). MoDEM: Mixture of Domain Expert Models. *Proceedings of the 22nd Annual Workshop of the Australasian Language Technology Association.* https://aclanthology.org/2024.alta-1.6

Merx, R., Vylomova, E., & Kurniawan, K. (2024). Generating bilingual example sentences with large language models as lexicography assistants. *Proceedings of the 22nd Annual Workshop of the Australasian Language Technology Association.* https://aclanthology.org/2024.alta-1.5/ **(Best Paper)**

Kurniawan, K., Mistica, M., Baldwin, T., & Lau, J. H. (2024). To Aggregate or Not to Aggregate. That is the Question: A Case Study on Annotation Subjectivity in Span Prediction. *Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis*, 362-368. https://aclanthology.org/2024.wassa-1.29/

Winata, G. I., Aji, A. F., Cahyawijaya, S., Mahendra, R., Koto, F., Romadhony, A., Kurniawan, K., Moeljadi, D., Prasojo, R. E., Fung, P., Baldwin, T., Lau, J. H., Sennrich, R., & Ruder, S. (2023). NusaX: Multilingual Parallel Sentiment Dataset for 10 Indonesian Local Languages. *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, 815–834. https://aclanthology.org/2023.eacl-main.57/ **(Outstanding Paper)**

Kurniawan, K., Frermann, L., Schulz, P., & Cohn, T. (2022). Unsupervised Cross-Lingual Transfer of Structured Predictors without Source Data. *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2041–2054. https://aclanthology.org/2022.naacl-main.149/

Aji, A., Winata, G. I., Koto, F., Cahyawijaya, S., Romadhony, A., Mahendra, R., Kurniawan, K., Moeljadi, D., Prasojo, R. E., Baldwin, T., Lau, J. H., & Ruder, S. (2022). One Country, 700+ Languages: NLP Challenges for Underrepresented Languages and Dialects in Indonesia. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 7226–7249. https://aclanthology.org/2022.acl-long.500

Kurniawan, K., Frermann, L., Schulz, P., & Cohn, T. (2021). PTST-UoM at SemEval-2021 Task 10: Parsimonious Transfer for Sequence Tagging. *Proceedings of the 15th International Workshop on Semantic Evaluation (SemEval-2021)*, 445–451. https://aclanthology.org/2021.semeval-1.54

Kurniawan, K., Frermann, L., Schulz, P., & Cohn, T. (2021). PPT: Parsimonious Parser Transfer for Unsupervised Cross-Lingual Adaptation. *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL): Main Volume*, 2907–2918. https://www.aclweb.org/anthology/2021.eacl-main.254

Kurniawan, K. (2019). KaWAT: A Word Analogy Task Dataset for Indonesian. *ArXiv:1906.09912 [Cs]*. http://arxiv.org/abs/1906.09912

Kurniawan, K., & Louvan, S. (2018). IndoSum: A New Benchmark Dataset for Indonesian Text Summarization. *2018 International Conference on Asian Language Processing (IALP)*, 215–220. https://doi.org/10.1109/IALP.2018.8629109

Kurniawan, K., & Aji, A. F. (2018). Toward a Standardized and More Accurate Indonesian Part-of-Speech Tagging. *2018 International Conference on Asian Language Processing (IALP)*, 303–307. https://doi.org/10.1109/IALP.2018.8629236

Kurniawan, K., & Louvan, S. (2018). Empirical Evaluation of Character-Based Model on Neural Named-Entity Recognition in Indonesian Conversational Texts. *Proceedings of the 2018 EMNLP Workshop W-NUT: The 4th Workshop on Noisy User-Generated Text*, 85–92. http://aclweb.org/anthology/W18-6112

Ikhwantri, F., Louvan, S., Kurniawan, K., Abisena, B., Rachman, V., Wicaksono, A. F., & Mahendra, R. (2018). Multi-Task Active Learning for Neural Semantic Role Labeling on Low Resource Conversational Corpus. *Proceedings of the Workshop on Deep Learning Approaches for Low-Resource NLP*, 43–50. http://aclweb.org/anthology/W18-3406

Employment

Research Fellow   Aug 2023 - present
School of Computing and Information Systems, University of Melbourne, Australia
- Working on training and evaluation methods of NLP models that embrace human label variation.
- Improved the accuracy of Justice Connect's legal area classifier for older people by 8%, contributing to JC's award of Not-for-Profit Technology Innovator of the Year in the Infoxchange

Australian Not-for-Profit Technology Awards.[1]

- Developed a neural model performing over 2 times better than chance in predicting text spans that support the assignment of a legal area to a problem description written by lay help-seekers.
- Improved the accuracy of JC's legal area classifier for Criminal Law by 5%.

Applied Scientist Intern   Mar - Sep 2022

Amazon, Melbourne, Australia

- Researched multiturn dialogue-based image retrieval.
- Successfully demonstrated that a published dataset designed for the task is solvable with only the last turn.
- Led a data annotation process for the task using large-scale Amazon customer behaviour data on AWS Sagemaker Ground Truth.
- Collaborated with annotators to iteratively improve annotation guidelines.
- Manager and mentor: Philip Schulz and Julien Monteil.

Research Scientist   Oct 2017 - May 2019

PT YesBoss Group Indonesia (Kata.ai), Jakarta, Indonesia

- Experimented on out-of-vocabulary (OOV) handling for Indonesian entity recognition on conversational texts, achieved a model having roughly 80% micro-averaged F_1 score even when OOV rate is more than 90%, and published the work at EMNLP Workshop on Noisy User-generated Text (W-NUT) 2018.
- Implemented the entity recognition model as a service for production use.
- Experimented on extractive summarization for Indonesian news. Achieved a model having 65% of ROUGE-1, 60% of ROUGE-2, and 64% of ROUGE-L scores.
- Implemented the extractive summariser as a Python library for production use.

Software Engineer   Apr 2015 - Jan 2016

PT Dwi Cermat Indonesia (Cermati), Jakarta, Indonesia

- Implemented auto-generation of minified CSS and JavaScript, auto-revert of broken build, and auto-tagging of successful build in production branch using Gulp and Travis CI.
- Implemented Slack webhook integration for error logs, Ansible deployment logs, and Fail2Ban logs.
- Wrote test fixtures creation module with Faker.js, reducing boilerplate code of setting up testing data.
- Implemented captcha and IP blocker with Fail2Ban to protect servers from malicious attacks.

Software Engineer   Jan - Apr 2015

PT Rekan Usaha Mikro Anda (RUMA), Jakarta, Indonesia

- Added SMS notification services to existing RUMA agents web management system.
- Developed balance deduction penalty system for RUMA agents that are dormant for a certain period of time.
- Fixed a critical bug in existing implementation regarding transaction report and agent transfer.
- Developed a simple middleware implementing RESTful API to be used for RUMA's Android application.

Junior Research Assistant, Intern   2013

National University of Singapore, Singapore

- Assisted a research in statistical machine translation under supervision of Ng Hwee Tou from NUS School of Computing.
- Implemented English-to-Chinese translation system prototype using Moses package based on the existing Chinese-to-English translation system.

---

[1] https://justiceconnect.org.au/wp-content/uploads/2024/12/JC24_Justice_Connect_Annual_Impact_Report_23-24.pdf, page 20

Teaching

Tutor   Feb - Jun 2025

Melbourne Business School, University of Melbourne, Australia
- Research DL frameworks and LLM APIs for the Natural Language Processing subject tutorials.
- Develop tutorial materials on RAG and LLM prompting using Google Gen AI SDK and ChromaDB vector database on Google Colab.

Supermarker "Computer Science Research Project Part 2"   Oct 2024

School of Computing and Information Systems, University of Melbourne, Australia
- Marked 36 master's research proposal presentations spanning all areas of computer science.

Head Tutor "Text Analytics for Business"   May 2024 - Jul 2024

Melbourne Business School, University of Melbourne, Australia
- Streamlined the computing environment setup by moving the tutorials to Google Colab.
- Scheduled the release of tutorial problems and solutions.
- Completed tutorial materials on syntactic parsing and lexical semantics.
- Simplified the BERT tutorial by using HuggingFace Transformers library's built-in trainer.
- Created solutions for the practice final exam.
- Delivered 8 on-site tutorial sessions in tandem with another tutor.
- Marked final assignment and exam.

Guest Lecture "Statistical Machine Learning"   Apr 2024

School of Computing and Information Systems, University of Melbourne, Australia
- Delievered a lecture on Transformers.

Tutor "Introduction to Machine Learning"   Jul - Dec 2023

School of Computing and Information Systems, University of Melbourne, Australia
- Delivered on-site tutorials.
- Marked students' assignment and exam submissions.

Tutor "Introduction to Machine Learning"   Feb - Jul 2023

School of Computing and Information Systems, University of Melbourne, Australia
- Delivered tutorials, both online and on-site.
- Assisted the subject coordinator in assembling assignments.
- Marked students' assignment and exam submissions.

Tutor "Advanced Machine Learning"   Aug 2021 - Feb 2022

Pacmann.ai, Jakarta, Indonesia
- Created and delivered materials to industry practitioners.

Project Supervisor "Data Science Project"   Mar - Dec 2021

School of Computing and Information Systems, University of Melbourne, Australia
- Supervised 2 groups working on NLP-related data science capstone projects.
- Attended supervision meetings and provided suggestions on methodology.

Tutor "Introduction to Machine Learning"   Feb - Jul 2020

School of Computing and Information Systems, University of Melbourne, Australia
- Delivered workshops, mostly online due to the pandemic.
- Marked students' assignment and exam submissions.

Teaching Assistant   Feb - Jun 2016

Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia
- Led the teaching assistant team for the Functional Programming class and prepared materials for the bi-weekly tutorials.

- Prepared project assignment materials and assisted students in weekly tutorials for the Advanced Programming class.
- Delivered tutorials for the Statistics and Probability class.

Teaching Assistant   2011 - 2015
Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia
- Was responsible for 15 - 20 students in Discrete Math, Linear Algebra, Statistics and Probability, and Data Structures and Algorithms classes.
- Delivered tutorials, assisted students in doing their assignments, and graded their works.

Supervision

Masters

Annisa Yusup, co-supervision with Jey Han Lau   Feb 2025 - present
Semi-supervised classification of Indonesian social media texts

Nahid Tajik, Akash Pandey, Avinash Rao (capstone)   2021
Adding NLP support to a cybersecurity risk platform

Liang Min, Minghui Sun, Jiacheng Yu, Zixuan Zeng, Xinle Chen (capstone)   2021
Automated tag suggestion for social media records

Undergraduates

Hannah Bansal, co-supervision with Lea Frermann   Mar 2025 - present
Neurosymbolic approaches for counterfactual reasoning

Toby Simonds, co-supervision with Jey Han Lau   Jul - Oct 2024
Investigating mixture of domain expert models to reduce inference cost

Talks

Empirical Meta-Evaluation of Metrics for Evaluation against Human Label Variation   2024
NLP Reading Group, University of Melbourne, Australia

Easier Transfer Learning for Predicting Sentence Structures   2022
3 Minute Thesis, CIS Doctoral Colloquium, Melbourne, Australia

Implementasi Teorema Matriks-Pohon dengan PyTorch   2021
PyCon ID, Python Indonesia, Online

PPT: Parsimonious Parser Transfer   2021
Bedah Paper, Indonesia Association for Computational Linguistics (INACL), Online

NLP for Chatbots   2018
Workshop on Machine Learning and Its Application, Telkom University, Bandung, Indonesia

Kata.ai's Best Practice for Machine Learning   2018
Tech in Asia - Product Development Conference, Menara Mandiri, Jakarta, Indonesia

Introduction to AI and NLP   2017
Facebook Developer Circle Surabaya: Facebook Goes to Campus, Surabaya, Indonesia

Predicting Within- and Inter-community Links in Networks 2014
102nd conference of The Japanese Society for Artificial Intelligence, Osaka, Japan

Projects

text2array, author <https://github.com/kmkurn/text2array>
Easy conversion of textual data to integer arrays. Intended as a replacement for torchtext and similar libraries.

rnnr, author <https://github.com/kmkurn/rnnr>
A library to help write neural network training and testing loops. Has built-in callbacks for dealing with early stopping, checkpointing, etc.

pytorch-crf, author <https://github.com/kmkurn/pytorch-crf>
Linear-chain conditional random field in PyTorch.

torchtext, past contributor <https://github.com/pytorch/text>
Data loaders and abstractions for NLP using PyTorch. Contributed to this project by implementing a field class to contain characters of words. This field is essential to use character embedding with `torchtext` and PyTorch.

id-nlp-resource, author <https://github.com/kmkurn/id-nlp-resource>
A list of Indonesian NLP resources. I gathered links to publicly available Indonesian NLP resources. As a low resource language, it can be hard to find NLP resources for Indonesian, which is why I created this repository.

NetworkX, past contributor <https://networkx.github.io>
A Python package for creation, manipulation, and study of networks and graphs. Contributed to this project by writing the link prediction module.

Honors and Awards

DAAD Postdoc AI Net Fellowship	2025
Melbourne Research Scholarship	2019
2nd Runner-Up, Gemastik 5 Programming Competition (national-level)	2012
4th Runner-Up, ITB Programming Contest (national-level)	2011

Professional Service

Publicity chair, *SEM 2025

Conference reviewer
ARR (2021–present), ALTA (2024, 2023), EMNLP (2021), EACL (2021), ACL (2021 (secondary)), SEALP (2025), SIGTYP (2024, 2021), SemEval (2021), ACL SRW (2022, 2020), NAACL SRW (2021), ACL-IJCNLP SRW (2021), AACL-IJCNLP SRW (2023, 2020)

Ethics reviewer
ARR (2025), EMNLP Industry Track (2024)

Mentor, Australasian Language Technology Association (ALTA)	2024–present
Co-organiser, CIS Research Fellows Day	2025
Mentor, "How to get into a PhD program", ML Summer School Indonesia	2020
System administrator, NLP group, University of Melbourne	2020–2022
Reading group organiser, NLP group, University of Melbourne	2020–2022

Certifications

Data Analysis and Statistical Inference by Duke University on Coursera	2015
Statistical Learning by Stanford University on Stanford Online	2015

Skills

Natural languages	Indonesian (native), English (fluent), Japanese (intermediate), Javanese (basic oral), Thai (basic oral)
Programming languages	Python, Bash, LaTeX
Operating Systems	macOS, Ubuntu, Arch Linux
Tools/Libraries	PyTorch, Git, Pandas, FlairNLP, Transformers

Links

Website	https://kkurniawan.com
LinkedIn	https://id.linkedin.com/in/kemalkurniawan
Semantic Scholar	https://www.semanticscholar.org/author/Kemal-Kurniawan/46199596
Google Scholar	https://scholar.google.com.au/citations?user=VqORXJwAAAAJ
Github	https://github.com/kmkurn
StackOverflow	https://stackoverflow.com/users/20671531/kmkurn
ResearchGate	https://www.researchgate.net/profile/Kemal-Kurniawan
Gitlab	https://gitlab.com/kmkurn

Referees

Name	Jey Han Lau
Organisation	University of Melbourne
Position	Senior Lecturer
Contact	laujh@unimelb.edu.au
Relationship	Jey Han is my main postdoctoral research supervisor.

Name	Timothy Baldwin
Organisation	Mohamed bin Zayed University of Artificial Intelligence
Position	Provost
Contact	tb@ldwin.net
Relationship	Tim is my postdoctoral research co-supervisor.

Name	Trevor Cohn
Organisation	Google Research Australia
Position	Research Scientist
Contact	tcohn@google.com
Relationship	Trevor was my principal PhD supervisor.

Name	Lea Frermann
Organisation	University of Melbourne
Position	Senior Lecturer
Contact	lea.frermann@unimelb.edu.au
Relationship	Lea was my PhD co-supervisor and the subject coordinator of Introduction to Machine Learning subject where I was employed as a tutor.

Name	Philip Schulz
Organisation	Amazon
Position	Applied Science Manager
Contact	phschulz@amazon.com
Relationship	Philip was a collaborator in my PhD research and my manager during my internship at Amazon.