TATSUYA AOKI, PH.D.

j20.at09303us@gmail.com ⋄ LinkedIn: a-tatsuya

SUMMARY

Senior Machine Learning Engineer with 5+ years of experience at Apple, specializing in On-Device LLMs, NLP, and Multilingual AI. Proven track record of deploying generative models to hundreds of millions of users (Apple Intelligence). Expert in optimizing transformer architectures for resource-constrained environments (mobile/edge) and handling complex CJK linguistic challenges.

EXPERIENCE

Apple Inc.   May 2022 - Present
Senior Machine Learning Engineer   Oct 2023 - Present
Machine Learning Engineer   May 2022 - Sept 2023

- **Apple Intelligence On-Device LLMs:** Led the development and deployment of on-device Large Language Models powering core features like *Writing Tools* and *Genmoji* for East Asian markets (Japanese, Korean, Chinese).
- **Model Optimization:** Architected a novel adaptation strategy to fit large foundation models onto mobile devices with strict memory and latency constraints, enabling high-fidelity generation without server dependency.
- **Evaluation Infrastructure:** Built robust tooling and automated evaluation pipelines for generative tasks, significantly reducing the iteration cycle time for model tuning.

Apple Japan Inc.   Oct 2020 - May 2022
Input Method Engineer

- **Autocorrect System:** Spearheaded language modeling improvements for Japanese, Korean, and Thai keyboards. Owned the full lifecycle from data curation and model training to C++ runtime integration.
- **Quick Path (Swipe Typing):** Played a key role in shipping the Quick Path interface for the Japanese market, ensuring high-accuracy path decoding.
- **Data-Driven Engineering:** Conducted large-scale analysis of user interaction logs to identify systemic errors in morphological analysis, resulting in a measurable improvement in candidate ranking quality.

EDUCATION

Institute of Science Tokyo (formerly Tokyo Institute of Technology)   March 2025
Ph.D. in Engineering

Tokyo Institute of Technology   March 2018
M.S. in Engineering

University of Tsukuba   March 2016
B.S. in Information Science and Media

TECHNICAL SKILLS

ML & NLP	LLMs (Llama, GPT), Transformers, BERT/RoBERTa, On-device Inference, Model Compression (Quantization, Pruning, Distillation), RLHF, Data-to-Text
Programming	Python, C++, Swift/Obj-C
Frameworks	PyTorch, TensorFlow, HuggingFace, CoreML, scikit-learn
Languages	Japanese (Native), English (Proficient), Mandarin Chinese (Elementary)

SELECTED PUBLICATIONS

[1] **Tatsuya Aoki**, J. H. Lau, H. Kamigaito, H. Takamura, T. Baldwin, and M. Okumura. Discovering unusual word usages with masked language model via pseudo-label training. *Journal of Natural Language Processing*, 2025.

[2] **Tatsuya Aoki**, A. Miyazawa, K. Aoki, K. Goshima, T. Ishigaki, I. Kobayashi, HiroyaTakamura, and Y. Miyao. Generating market comments referring to external resources. In *International Conference on Natural Language Generation*, 2018.

[3] **Tatsuya Aoki**, R. Sasano, H. Takamura, and M. Okumura. Distinguishing japanese non-standard usages from standard ones. In *Empirical Methods in Natural Language Processing*, 2017.