

# YIPING JIN

Google Scholar | GitHub  
(+34) 637235961 ◇ yiping [at] u-nus.edu  
C/ Jonqueres 4, 1-2 08003 Barcelona

## EDUCATION

---

### **Chulalongkorn University, Thailand**

*Jan 2019 - Apr 2022*

Ph.D. in Computer Science

Dissertation title: Natural Language Processing for Digital Advertising

Recipient of 100th Anniversary Doctoral Scholarship

Teach machine learning and deep learning classes

### **Chulalongkorn University, Thailand**

*Jan 2017 - Dec 2018*

M.Sc. in Computer Science

Thesis title: Lightly-Supervised Learning Methods for One-Class Text Classification

Ranked No. 1 among all Master's students in recent years

Overall GPA 4 / 4

### **National University of Singapore**

*Aug 2009 - May 2013*

B.Comp. in Computer Science (1st Class Honors Degree)

Thesis title: Mining Definition of Term in Scientific Articles

Thesis advisor: Assoc Prof Min-Yen Kan

Recipient of Singapore Ministry of Education Full Scholarship

Dean's List Award 3 out of 7 semesters

Overall CAP: 4.68 / 5

## WORK EXPERIENCE

---

### **Pompeu Fabra University**

*Oct 2022 - NOW*

*Visiting Postdoc Researcher*

*Spain*

- Research into hate speech detection
- Advisor: Prof Leo Wanner

### **Knorex**

*Jan 2021 - Dec 2022*

*Head of R&D/Staff Research Scientist*

*Singapore*

- Supervise researchers in NLP and data science
- Set R&D roadmap

### **Knorex**

*Jan 2018 - Dec 2020*

*Senior Research Scientist*

*Singapore*

- Lead of NLP team
- Interim lead of data science team (April 2017 - Oct 2018)
- Deliver research grant projects

### **Knorex**

*Sep 2015 - Dec 2017*

*Senior Research Engineer*

*Singapore*

- Research into NLP and information extraction projects
- Build up data science team from scratch
- Lead the team to work on AI-powered real-time bidding system

**Baidu-I<sup>2</sup>R Research Centre, A-STAR**  
*Research Engineer*

Jan 2014 - June 2015  
*Singapore*

- Research into NLP and information extraction projects
- Advisor: Dr. Jian Su

**National University of Singapore**  
*Undergraduate Researcher & Research Assistant*

Nov 2012 - Dec 2013  
*Singapore*

- Developed spatio-temporal data mining and visualization tools
- Developed cloud-based medical image classification system

## TECHNICAL STRENGTHS

---

<b>Expertises</b>	Natural Language Processing
<b>Research Focus</b>	Information Retrieval, Text Mining, Natural Language Generation
<b>Computer Languages</b>	Python, Java
<b>ML Libraries</b>	Scikit-Learn, NLTK, Spacy, Mallet, etc.
<b>DL Libraries</b>	HuggingFace, PyTorch, TensorFlow, Fast.ai

## GLOBAL EXPERIENCE

---

**ETH Zurich**

Feb 2012 - July 2012

- Student Exchange Programme

**Universität Leipzig, Germany**

July 2011 - Aug 2011

- Summer Language Course
- Recipient of German Academic Exchange Service (DAAD) Scholarship
- Topped the class among many students majored in German studies

**Universität Ulm, Germany**

June 2010 - July 2010

- International Summer Programme
- Recipient of Baden-Württemberg Scholarship, Germany
- Grade: A for all the courses

## LANGUAGE SKILLS

---

- Fluent spoken & written English and Mandarin
- Working proficiency for German and Spanish (C1)
- Conversational Thai (B1)
- Basic Catalan (A1)

## PROFESSIONAL SERVICES

---

- EMNLP 2021-2023: reviewer (Areas: Information Retrieval and Text Mining, Generation)
- ACL 2023: reviewer (Area: Information Retrieval and Text Mining)
- ACL ARR: reviewer since Jan 2022 (Areas: Information Retrieval and Text Mining, Generation)
- NLE Journal: reviewer since Aug 2023
- NLPCC 2022: reviewer (Area: Information Extraction)

## COMPLETED RESEARCH GRANT PROJECTS

---

### Enterprise Development Grant

Apr 2022 - Sep 2022

*Lead Research Scientist*

*Singapore*

- Project title: Automatic Planning & Execution of Cross-Channel Digital Advertising Campaigns with Knorex KAIROS Autopilot
- Grant giver: Infocomm Media Development Authority

### Capability Development Grant

Jan 2019 - Sep 2019

*Senior Research Scientist*

*Singapore*

- Project title: SENTINEL: Intelligence-as-a-Service Platform for Content Moderation at Scale
- Grant giver: Infocomm Media Development Authority

### Capability Development Grant

July 2015 - June 2016

*Senior Research Engineer*

*Singapore*

- Project title: KGen: SaaS platform for Intelligent Auto Generation of Contextual Summaries and Questions in Natural Language
- Grant giver: Infocomm Development Authority
- Responsibility: Develop headline and question generation systems

## REFEREES

---

- Prof Leo Wanner, `leo.wanner [at] upf.edu`
- Asst Prof Dittaya Wanvarie, `Dittaya.W [at] chula.ac.th`
- Assoc Prof Min-Yen Kan, `kanmy [at] comp.nus.edu.sg`
- Phu Le (Group General Manager), `le_phu [at] knorex.com`
- Dr. Su Jian sujian `[at] i2r.a-star.edu.sg`

## PATENTS

---

Yiping Jin. "Cross-domain contextual targeting without any in-domain labelled data." *U.S. Patent No. 11,093,969*. 17 Aug. 2021.

## PUBLICATIONS

---

Yiping Jin, Min-Yen Kan, Jun-Ping Ng and Xiangnan He (2013). Mining Scientific Terms and their Definitions: A Study of the ACL Anthology. *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP 2013)*. Seattle, USA.

Yiping Jin, Phu Le (2016). Selecting Domain-Specific Concepts for Question Generation With Lightly-Supervised Methods. *Proceedings of the 9th International Natural Language Generation Conference (INLG)*. Edinburgh, Scotland.

Yiping Jin, Dittaya Wanvarie, Phu Le (2017). Combining Lightly-Supervised Text Classification Models for Accurate Contextual Advertising. *Proceedings of the 8th International Joint Conference on Natural Language Processing (IJCNLP 2017)*. Taipei, Taiwan.

Bao-Dai Nguyen-Hoang, Bao-Tran Pham-Hong, Yiping Jin, Phu T.V.Le (2018). Genre-Oriented Web Content Extraction with Deep Convolutional Neural Networks and Statistical Methods. *Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation (PACLIC 32)*. Hong Kong, China.

- Yiping Jin, Dittaya Wanvarie, Phu Le (2019). Bridging the Gap between Research and Production with CODE. *Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2019)*. Macau, China.
- Nontawat Charoenphakdee, Jongyeong Lee, Yiping Jin, Dittaya Wanvarie, Masashi Sugiyama (2019). Learning Only from Relevant Keywords and Unlabeled Documents. *Proceedings of Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Hong Kong, China.
- Yiping Jin, Dittaya Wanvarie, Phu Le (2020). Learning From Noisy Out-of-Domain Corpus Using Dataless Classification. *Natural Language Engineering*. Cambridge University Press, UK.
- Yiping Jin, Akshay Bhatia, Dittaya Wanvarie (2021). Seed Word Selection for Weakly-Supervised Text Classification with Unsupervised Error Estimation. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop (NAACL SRW)*. Online.
- Yiping Jin, Vishakha Kadam, and Dittaya Wanvarie (2021). Bootstrapping Large-Scale Fine-Grained Contextual Advertising Classifier from Wikipedia. *Proceedings of the Fifteenth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-15)*.
- Yiping Jin, Akshay Bhatia, Dittaya Wanvarie, Phu Le (2022). Towards Improving Coherence and Diversity of Slogan Generation. *Natural Language Engineering*. Cambridge University Press, UK.
- Yiping Jin, Vishakha Kadam, and Dittaya Wanvarie (2022). Plot Writing From Pre-Trained Language Models. *Proceedings of the 15th International Conference on Natural Language Generation (INLG)*. Maine, USA.
- Vishakha Kadam, Yiping Jin, and Bao-Dai Nguyen-Hoang (2022). Automated Ad Creative Generation. *Proceedings of the 15th International Conference on Natural Language Generation (INLG): System Demonstrations*. Maine, USA.
- Yiping Jin, Leo Wanner, Vishakha Kadam, and Alexander Shvets (2023). Towards Weakly-Supervised Hate Speech Classification Across Datasets. *Proceedings of the 7th Workshop on Online Abuse and Harms (WOAH 2023)*. Toronto, Canada.
- Yiping Jin, Leo Wanner, and Alexander Shvets (2024). GPT-HateCheck: Can LLMs Write Better Functional Tests for Hate Speech Detection? *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. Turin, Italy.
- Chen Huang, Yiping Jin, Ilija Ilievski, Wenqiang Lei, and Jiancheng Lv (2024). ARAIDA: Analogical Reasoning-Augmented Interactive Data Annotation. To appear in *The 62nd Annual Meeting of the Association for Computational Linguistics (ACL 2024)*. Bangkok, Thailand.