

# SANDESH SWAMY

---

[LinkedIn](#) | [Personal Website](#) | [Google Scholar](#) | Phone: 614.815.2292 | ssandesh.2991@gmail.com

Senior Applied Scientist with 10+ years of experience in production machine learning, AI safety, and pre-/post-training LLMs. Technical lead for AWS AgentCore Policy, Multimodal Guardrails, contextual grounding guardrails, and AWS' chatbot. Published researcher in LLM safety, adversarial robustness, and agentic systems (EMNLP, EACL, ICLR). Expert in building and leading ML teams to ship safety-critical AI systems at scale.

## TECHNICAL SKILLS

---

**Programming:** Python (expert), Java, C, C++ (proficient)

**ML/AI Frameworks:** PyTorch, HuggingFace Transformers, numpy, scikit-learn, pandas

**LLM Safety & Alignment:** Guardrail architectures, adversarial robustness, red-teaming, jailbreak detection, content moderation, RLHF, prompt engineering, safety evaluation frameworks, policy enforcement systems

**Production ML Systems:** A/B testing, data annotation pipelines, multimodal AI, retrieval-augmented generation (RAG), LM training, LM fine-tuning with SFT and post-training techniques.

**Research & Communication:** Experimental design, statistical analysis, peer review, technical writing, cross-functional collaboration, team leadership

## WORK EXPERIENCE

---

### Amazon, Senior Applied Scientist

July 2017 - Present

- Led safety architecture and science strategy for AgentCore Policy Engine, managing a team of 6 scientists to develop production-scale guardrails for agentic AI systems—designed multi-layered policy enforcement preventing harmful behaviors while preserving utility (200 customers in 1 week of launch, ~500 avg requests per day, with 30s p90 latencies).
- Architected end-to-end ML pipeline for Multimodal Guardrails (text/image), leading a 5-person science team across data quality frameworks, distributed training infrastructure, and safety evaluation—deployed models processing 4 million requests/week with sub-50ms p90 latency.
- Lead scientist for Amazon Q's conversational search and Q&A, developing RAG-based system for natural language queries across AWS resources—deployed to 200K customers, achieving >70% exact match (EM) accuracy on enterprise queries.
- Pioneered AWS's first conversational AI chatbot translating natural language to executable CLI commands and resource queries—built intent classification and command generation system achieving >92% exact match accuracy and reducing developer query resolution time by 67% through model distillation.

- Led neural model deployment for Alexa skills across 15+ international markets (Asia, Europe, North America), managing cross-lingual transfer learning and locale-specific fine-tuning—launched to 10 million users while reducing locale onboarding time by 75%.
- Architected domain-adaptive training pipeline for Alexa skills, implementing few-shot learning and meta-learning techniques—reduced SEMER (semantic error rate) by 3% over production baselines.
- Co-led Alexa's transition from legacy rule-based to neural skill understanding across 50+ skill categories, coordinating science, engineering, and product teams, deployed to 100M+ devices, reduced semantic error rates (SEMER) by 17% relative while serving 5M requests/week.

**The Ohio State University, Graduate Research Assistant**

**May 2016 - May 2017**

- Developed machine learning methods for event forecasting from social media, creating novel dataset for prediction credibility assessment and training models to forecast real-world outcomes (elections, sports) from Twitter user predictions—published at EMNLP 2017.

**Cisco Systems, Software Developer**

**August 2013 - July 2015**

- Core contributor to the development of CLI parser, USB console, and LED modules for Cat2k/3k range of switches.

**SELECTED PUBLICATIONS, THESIS, AND PATENTS**

---

- Rheeya Uppaal, Phu Mon Htut, Min Bai, Nikolaos Pappas, Zheng Qi, **Sandesh Swamy**, “*Journey Before Destination: On the importance of Visual Faithfulness in Slow Thinking*”, **EACL 2026**.
- Devang Kulshreshtha, Wanyu Du, Raghav Jain, Srikanth Doss, Hang Su, **Sandesh Swamy**, Yanjun Qi, “*The Subtle Art of Defection: Understanding Uncooperative Behaviors in LLM based Multi-Agent Systems*”, **EACL 2026**.
- [PATENT]**Sandesh Swamy**, Rashmi Gangadharaiah, James W Horsley, Abhijit S Barde, Jonathan James Pezzino, “Provider network user console with natural language querying feature”, [Granted 2025](#).
- [PATENT]**Sandesh Swamy**, Rashmi Gangadharaiah, “Automatic user console question generation”, [Granted 2025](#).
- Dhruv Agarwal, Manoj Ghuhan Arivazhagan, Rajarshi Das, **Sandesh Swamy**, Rashmi Gangadharaiah, “Searching for Optimal Solutions with LLMs via Bayesian Optimization”, **ICLR 2025**.
- **Sandesh Swamy**, Narges Tabari, Chacha Chen, Rashmi Gangadharaiah, “Contextual dynamic prompting for response generation in task-oriented dialog systems”, **EACL 2023**.
- **Sandesh Swamy**, Alan Ritter, Marie-Catherine de Marneffe, “" i have a feeling trump will win.....": Forecasting Winners and Losers from User Predictions on Twitter”, **EMNLP 2017**.
- [THESIS] **Sandesh Swamy**, “Forecasting event outcomes from user predictions on Twitter”, **Master’s Thesis, 2017**.

## EDUCATION

---

**The Ohio State University** May 2017  
GPA: 3.7/4.0  
*Master of Science in Computer Science (NLP and Machine Learning)*  
*Advisers: Prof. Alan Ritter and Prof. Marie-Catherine de Marneffe*

**R.V. College of Engineering** July 2013  
GPA: 9.4/10.0  
*Bachelor of Engineering in Computer Science and Engineering*

## RESEARCH COMMUNITY SERVICE

---

**Conference Reviewing:** COLING (2018-2024), ACL-SRW (2018, 2020), EMNLP (2019-2025), ACL (2020, 2023), ARR (since inception), AMLC (Amazon internal ML conference) (2017-2023), NAACL 2022 Session Chair.