

Nesreen K. Ahmed, Ph.D.

 n.kamel@gmail.com  nkahmed
 <http://nesreenahmed.com/>  Google Scholar

Summary

Senior Principal AI Scientist and Systems Architect with 20+ years of experience leading foundational AI research and large-scale intelligent systems development across industry and production environments. At Cisco AI Research and Intel Labs, led strategic research and engineering initiatives resulting in 100+ publications, 14 patents, and production AI systems spanning agentic AI, LLM reasoning, graph learning, and AI-driven HPC code generation, translation, and code optimization. Recognized for driving high-impact technical agendas and cross-functional innovation — from co-leading the DARPA HIVE program on intelligent programming systems to architecting enterprise-scale multi-agent AI frameworks for real-world deployment. Combines deep technical expertise, research vision, and cross-organizational leadership with a strong track record of translating advanced research into scalable platforms and measurable business impact. Active leader in the AI research community through conference and workshop leadership, editorial service, and international research initiatives, including organizing Dagstuhl seminars and premier venues such as SIGKDD, CIKM, Supercomputing, and IPDPS.

Technical Strategy, Research & Systems Impact

- **Agentic AI Systems:** Leads research and architecture development for reasoning, memory, planning, and coordination in large-scale multi-agent AI systems at Cisco AI Research and Outshift, including orchestration-layer design, inter-agent alignment and validation, observability, security, and enterprise workflow automation; filed 7 patents in 2025 spanning agent orchestration, reasoning supervision, and autonomous system safety.
- **AI for Programming Systems:** Led Intel Labs research on AI-driven HPC code generation, code-to-code translation, and compiler optimization, including co-leading the DARPA HIVE program on intelligent programming systems.
- **Large-Scale AI & Graph Learning Infrastructure:** Architected graph-based machine learning frameworks for knowledge graph reasoning, personalization, and automated program analysis across heterogeneous computing environments.
- **Scientific Leadership:** Authored 100+ publications in leading venues including NeurIPS, ICML, ICLR, ACL, EMNLP, WWW, and KDD, with contributions spanning agentic AI, graph learning, and AI systems.
- **Innovation & Intellectual Property:** Inventor on 14 patents across multi-agent systems, AI-driven software optimization, and autonomous AI orchestration platforms (4 granted, 10 pending).

Education

- 2008 – 2015  **Ph.D., Computer Science, Purdue University**
Dissertation: Scaling Up Network Analysis and Mining: Statistical Sampling, Estimation, and Pattern Discovery.
- 2012 – 2014  **M.S., Computer Science & Statistics, Purdue University**
- 2003 – 2006  **M.S., Computer Science, Cairo University**
- 2001  **B.S., Computer Science, Cairo University**
Thesis: Arabic Speech Recognition using Hidden Markov Models.

Experience

2024 – present	Senior Principal Engineer & Scientist, Cisco AI Research, Outshift, San Jose, CA. • Leading technical strategy, architecture, and applied research for large-scale multi-agent AI systems spanning reasoning, memory, planning, orchestration, inter-agent alignment, observability, security, and enterprise workflow automation. • Developed Agent-ToM, a learning-based monitoring framework grounded in Theory-of-Mind reasoning for detecting covert malicious behavior in autonomous LLM agents; introduced a Reason–Verify–Refine inference pipeline with persistent semantic guardrail memory, achieving state-of-the-art performance on SHADE-Arena while reducing inference cost to 2 LLM calls versus 8–17 for ensemble-based approaches. • Built evaluation and observability frameworks for agent reliability, drift detection, and reasoning correctness in LLM systems, establishing internal benchmarks adopted across Cisco AI teams. • Designed memory-augmented architectures for long-horizon reasoning, retrieval, and adaptive coordination in enterprise AI workflows. • Filed 7 patents in 2025 covering agent orchestration, theory-of-mind reasoning, and autonomous system safety.
2015 – 2024	Senior Staff Research Scientist & Technical Lead, Intel Labs, Santa Clara, CA. • Led strategic research initiatives in AI-driven code generation, compiler optimization, and program analysis using machine learning and reinforcement learning, with publications at NeurIPS, ICML, EMNLP, and related top-tier venues. • Developed domain-specific language models and code translation systems for HPC, including CodeRosetta (NeurIPS), AutoParLLM (NAACL), MonoCoder, OmpGPT, and MPIrigen, advancing state-of-the-art capabilities in parallel code generation, translation, and optimization. • Pioneered AI-driven compiler optimization with NeuroVectorizer (CGO 2020), among the first frameworks applying deep reinforcement learning to compiler vectorization decisions, contemporaneous with Google’s MLGO. • Developed PerfoGraph (NeurIPS), a graph-based program representation framework enabling automated parallelization and optimization in heterogeneous computing systems; open-sourced and adopted externally. • Led Intel’s AI-for-compilers research roadmap and co-led the DARPA HIVE proposal and research program on intelligent programming systems. • Established cross-organizational collaborations with academia and industry resulting in 50+ publications and multiple patents across a 9-year tenure.
2008 – 2015	PhD Candidate & Researcher, Purdue University, West Lafayette, IN.
2014	Research Intern, Meta, Menlo Park, CA.
2013	Research Intern, Intel Corp, Folsom, CA.
2012	Research Intern, Adobe Research, San Jose, CA.
2005 – 2008	Research Fellow, Data Mining & Computer Modeling Center of Excellence in Cairo, United Nations Development Program.

Selected Systems & Open Source

- **Agent-ToM** — First learning-based agent monitoring framework grounded in Theory-of-Mind reasoning; provides observability, monitoring, and security for autonomous LLM agents in enterprise environments; achieves state-of-the-art performance on SHADE-Arena. arXiv:2605.24216
- **NeuroVectorizer** — First deep reinforcement learning framework for end-to-end compiler vectorization; among the first RL-based compiler optimization systems, contemporaneous with Google's MLGO. CGO 2020. github.com/intel/neuro-vectorizer
- **PerfoGraph** — Language-agnostic LLVM-based graph representation framework for program synthesis and optimization; numerically aware, captures composite data structures, open-sourced and adopted externally. NeurIPS 2023. github.com/tehranixyz/perfograph
- **CodeRosetta** — Encoder-decoder transformer for unsupervised HPC code translation (C++/CUDA, C++/Fortran); outperforms closed-source LLMs by 22 BLEU points on C++ to CUDA translation and is the first model for Fortran to parallel C++ translation. NeurIPS 2024. coderosetta.com
- **PGD** — First parallel framework to extract and count all graphlet motifs at billion-edge scale, achieving exact counts in constant time per edge via combinatorial arguments; on average 460x faster than state-of-the-art methods across 300+ networks. Best Paper Runner-Up, ICDM 2015. graphlets.org
- **PGraphRAG** — Personalized graph-based retrieval-augmented generation framework and benchmark for personalized LLM reasoning and user-centric AI assistants. pgraphrag-benchmark.github.io
- **GLEMOS** — First comprehensive benchmark for instantaneous graph learning model selection; covers 366 models across 457 graphs for link prediction and node classification; companion benchmark to MetaGL. NeurIPS 2023 Datasets & Benchmarks Track. github.com/facebookresearch/glemos
- **Network Repository** — First and largest network data repository with interactive analytics and visualization; 500+ graph datasets serving as a hub for benchmarking graph ML and graph mining algorithms. AAAI 2015. **3,000+ citations**. networkrepository.com

Selected Publications

Author of **100+ publications** with **15,000+ citations**; complete publication list available on Google Scholar.

- 1 **N. K. Ahmed** and N. Nafisi, *Agent-ToM: Learning to monitor autonomous LLM agents via Theory-of-Mind reasoning*, arXiv preprint arXiv:2605.24216, 2026.
- 2 H. Ping, A. Bhattacharjee, P. Zhang, S. Li, W. Yang, A. Cheng, X. Zhang, J. Thomason, A. Jannesari, **N. K. Ahmed**, et al., "VeriMoA: A mixture-of-agents framework for Spec-to-HDL generation," in *Proceedings of Machine Learning and Systems (MLSys)*, 2026.
- 3 S. Au, C. J. Dimacali, O. Pedirappagari, N. Park, F. Dernoncourt, Y. Wang, N. Kanakaris, H. Deilamsalehy, R. A. Rossi, and **N. K. Ahmed**, "Personalized graph-based retrieval for large language models," in *Proceedings of the 39th Pacific Asia Conference on Language, Information and Computation*, 2025.
- 4 Y. Li et al., "AutoSDT: Scaling data-driven discovery tasks toward open co-scientists," in *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2025.
- 5 Q. I. Mahmud, A. TehraniJamsaz, H. D. Phan, **N. K. Ahmed**, and A. Jannesari, "AutoParLLM: GNN-guided automatic code parallelization using large language models," in *2025 Annual Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics (NAACL)*, 2025.
 URL: <https://arxiv.org/pdf/2310.04047>.

- 6 Y. Wang, R. A. Rossi, N. Park, H. Chen, **N. K. Ahmed**, P. Trivedi, F. Deroncourt, D. Koutra, and T. Derr, "Large generative graph models," in *International Conference on Learning Representations (ICLR)*, 2025.  URL: <https://arxiv.org/abs/2406.05109>.
- 7 K. Cheng, **N. K. Ahmed**, T. L. Willke, and Y. Sun, "Structure-guided prompt: Instructing large language model in multi-step reasoning by exploring graph structure of the text," in *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2024.
- 8 I. O. Gallegos, R. A. Rossi, J. Barrow, M. M. Tanjim, S. Kim, F. Deroncourt, T. Yu, R. Zhang, and **N. K. Ahmed**, "Bias and fairness in large language models: A survey," *Computational Linguistics*, vol. 50, no. 3, pp. 1097–1179, Sep. 2024.  DOI: 10.1162/coli_a_00524.
- 9 T. Kadosh, N. Hasabnis, V. A. Vo, N. Schneider, N. Krien, M. Capota, A. Wasay, **N. K. Ahmed**, T. Willke, G. Tamir, et al., "MonoCoder: Domain-specific code language model for HPC codes and tasks," in *IEEE High Performance Extreme Computing Conference (HPEC)*, 2024.
- 10 A. TehraniJamsaz, A. Bhattacharjee, L. Chen, **N. K. Ahmed**, A. Yazdanbakhsh, and A. Jannesari, "CodeRosetta: Pushing the boundaries of unsupervised code translation for parallel programming," in *The Thirty-eighth Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2024.
- 11 K. Cheng, **N. K. Ahmed**, and Y. Sun, "Neural compositional rule learning for knowledge graph reasoning," in *International Conference on Learning Representations (ICLR)*, May 2023.
- 12 N. Park, R. A. Rossi, **N. K. Ahmed**, and C. Faloutsos, "META-GL: Evaluation-free selection of graph learning models via meta-learning," in *International Conference on Learning Representations (ICLR)*, May 2023.
- 13 A. TehraniJamsaz, Q. I. Mahmud, L. Chen, **N. K. Ahmed**, and A. Jannesari, "PerfoGraph: A numerical aware program graph representation for performance optimization and program analysis," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 36, Dec. 2023.
- 14 A. Haj-Ali, **N. K. Ahmed**, T. Willke, S. Shao, K. Asanovic, and I. Stoica, "NeuroVectorizer: End-to-end vectorization with deep reinforcement learning," in *CGO 2020: Proc. of the International Symposium on Code Generation and Optimization*, Feb. 2020.
- 15 R. A. Rossi and **N. K. Ahmed**, "Role discovery in networks," *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, vol. 27, no. 4, pp. 1112–1131, 2015, ISSN: 1041-4347.  DOI: 10.1109/TKDE.2014.2349913.

Selected Patents

- 1 **N. K. Ahmed**, "Federated Agent Discovery for Decentralized Heterogeneous Registries," Cisco Systems, 2025, Filed.
- 2 **N. K. Ahmed** and P. Bosch, "Modular Theory-of-Mind Reasoning for Multi-Agent AI Systems," 19/356842, Cisco Systems, 2025, Filed.
- 3 **N. K. Ahmed**, S. Pant, and M. Viswanathan, "Autonomous Dynamic Optimization in Multi-Agent Systems," Cisco Systems, 2025, Filed.
- 4 **N. K. Ahmed**, R. Pishehvar, B. Riordan, S. Pant, and M. Viswanathan, "Agent Selection for Task Execution in a Multi-Agent System," Cisco Systems, 2025, Filed.
- 5 P. Bosch, **N. K. Ahmed**, and A. Surcouf, "Systems and Methods for Establishing Security in Agentic Applications," 19/326224, Cisco Systems, 2025, Filed.
- 6 P. Bosch, R. Kompella, **N. K. Ahmed**, A. Duminuco, M. Yannuzzi, and A. Surcouf, "Improving Security of Agentic Applications Using Theory-of-Mind Reasoning," 19/425429, Cisco Systems, 2025, Filed.

- 7 M. Capota, G. MA, S. Zhou, N. Hasabnis, and **N. K. Ahmed**, "Execution Placement of a Computer Program Using a Trained Machine-Learning Model," US12307223B2, Intel Corp, May 2025, Granted.
- 8 J. Tang, C. Wang, K. Ponnambalam, and **N. K. Ahmed**, "Metric Selection to Evaluate Multi-Agent Systems," Cisco Systems, 2025, Filed.

Selected Honors and Awards

- 2025 Best Paper Award, COLM'25 LLM for Scientific Discovery Workshop — AutoSDT.
- 2024 Outstanding Paper Award, IEEE High Performance Extreme Computing Conference (HPEC 2024).
- 2022 Intel Labs Recognition Award for Outstanding Contributions.
- 2020 Intel Labs Recognition Award for Outstanding Contributions; Intel Labs Recognition Inventor Award.
- 2019 Intel Labs Recognition Award for Outstanding Contributions.
- 2017 Finalist, IEEE/Amazon/DARPA Graph Challenge for High-Performance Graph Analytics, HPEC.
- 2015 Best Paper Candidate, IEEE International Conference on Data Mining (ICDM).
- 2014 Rising Stars Award, University of California Berkeley — Awarded to top female PhD candidates and postdoctoral researchers in EECS.
- 2012 Best Paper Runner-Up Award, SIGKDD Workshop on Big Data, Streams and Source Mining.
- 2005–2008 Research Fellowship, United Nations Development Program (UNDP).

Research Leadership & Professional Service

Conference & Workshop Organization

- 2027 Seminar Organizer — Dagstuhl Seminar on Intelligent Design Automation for the Open Silicon Era.
- 2025 Workshop Co-Chair and Organizer — Machine Learning on Graphs in the Era of Generative AI (SIGKDD).
- 2023 Seminar Organizer — Dagstuhl Seminar on Scalable Graph Mining and Learning.
PhD Consortium Co-Chair — ACM SIGKDD Conference on Knowledge Discovery and Data Mining.
- 2022 Demo Track Co-Chair — ACM International Conference on Information and Knowledge Management (CIKM).
- 2021–2026 Workshop Co-Chair and Organizer — Graphs, Architectures, Programming, and Learning (GrAPL), IEEE IPDPS.
- 2018 Program Committee Chair — IEEE Big Data Conference, Industry and Government Track.

Journal Editorial Board

- 2020–2021 Associate Editor, IEEE Transactions on Neural Networks and Learning Systems (TNNLS).
- 2019–2020 Associate Editor, IEEE Signal Processing Magazine.
- 2018–present Member, Frontiers in Machine Learning and Artificial Intelligence.

Conference Committee

- 2015–present Regular SPC/TPC member for leading AI and ML conferences, including NeurIPS, ICML, ICLR, ACL, EMNLP, SIGKDD, AAI, and WWW.