李朋軒 [liˇ pʰəŋˊ ɕyɛnˊ]

Li Peng-Hsuan

886 965-444-660
jacobvsdanniel@gmail.com
jacobvsdanniel.github.io

Education

2020 **University Fellowship**, *Department of Linguistics*, Ohio State University.
- GPA 4.00/4.00, TOEFL 112/120, GRE 335/340.
- University Fellowship for the doctoral program.

2015–2017 **M.Sc.**, *Department of Computer Science & Information Engineering*, National Taiwan University.
- GPA 3.90/4.30, ranking 70/152.

2011–2015 **B.Sc.**, *Department of Computer Science & Information Engineering*, National Taiwan University.
- GPA 4.08/4.30, ranking 4/111.
- The Presidential Awards: Autumn 2013, Autumn 2012, Spring 2012.

Experience

2021–2025 **ML Engineer**, *Taiwan AI Labs* 台灣人工智慧實驗室.
- Research publications.
  – Knowledge graph (BIB)
  – Biomedical literature mining (NAR)
- Developed LORE (https://github.com/ailabstw/LORE).
  – LLM-based knowledge graph and embedding from literature.
  – Entity-entity association prediction with sparse labels.
- Developed pubmedKB (https://github.com/jacobvsdanniel/pubmedkb_core).
  – End-to-end relation extraction for biomedical literature.
  – Entity recognition and normalization for genes, variants, diseases, chemicals, and populations.
  – Open relation extraction, closed relation extraction, and odds ratio statistics extraction.
- Other research and developments
  – (Main developer) plant gene function annotation: collaborate with IPB, NTU 台大植科所
  – (Main developer) phenotype-to-gene annotation: collaborate with Taipei Veterans General Hospital 台北榮總
  – (Joint research) polygenic risk prediction; T-cell receptor binding; single-cell gene regulatory networks

2018–2020 **Research Assistant**, *Institute of Information Science, Academia Sinica* 中央研究院.
- Research publications.
  – Named entity recognition (AAAI, EMNLP)
  – Relation extraction (ACL)
  – Knowledge graph (LREC)
- Developed CkipTagger (https://github.com/ckiplab/ckiptagger).
  – A state-of-the-art, light weight, customizable neural package for traditional Chinese.
  – Integrating word segmentation, part-of-speech tagging, and named entity recognition.
- Competed in the 2018-2019 Formosa Grand Challenge 科技大擂台.
  – Spoken question answering for traditional Chinese.
  – 2nd-place in the short-answer track, 3rd-place in the multiple-choice track, out of 143 teams.
  – Developed the strongest models for the team in both tracks; model ensemble; speech recognition denoising and segmentation.
- Competed in the IJCNLP-2017 DSAP shared task: sentiment analysis for Chinese phrases.
  – Developed all the models; 1st-place in the arousal track, 3rd-place in the combined track, out of 13 teams.

2014–2016 **QA Engineer**, *Trend Micro* 趨勢科技.
- Developed integrated test automation for an enterprise Intrusion Prevention System (IPS).
- For the user-awareness module: Active Directory and LDAP test environments.
- For the TLS stack: man-in-the-middle connections, handshake protocols, data encryption protocols.
- For the application traffic scan engine: extract packets in anomalous applications.

Summer 2013 **Intern Engineer**, *InfoKeyVault Technology* 銓安智慧科技.
- Porting a secure, encrypted custom Windows file system to a custom Linux Virtual File System.

Part-time **Teaching Experience**.
- Spring 2016, Teaching Assistant (Introduction to AI), *Department of CSIE, NTU* 台灣大學.
  – Designing program assignments and test automation; tutoring and reviewing term projects and term exams.
- 2011-2013, Math Tutor, *HER-JER Math After-class School* 赫哲數學.
  – Tutoring high school students; writing detailed solutions for term exams and college entrance exams.

Publications

2025 **A large language model framework for literature-based disease–gene association prediction**.
- *Peng-Hsuan Li*, Yih-Yun Sun, Hsueh-Fen Juan, Chien-Yu Chen, Huai-Kuang Tsai, Jia-Hsin Huang.
- Briefings in Bioinformatics (BIB).

2022 **pubmedKB: An Interactive Web Server to Explore Biomedical Entity Relations from Biomedical Literature**.
- *Peng-Hsuan Li*, Ting-Fu Chen, Jheng-Ying Yu, Shang-Hung Shih, Chan-Hung Su, Yin-Hung Lin, Huai-Kuang Tsai, Hsueh-Fen Juan, Chien-Yu Chen, and Jia-Hsin Huang.
- Nucleic Acids Research (NAR).

2020 **Why Attention? Analyze BiLSTM Deficiency and Its Remedies in the Case of NER**.
- *Peng-Hsuan Li*, Tsu-Jui Fu, Wei-Yun Ma.
- In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI).

2020 **CA-EHN: Commonsense Analogy from E-HowNet**.
- *Peng-Hsuan Li*, Tsan-Yu Yang, Wei-Yun Ma.
- In Proceedings of The 12th Language Resources and Evaluation Conference (LREC).

2019 **GraphRel: Modeling Text as Relational Graphs for Joint Entity and Relation Extraction**.
- Tsu-Jui Fu, *Peng-Hsuan Li*, Wei-Yun Ma.
- In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL).

2019 **HWE: Word Embedding with Heterogeneous Features**.
- Jhih-Sheng Fan, Mu Yang, *Peng-Hsuan Li*, Wei-Yun Ma.
- In Proceedings of 2019 IEEE 13th International Conference on Semantic Computing (ICSC).

2017 **Leveraging Linguistic Structures for Named Entity Recognition with Bidirectional Recursive Neural Networks**.
- *Peng-Hsuan Li*, Ruo-Ping Dong, Yu-Siang Wang, Ju-Chieh Chou, Wei-Yun Ma.
- In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP).

2017 **CKIP at IJCNLP-2017 Task 2: Neural Valence-Arousal Prediction for Phrases**.
- *Peng-Hsuan Li*, Wei-Yun Ma, Hsin-Yang Wang.
- In Proceedings of the IJCNLP 2017, Shared Tasks.