ALEX BOYD

✉ alexjamesboyd2@gmail.com ☐ ajboyd2 🔗 alexboyd-research.github.io

SUMMARY

Machine learning researcher with broad experience spanning deep learning R&D, including model development, large-scale data curation, and end-to-end system design. Background includes published work in top ML venues, internships at Apple, Microsoft Research, and NVIDIA, and industry work building retrieval systems, agentic workflows, and forecasting models. Skilled at translating research ideas into polished, reliable, and rapidly deployable solutions across natural language, multimodal understanding, and healthcare applications.

EXPERIENCE

GE HEALTHCARE — Bellevue, WA
AI Research Scientist — 2024 – Present

- MULTIMODAL AGENTIC SYSTEMS (END-TO-END): Architected and deployed a GenAI agent for medical device troubleshooting, integrating LLM tool-calling (REST), Docling-based document parsing, and vector search.
  - Built the full-stack solution (FastAPI, PostgreSQL, Svelte) with small team in 2.5 months, achieving $3M+ in operational savings via improved technician efficiency.
- INTERNAL ML FRAMEWORK: Developed a Python package supporting automated ML workflows, agentic task orchestration, and parallelized experimentation. This tooling enabled the rapid prototyping and deployment of the device troubleshooting agent.
- RAG FOR CLINICAL ONCOLOGY: Designed and implemented a retrieval-augmented generation (RAG) system automating clinical trial matching, synthesizing unstructured clinical notes with eligibility rules via semantic (vector) search.
- TIME-SERIES FORECASTING R&D: Led research on zero-shot and few-shot time-series methods to predict hospital operational demand, evaluating state-space, deep learning, and foundation-model approaches.

APPLE — Cupertino, CA
Applied ML Research Intern | *Document AI* — 2021

- Designed few-shot learning methods enabling Document AI systems to generalize to new document types with minimal labels.
- Developed models for automated document parsing and metadata extraction to improve enterprise workflows.
- Prototyped sequence-to-structure architectures to support heterogeneous document formats.

MICROSOFT RESEARCH — Redmond, WA
Deep Learning Research Intern | *Program Synthesis* — 2020

- Built a program-synthesis framework for extracting temporal expressions from natural-language emails to support automatic scheduling suggestions.
- Designed a DSL and interpreter for representing complex time intervals and resolving ambiguous user intent.
- Developed a hybrid training regime combining supervised learning with REINFORCE using a Wasserstein reward.

NVIDIA — Santa Clara, CA
Applied Deep Learning Research Intern | *Conversational AI* — 2019 – 2020

- Researched conversational, SFT methods to improve responses, long-context coherence, and persona control.
- Constructed large-scale multi-turn datasets by mining millions of Reddit conversations, increasing naturalness and topical coverage.

WORKDAY — Pleasanton, CA
Machine Learning Intern | *NL2SQL* — 2018
Data Science Intern | *Process Mining* — 2017

EDUCATION

UNIVERSITY OF CALIFORNIA, IRVINE | *Ph.D. Statistics* — 2018 – 2024
Advisors: **Dr. Padhraic Smyth & Dr. Stephan Mandt** - NSF Graduate Fellow — GPA: 4.00 / 4.00

CALIFORNIA POLYTECHNIC STATE UNIVERSITY | *B.S. Software Engineering* — 2014 – 2018
Advisor: **Dr. Dennis Sun** - *summa cum laude* — GPA: 4.00 / 4.00

SKILLS

ML/AI Specialties: Temporal point processes, generative models (VAEs, LLMs), RAG systems, reinforcement learning, few-shot learning, sequence modeling, probabilistic modeling, (approximate) Bayesian methods

Languages & Frameworks: Python, PyTorch, Lightning, JAX, TensorFlow, C/C++, CUDA, R, Java, SQL

Deployment/Tools: FastAPI, Svelte, Docker, Spark, AWS, Git

Data & Research: Large-scale data curation, synthetic data generation, model evaluation pipelines, statistical analysis, A/B testing, experimental design

PUBLICATIONS

IN SUBMISSION

1. **A. Boyd**, A. Warrington, T. Kass-Hout, P. Bhatia, and C. Xiao, 'Hyper Hawkes Processes: Interpretable Models of Marked Temporal Point Processes,' *(currently under review)*, 2025.

2. A. Chang, L. Huang, **A. Boyd**, P. Bhatia, T. Kass-Hout, C. Xiao, and F. Ma, 'MedSight: Towards Grounded Visual Comprehension in Medical Large Vision-Language Models,' *(currently under review)*, 2025.

3. C. Li, A. Elmahdy, **A. Boyd**, Z. Wang, A. Garcia, P. Bhatia, T. Kass-Hout, C. Xiao, and M. Hong, 'ST-PPO: Stabilized Off-Policy Proximal Policy Optimization for Multi-Turn Agents Training,' *(currently under review)*, 2025.

PEER-REVIEWED

1. Y. Chang*, **A. Boyd***, C. Xiao, T. Kass-Hout, P. Bhatia, P. Smyth, and A. Warrington, 'Deep Continuous-Time State-Space Models for Marked Event Sequences,' *Advances in Neural Information Processing Systems 38 (NeurIPS)*, 2025. **[Spotlight Presentation]**

2. A. Chang, L. Huang, **A. Boyd**, P. Bhatia, T. Kass-Hout, C. Xiao, and F. Ma, 'Focus on What Matters: Enhancing Medical Vision-Language Models with Automatic Attention Alignment Tuning,' *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (ACL)*, 2025.

3. M. Kelly, **A. Boyd**, S. Showalter, M. Steyvers, and P. Smyth, 'Bayesian Inference for Correlated Human Experts and Classifiers,' *Proceedings of the 42nd International Conference on Machine Learning (ICML)*, 2025.

4. E. Wong-Toi, **A. Boyd**, V. Fortuin, and S. Mandt, 'Understanding Pathologies of Deep Heteroskedastic Regression,' *Proceedings of the 40th Conference on Uncertainty in Artificial Intelligence (UAI)*, 2024. **[Oral Presentation]**

5. S. Showalter*, **A. Boyd***, P. Smyth, and M. Steyvers, 'Bayesian Online Learning for Consensus Prediction,' *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics (AIStats)*, 2024.

6. Y. Chang, **A. Boyd**, and P. Smyth, 'Probabilistic Modeling for Sequences of Sets in Continuous-Time,' *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics (AIStats)*, 2024. **[Oral Presentation]**

7. **A. Boyd**, Y. Chang, S. Mandt, and P. Smyth, 'Inference for Mark-Censored Temporal Point Processes,' *Proceedings of the 39th Conference on Uncertainty in Artificial Intelligence (UAI)*, 2023. **[Spotlight Presentation]**

8. **A. Boyd**, Y. Chang, S. Mandt, and P. Smyth, 'Probabilistic Querying of Continuous-Time Event Sequences,' *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics (AIStats)*, 2023.

9. **A. Boyd***, S. Showalter*, S. Mandt, and P. Smyth, 'Predictive Querying for Autoregressive Neural Sequence Models,' *Advances in Neural Information Processing Systems 35 (NeurIPS)*, 2022. **[Oral Presentation]**

10. A. Alexos*, **A. Boyd***, and S. Mandt, 'Structured Stochastic Gradient MCMC,' *Proceedings of the 39th International Conference on Machine Learning (ICML)*, 2022.

11. A. Li, **A. Boyd**, P. Smyth, and S. Mandt, 'Detecting and Adapting to Irregular Distribution Shifts in Bayesian Online Learning,' *Advances in Neural Information Processing Systems 34 (NeurIPS)*, 2021.

12. P. Putzel, H. Do, **A. Boyd**, H. Zhong, and P. Smyth, 'Dynamic Survival Analysis for EHR Data with Personalized Parametric Distributions,' *Proceedings of the 6th Machine Learning for Healthcare Conference (MLHC)*, 2021.

13. **A. Boyd**, P. Smyth, R. Bamler, and S. Mandt, 'User-Dependent Neural Sequence Models for Continuous-Time Event Data,' *Advances in Neural Information Processing Systems 33 (NeurIPS)*, 2020.

14. **A. Boyd***, R. Puri*, M. Shoeybi, M. Patwary, and B. Catanzaro, 'Large Scale Multi-Actor Generative Dialog Modeling,' *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2020.

* = equal contributions