Peinan Zhang

Senior Research Scientist working on NLP at CyberAgent AI Lab, Shibuya, Japan
zhang_peinan@cyberagent.co.jp peinan.cc in/peinan-zhang peinan

ABOUT ME

I am a senior research scientist with a strong background in NLP and a proven track record in bridging academic research with real-world advertising applications.

Background: I specialized in natural language processing (NLP). After starting my career as a machine learning engineer, I currently serve as a researcher, engineer, and manager. My current focus is on (1) applying NLP and machine learning technologies to the advertising domain, and (2) maximizing the performance of my team.

Work: As an active contributor, I incorporate methods from both domestic and international research that align with our business strategies, implement prototypes, and validate their effectiveness. I also publish and present new ideas and methods derived from this process at conferences, bridging the gap between industry and academia. Recently, my responsibilities have expanded to include management tasks, where I focus on maximizing team performance.

Achievements: I led a research project focused on automatically generating effective advertising texts, which culminated in the launch of a product called KiwamiTD centered around this capability. I also transformed product challenges into research topics, resulting in papers accepted at prestigious international conferences such as ACL, NAACL, and EMNLP, which are among the top-tier conferences in the NLP field.

EDUCATION

Masters in Engineering Apr 2014 — Mar 2016
Tokyo Metropolitan University, Tokyo, Japan

- Researched sentiment analysis tasks with deep learning techniques
- *Related coursework*: Majored in Natural Language Processing, Machine Learning, Artificial Intelligence, and Computer Science
- *Thesis:* Japanese Sentiment Classification with Stacked Denoising Auto-Encoder using Distributed Word Representation
- Adviser: Mamoru Komachi
- *University Information*: 1-1 Minami-Osawa, Hachioji-shi, Tokyo, 192-0397, Japan, +81-42-677-1111, https://www.tmu.ac.jp/english/index.html

Bachelors in Engineering Apr 2010 — Mar 2014
Tokyo Metropolitan University, Tokyo, Japan

- Researched dialogue and discourse with deep learning techniques
- *Related coursework*: Majored in Natural Language Processing, Machine Learning, Artificial Intelligence, and Computer Science
- *Thesis:* Dependency-Aware Deep Learning for Sentiment Classification
- Adviser: Mamoru Komachi

High School Diploma Apr 2007 — Mar 2010
Kaijo High School, Tokyo, Japan

- *School Information*: 3-6-1 Okubo, Shinjuku-ku, Tokyo, 169-0072, Japan, +81-3-3209-5880, https://www.kaijo.ed.jp/

EXPERIENCE

Senior Research Scientist Oct 2022 — Present
Research Scientist Jun 2018 — Oct 2022
CyberAgent, Inc., Tokyo, Japan

- Applying NLP and machine learning techniques, such as large language models (LLM), natural language generation, few-/zero-shot learning, and pre-training/fine-tuning, to advertising domain
- Implementing and validating research methods aligned with business strategies
- Developing product, such as KiwamiTD, for automatic advertising text generation
- Publishing papers at top-tier conferences including ACL, NAACL, and EMNLP
- Managing research team and maximizing team performance
- *Company Information*: Abema Towers, 40-1, Udagawa-Cho, Shibuya-Ku, Tokyo, 150-0042, Japan, +81-3-5459-0202, https://www.cyberagent.co.jp/en/

Machine Learning Engineer Mar 2016 — Apr 2018
Yahoo! Japan (now LY Corporation), Tokyo, Japan

- Implemented personalization feature for push notifications for over 10 million users
- Applied NLP/ML technologies with Hadoop for large-scale data processing
- Deployed solution as in-house application
- *Company Information*: Kioi Tower, 1-3 Kioicho, Chiyoda-ku, Tokyo, 102-8282, Japan, +81-3-6779-4900, https://www.lycorp.co.jp/en/

Internship Jul 2014 — Dec 2014
Honda Research Institute Japan Co., Ltd., Saitama, Japan

- Developed speech act classification system
- Built language model for speech recognition using Nested Pitman-Yor Language Model
- *Company Information*: 8-1 Honcho, Wako-shi, Saitama 351-0188, Japan, https://www.jp.honda-ri.com/en/

PUBLICATIONS

Journal Articles

1. Ukyo Honda, Tatsushi Oka, Peinan Zhang and Masato Mita. 2024. **Not Eliminate but Aggregate: Post-Hoc Control over Mixture-of-Experts to Address Shortcut Shifts in Natural Language Understanding**. *Transactions of the Association for Computational Linguistics (TACL)* (Presented at EMNLP 2024).

2. Yifan Zhou, Wenyuan Qi, Yuyin Zhang and Peinan Zhang. 2021. **Investigation on spray cyclic variations under idle operation of engine using optical diagnostics and statistical methods**. *International Journal of Engine Research 2021*.

Conference Papers

1. Murakami Soichiro, Peinan Zhang, Hidetaka Kamigaito, Hiroya Takamura, and Manabu Okumura. 2025. **AdParaphrase v2.0: Generating Attractive Ad Texts Using a Preference-Annotated Paraphrase Dataset**. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL 2025 Findings)*.

2. Peinan Zhang, Yusuke Sakai, Masato Mita, Hiroki Ouchi and Taro Watanabe. 2025. **AdTEC: A Unified Benchmark for Evaluating Text Quality in Search Engine Advertising**. *Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2025)*.

3. Soichiro Murakami, Peinan Zhang, Hidetaka Kamigaito, Hiroya Takamura and Manabu Okumura. 2025. **AdParaphrase: Paraphrase Dataset for Analyzing Linguistic**

Features toward Generating Attractive Ad Texts. *Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2025 Findings)*.

4. Hidetaka Kamigaito, Soichiro Murakami, Peinan Zhang, Hiroya Takamura and Manabu Okumura. 2024. **Generating Attractive Ad Text by Facilitating the Reuse of Landing Page Expressions**. *International Conference on Natural Language Generation (INLG 2024)*.
5. Yuu Jinnai, Ukyo Honda, Tetsuro Morimura and Peinan Zhang. 2024. **Generating Diverse and High-Quality Texts by Minimum Bayes Risk Decoding**. *Annual Meeting of the Association for Computational Linguistics (ACL 2024 Findings)*.
6. Masato Mita, Soichiro Murakami, Akihiko Kato and Peinan Zhang. 2024. **Striking Gold in Advertising: Standardization and Exploration of Ad Text Generation**. *Annual Meeting of the Association for Computational Linguistics (ACL 2024)*.
7. Tetsuro Morimura, Kazuhiro Ota, Kenshi Abe and Peinan Zhang. 2024. **Policy Gradient Algorithms with Monte Carlo Tree Learning for Non-Markov Decision Processes**. *Reinforcement Learning Conference (RLC 2024)*.
8. Ye Xiong, Hidetaka Kamigaito, Soichiro Murakami, Peinan Zhang, Hiroya Takamura and Manabu Okumura. 2024. **Grasping Both Query Relevance and Essential Content for Query-focused Summarization**. *International Conference on Research and Development in Information Retrieval (SIGIR 2024)*.
9. Sho Hoshino, Akihiko Kato, Soichiro Murakami and Peinan Zhang. 2024. **Cross-lingual Transfer or Machine Translation? On Data Augmentation for Monolingual Semantic Textual Similarity**. *Language Resources and Evaluation Conference - International Conference on Computational Linguistics (LREC-COLING 2024)*.
10. Go Inoue, Akihiko Kato, Masato Mita, Ukyo Honda and Peinan Zhang. 2024. **CAMERA^3: An Evaluation Dataset for Controllable Ad Text Generation in Japanese**. *Language Resources and Evaluation Conference - International Conference on Computational Linguistics (LREC-COLING 2024)*.
11. Soichiro Murakami, Peinan Zhang, Sho Hoshino, Hidetaka Kamigaito, Hiroya Takamura and Manabu Okumura. 2022. **Aspect-based Analysis of Advertising Appeals for Search Engine Advertising**. *Annual Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT 2022 Industry Track)*.
12. Shunyo Kawamoto, Yu Sawai*, Kohei Wakimoto* and Peinan Zhang*. 2021. **FAST: Fast Annotation tool for SmarT devices**. *Conference on Empirical Methods in Natural Language Processing (EMNLP 2021 System Demonstrations)*.
13. Hidetaka Kamigaito*, Peinan Zhang*, Hiroya Takamura and Manabu Okumura. 2021. **An Empirical Study of Generating Texts for Search-Engine Advertising**. *Annual Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT 2021 Industry Track)*.
14. Peinan Zhang and Mamoru Komachi. 2015. **Japanese Sentiment Classification with Stacked Denoising Auto-Encoder using Distributed Word Representation**. *Pacific Asia Conference on Language, Information and Computation (PACLIC 2015)*.

Preprints

1. Akihiko Kato, Masato Mita, Soichiro Murakami, Ukyo Honda, Sho Hoshino and Peinan Zhang. 2024. **FaithCAMERA: Construction of a Faithful Dataset for Ad Text Generation**. *arXiv*.

2. Soichiro Murakami, Sho Hoshino and Peinan Zhang. 2023. **Natural language generation for advertising: A survey**. *arXiv*.

Non-Peer-Reviewed Publications in Japanese

- 大竹 啓永, 張 培楠, 坂井 優介, 三田 雅人, 大内 啓樹, 渡辺 太郎. 2025. **広告画像ランキングによる視覚言語モデルの評価**. 言語処理学会 2025
- 本多 右京, 岡 達志, 張 培楠, 三田 雅人. 2025. **Mixture-of-Experts の悲観的な統合による頑健な自然言語理解**. 言語処理学会 2025
- 山田 康介, 張 培楠. 2025. **訓練不要な条件付きテキスト埋め込み**. 言語処理学会 2025
- 加藤 明彦, 三田 雅人, 村上 聡一朗, 本多 右京, 星野 翔, 張 培楠. 2025. **FaithCAMERA: 広告文生成タスクのための忠実性を担保した評価データセットの構築**. 言語処理学会 2025
- 村上 聡一朗, 張 培楠, 上垣外 英剛, 高村 大也, 奥村 学. 2025. **AdParaphrase: 魅力的な広告表現の分析を目的とした広告文言い換えデータセット**. 言語処理学会 2025
- 村上 聡一朗, 菊田 洸, 張 培楠, 上垣外 英剛, 高村 大也, 奥村 学. 2023. **原文の書き換えによる広告文生成**. 言語処理学会 2023
- 三田 雅人, 村上 聡一朗, 張 培楠. 2023. **広告文生成タスクの規定とベンチマーク構築**. 言語処理学会 2023
- 星野 翔, 張 培楠. 2023. **自然言語生成におけるタスク横断自動評価のメタ分析**. 言語処理学会 2023
- 張 培楠, 坂井 優介, 三田 雅人, 大内 啓樹, 渡辺 太郎. 2023. **AdGLUE: 広告言語理解ベンチマーク**. 言語処理学会 2023
- 加藤 明彦, 大田 和寛, 村上 聡一朗, 三田 雅人, 本多 右京, 張 培楠. 2023. **広告データセットに内在する幻覚の分析**. 言語処理学会 2023
- 森村 哲郎, 大田 和寛, 阿部 拳之, 張 培楠. 2022. **ビームサーチ推論のための強化学習**. IBIS 2022
- 村上 聡一朗, 星野 翔, 張 培楠, 上垣外 英剛, 高村 大也, 奥村 学. 2022. **LP-to-Text: マルチモーダル広告文生成**. 言語処理学会 2022
- 大曽根 宏幸, 張 培楠. 2021. **GPT-2 の転移学習によるキーワードを考慮した広告文生成**. 人工知能学会 2021
- 川本 峻頌, 澤井 悠, 張 培楠, 脇本 宏平. 2021. **FAST: スマートデバイス用の高速なアノテーションツール**. 人工知能学会 2021
- 澤井 悠, 張 培楠, 吉本 暁文. 2020. **自動生成された広告文の人手評価における評価指標と支援ツールの提案**. 人工知能学会 2020
- 脇本 宏平, 川本 峻頌, 張 培楠. 2020. **インターネット広告におけるキーワードに基づく広告文の自動生成**. 人工知能学会 2020
- 川本 峻頌, 張 培楠. 2020. **スタイル制御を考慮した多様な広告文生成**. 言語処理学会 2020
- 上垣外 英剛, 張 培楠, 高村 大也, 奥村 学. 2019. **広告効果を報酬とした強化学習に基づく広告文の自動生成**. WebDB Forum 2019
- 張 培楠. 2019. **EC サイトにおける商品タイトルからの商品名抽出**. 人工知能学会 2019

Note: Asterisk (*) denotes equal contribution. Further information can also be found on the personal page, ACL Anthology, and Google Scholar.

TALKS, WRITINGS, AND PRESS

- zenn.dev. **戦いは数だよ兄貴！ Google Cloud Batch による並列化入門 〜LLM の前処理を例に〜**. 9 December 2024.
  https://zenn.dev/peinan/articles/721a908599dcf7

- CyberAgent Press Releases. **AI Lab、計算言語学・自然言語処理分野に関する主要ジャーナル「Transactions of the Association for Computational Linguistics」にて論文採択 ―短絡的な予測への依存を抑制する手法を提案―**. 30 July 2024.
  https://www.cyberagent.co.jp/news/detail/id=30555
- CyberAgent Press Releases. **AI Lab、自然言語処理分野のトップカンファレンス「ACL 2024」にて 3 本の論文採択**. 24 July 2024.
  https://www.cyberagent.co.jp/news/detail/id=30532
- CyberAgent Press Releases. **AI Lab、情報検索・推薦システム分野のトップカンファレンス「SIGIR 2024」にて論文採択 ー原文書の重要情報とクエリ関連性を捉えたクエリ指向要約モデルを提案ー**. 27 June 2024.
  https://www.cyberagent.co.jp/news/detail/id=30417
- CyberAgent Press Releases. **AI Lab、自然言語処理分野の国際会議「LREC-COLING 2024」にて 4 本の論文採択 ー広告文の自動生成やテキスト音声合成技術に関する分析・新手法を提案ー**. 10 April 2024.
  https://www.cyberagent.co.jp/news/detail/id=30046
- CyberAgent Press Releases. **AI Lab、自然言語処理分野のトップカンファレンス「NAACL-HLT 2022」にて主著論文採択 ― 効果の高い広告訴求を分析 ―**. 28 April 2022.
  https://www.cyberagent.co.jp/news/detail/id=27559
- CyberAgent Developer Conference 2022. **自然言語処理を用いた効果的な広告テキストの自動生成**. 24 March 2022.
  https://cadc.cyberagent.co.jp/2022/program/automatically-generate-effective-sales-copy/
- CyberAgent Press Releases. **自然言語処理分野のトップカンファレンス「EMNLP 2021」の System Demonstration Track にて論文採択 ― モバイル端末用の効率的なアノテーションツールを提案 ―**. 12 October 2021.
  https://www.cyberagent.co.jp/news/detail/id=26746
- CyberAgent Press Releases. **AI Lab、自然言語処理分野のトップカンファレンス「NAACL-HLT 2021」にて共著論文採択　– 広告効果を考慮した広告文生成手法を提案 –**. 14 April 2021.
  https://www.cyberagent.co.jp/news/detail/id=26075
- CyberAgent Way. **効果を出す「AI の共同研究」 東京工業大学 奥村・高村研究室と創る自然言語処理の未来**. 23 December 2020.
  https://www.cyberagent.co.jp/way/features/list/detail/id=25604
- CyberAgent IR Channel at YouTube. **「効果を出す AI - サイバーエージェントの AI 研究とビジネス実装力 -」 CyberAgent IR チャンネル 第 23 弾**. 3 December 2020.
  https://www.youtube.com/watch?v=XuxNcucXSMU
- CyberAgent Press Releases. **AI 事業本部「極予測 AI チーム」のデータサイエンティストによる論文が人工知能学会全国大会 (JSAI2020) において優秀賞を受賞**. 5 August 2020.
  https://www.cyberagent.co.jp/techinfo/news/detail/id=25008
- CyberAgent Press Releases. **AI で数億キーワードの品質スコアを自動で改善　効果の出せる広告テキストを自動生成「極予測 TD」の提供を開始**. 22 May 2020.
  https://www.cyberagent.co.jp/news/detail/id=24670
- テキストアナリティクス・シンポジウム. **ウェブサービス事業者における研究開発インターン**. 27 September 2019.
  https://www.ieice.org/ nlc/wiki/wiki.cgi
- CyberAgent AI Lab Blog. **人工知能学会 2019 参加報告**. 25 June 2019.
  https://cyberagent.ai/blog/research/2219
- Speakerdeck. **A Deep Reinforced Model for Abstractive Summarization**. 29 March 2019.

https://speakerdeck.com/peinan/a-deep-reinforced-model-for-abstractive-summarization

- CyberAgent AI Lab Blog. **木星を継ぐもの ～JupyterLab よこんにちは～**. 20 March 2019.
  https://cyberagent.ai/blog/research/10260
- CyberAgent AI Lab Blog. **Tweet ベースマッチングシステムを支える技術**. 5 February 2019.
  https://cyberagent.ai/blog/research/10257
- テキストアナリティクス・シンポジウム. **ACL 参加報告**. 7 September 2018.
  https://speakerdeck.com/peinan/acl-can-jia-bao-gao
- CyberAgent AI Lab Blog. **ACL 2018 参加報告**. 13 August 2018.
  https://cyberagent.ai/blog/research/511
- CyberAgent AI Lab Blog. **人工知能学会 2018 参加報告**. 18 June 2018.
  https://cyberagent.ai/blog/research/86

Note: Further information can also be found on the personal page.

AWARDS

- Paper “Mixture-of-Experts の悲観的な統合による頑健な自然言語理解” received **委員特別賞** at 言語処理学会 2025.
- Paper “FaithCAMERA: 広告文生成タスクのための忠実性を担保した評価データセットの構築” received **委員特別賞** at 言語処理学会 2025.
- Paper “広告文生成タスクの規定とベンチマーク構築” received **委員特別賞** at 言語処理学会 2023.
- Paper “インターネット広告におけるキーワードに基づく広告文の自動生成” received **優秀賞** at 人工知能学会 2021.
- Received **Best Engineer Award** at CyberAgent AWARDS 2021 (全社表彰)

PATENTS

1. ヤフー株式会社. (2022). 情報管理システム、情報管理方法、およびプログラム. 特開 2022-087118.
2. 株式会社サイバーエージェント. (2021). 広告文自動作成システム. 特開 2021-140228.
3. ヤフー株式会社. (2020). 情報管理システム、情報管理方法、およびプログラム. 特開 2020-042481.
4. ヤフー株式会社. (2019). 情報処理システム、情報処理方法、およびプログラム. 特開 2019-164654.
5. ヤフー株式会社. (2018). 決定装置、決定方法、及び決定プログラム. 特開 2018-156460.
6. ヤフー株式会社. (2018). 表示プログラム、表示方法、端末装置、生成装置、生成方法、及び生成プログラム. 特開 2018-156421.
7. ヤフー株式会社. (2018). 表示プログラム、表示方法、端末装置、情報処理装置、情報処理方法、及び情報処理プログラム. 特開 2018-156186.

SKILLS

Computational

- **Python** (Over 10 years of experience, including proficiency with mainstream data science libraries such as PyTorch, NumPy, Pandas, and others, as well as building and publishing open-source Python libraries.)

- **Google Cloud Platform** (Experience building infrastructure for large-scale projects)
- Miscellaneous
  - *Python Libraries*: FastAPI, Pydantic, PydanticAI, OpenAI, HuggingFace, Jupyter, Plotly, Streamlit, Gradio, Reflex, etc.
  - *Programming Languages and Frameworks*: ShellScript (Bash/Zsh), LaTeX, Golang, Make, JavaScript (TypeScript), React, Next.js, etc.
  - *Editors*: Vim, VSCode (Cursor), PyCharm, etc.
  - *Tools for Development*: macOS, Linux/Unix, Git, GitHub, GitHub Actions, Docker, Kubernates, uv (Python package manager), etc.
  - *Tools for Productivity*: Raycast, Notion, Overleaf, Figma, Arc, Slack, etc.
  - *Other Tools*: Illustrator, Photoshop, Lightroom, etc.
- For more information, please refer to my GitHub profile at `https://github.com/peinan`

Languages

- **Japanese**: Native proficiency (Lived in Japan for over 20 years since age 10)
- **Chinese**: Native proficiency (Born in China and lived there until age 10)
- **English**: Advanced proficiency (Lived in the United States for one year when I was 12)
  - TOEFL iBT: 94 (R: 25, W: 24, L: 23, S: 22)