Nicholas Stephen Kersting, Ph.D.

Personal

E-mail: 1054h34@gmail.com
LinkedIn: linkedin.com/in/nicholas-kersting-b9a35330/
Website: quantumrepoire.com/nk
GitHub: github.com/nkersting

Education

University of California, Berkeley, CA **1997.8 - 2002.5**
Ph.D. in Particle Physics

Carnegie Mellon University, Pittsburgh, PA **1993.9 - 1997.6**
B.S. Physics (Magna Cum Laude)

Amherst Central High School, Amherst, NY **1989.9 - 1993.6**

Employment History

Iron Action AI, LLC , Austin, TX **2022.8 - present**
(CEO) **ironactionai.com**

- Safe, Robust, and Trustworthy AI
  - Realtime Machine Learning Model Robustness
    → **github.com/nkersting/harmonic_robustness/tree/main/harmonic_tester**
  - LLM Hallucination Detection
    → **github.com/nkersting/harmonic_robustness/tree/main/gamma_tester**
  - LLM grounded-ness / logic verification
  - Robust robotic motion
- Optimized Human/AI communication
  - Total privacy-preserving interest matching
    → prototype at **quantumrepoire.com**
- AI/ML Advisory
  - Paket (www.paket.tv)
  - Storyboard Capital

Visa, Inc., Austin, TX **2021.11 - 2025.1**
(Principal Machine Learning Engineer)

- Generative AI Projects
  - ∗ Client-facing GenAI external products
    - → Paris 2024 Olympics chatbot (within "Visa Go" IOS/Android app): GPT-4-based, leveraging own anti-hallucination metric (arxiv.org/abs/2404.19708)
    - → GenAI Quality Service: in-house and 3rd party metrics to gauge reliability/safety for RAG and LLM-coding applications
  - ∗ ChatGPT/GPT-4-embedded internal products
    - → VisaSecureGPT (secure wrapper around OpenAI services; rolled out to 25k employees; use cases: summary, search, ideation, coding, ...)
    - → VisaGPT (user-defined RAG application; use cases: summary, search training, internal documents)
    - → VisaTableGPT (Smart Spreadsheet analysis; agentic RAG and feature generation; use cases: client survey analysis, transaction summary)
    - → VisaMarketGPT (RAG-based marketing tool; use cases: sales presentation mock-ups, ad design, market analysis)
    - → VisaFormGPT (Agentic form-filling tool cross-referencing historical records; use cases: legal/compliance forms, security assessment forms, ...)
  - ∗ Local Deployment of Open Source LLMs
    - → Stable Diffusion (PoC for testing potential for Marketing)
    - → Mosaic MPT-7B (PoC for testing potential for QA system)
  - ∗ Payment Foundation Model: On-prem training of in-house LLM for fraud/risk applications; Exploration in new transformer architectures, multitask learning, NLP techniques
  - ∗ Documentation/Pedagogical material for C-level stakeholders
- Onboarding/Maintenance of AI Models for Credit Card Risk
  - ∗ Main Architect for flagship models processing $\sim$ 100M tx ($\sim$ \$10B) / day
  - ∗ Authoring architectural documents, coordination with Research and Business teams; Overseeing production upgrades, monitoring issues and response
- Visa University Faculty Member
  - ∗ Teaching course on Value Added Services; Assisting with labs and quantitative analysis of data sets
  - ∗ Delivering Tech Talks on current topics (LLMs, Responsible AI, Quantum Computing...)
- Intellectual Property Development
  - → 7 Filed Patents (see below)
  - → 5 Technical Innovations: Intelligent Form-Filling Assistant GPT-powered

spreadsheet analyzer, Automated Bias Detection in LLM, Self-Explainable AI Models, Variety-based Data Explorer

→ 1 Defensive Publication: Secure Search with Embeddings

- Headed Responsible AI Initiative: Fundamental Research on SecureML and Robust Production Models, Federated Learning and Privacy-Preserving ML, Explainable AI (XAI); Led Team for Gold Medal in 2022 Visa Global Hackathon
- Headed Batch Platform Initiative: Architecture of generic batch computing capabilities of AI platform; Oversaw scrum team and high-level reporting
- Intern Guidance: Lead Team of interns on E2E LLM QA project; Guided interns on Responsible AI project (LLM Hallucination), AI Agents
- New Hire Onboarding Initiative: Completely revised 50+ new hires' onboarding procedure; Organized scheduled seminars, projects, documenting.

Neurothink / Lanai, Austin, TX **2021.10 - 2024.2**

(Head of AI Research) **neurothink.io, lanailabs.ai**

- Research in Fundamental Neural Net Architectures: Variety-Driven (Max-Variety) models, XAI models, Human-AI-generated content disambiguation
- Platform for "Radically Accessible ML": Performance Testing with A100s, User Testing with University/Government cohorts, Basic Model Testing (MLP, Transformer, CNN, AE, VAE, ...), Pedagogical presentations to clients
- Deep Fake Detection
  - ∗ Devised XAI algorithm to distinguish AI from Human-written text
    - → Commercial Chrome Plugin (see **lanai.ai**) to test probability of webpage text being AI-generated; Identifies ChatGPT-generated content with F1=82%
  - ∗ Physics-based Deep Fake detection research: AI text scoring, AI-generated image(e.g., face) detection, AI-generated copyright violating music detection
    - → prototype at **quantumrepoire.com/variety.html**
- AI Trustworthiness & Robustness: Geometric methods for production robustness monitoring, AI trustworthiness metrics, XAI metrics, Adversarial vulnerability metrics

Ant Financial, NY, NY and San Mateo, CA **2018.1 - 2021.10**

(Principal for Intelligent Products & Solutions, Staff Engineer)

- Full stack conversational chatbots
  - ∗ Design and implementation of financial chatbot for Ant Fortune app
    - → Back-end FAQ component of full product in team of 70+ engineers; 0-1 coding, testing, implementation, and maintenance
    - → Attained 90% precision, 90% recall with $< 30ms$ latency per query;

99.999% resilience for high volume (1M requests/day) chatbot in Alipay app
  - Platform for chatbot creation: Design and launch customized FAQ chatbot in minutes; Back-end advanced NLP Models (WAM, BERT, Transformer); Real time visual monitoring, diagnostics & reporting; Full testing suite: smoke, regression, stress, periodic, ...
  - Research in Multi-task Learning for Dialog Response Generation
- Engineering Team Product Coordination
  - Organizing and customizing China-developed products with 500+ person AI engineering team: Financial AI modeling platform, Automated PDF document entity extraction, Image-recognition-based insurance claims, Sentiment analysis platform
  - Remote and on-site product demonstration to financial customers
  - Authoring and vetting international product documentation
- International Customer Outreach and Business Development
- → Outreach presentations to very large (AUM > $\$100B$) high profile financial customers (e.g. World Bank)
- → Business representation at international conferences (e.g. CB Insights Fintech)
- → Liaison between Chinese AI specialists and U.S. financial customers

IBM Research, Yorktown Heights, NY **2015.1 - 2018.1**
(Senior Software Engineer)

- Watson Cognitive Computing / Artificial Intelligence
  - Researching Fundamental NLP
  - Focus Topics: Semantic Word Embeddings, Information Retrieval, Dialogue Question/Answering, Natural Language Generation and Understanding
- Statistical and Neural Machine Translation (SMT, NMT)
  - Language Modeling and Data Processing for > 100 World Languages (personally added Somali, Turkish, Dutch, Polish, Traditional Chinese to Watson Languages); Data Curation, Maintenance, Distribution (billions of words per language)
  - Algorithmic Retrieval & Quality Filtration of Web and Corpus Data (millions of words per language per day)
  - → Collaboration with ∼ 20 Language Specialists to improve NMT quality, personally specializing in Russian and Chinese, improved BLEU score to beat GoogleTranslate in Italian, Korean, Japanese
  - → Managing team ∼ 10 Language Annotators
- Autonomous Web Crawling
- → Code/maintain architecture for autonomous web crawler (∼10M URLs/day)
  - Sys Admin, distributed database storage and retrieval (∼ 100 TB)
  - Code and maintain downstream NLP pipe: text extraction, filtration, anal-

ysis & storage

- Intellectual Property Development

→ 5 patents filed (see below)

Princeton Consultants, Inc., Princeton, NJ **2010.8 - 2015.1**

(Consultant)

- Network Planning Software for UPS: adapting legacy COBOL logic to modern C++; Client consultation and documentation

→ High-performance development: improved global network traversal from 5 min to $5\mu s$, update response time in seconds

- Railroad Signal/Switch Testing Software

→ Led team of 5 software engineers with Agile life-cycle

→ Developed GUI for signal engineers allowing 10-hour jobs to run in minutes: (**princeton.com/sats**); Marketed to several major railways: BNSF, CSX, NS, UP

- Social Networking Server

→ Innovated privacy-preserving set comparison algorithm (see arxiv.org/abs/1308.3294)

→ Interactive prototype at **quantumrepoire.com**

Sichuan University, Chengdu, P.R.C. **2006.6 - 2010.8**

(Professor of Physics)

- Teaching (4 courses/semester) and advising (20 graduate students)
- Experiment administration/planning (China Dark Matter Experiment (CDEX))
- Research in particle physics
  - Pioneered statistical techniques to find new particles with 100-fold accuracy
- Statistical analysis of TB-size data
  - Experimental simulation and analysis: generated 10 years' LHC data
  - Monte Carlo parallel computing

Tsinghua and Sichuan Universities, P.R.C. **2002.9 - 2006.6**

(Post-Doctoral Researcher)

- Investigated new particle physics models; Generation and analysis of large datasets; Managed student research groups

LBNL, Berkeley, CA **1999.5 - 2002.5**

(Lawrence Berkeley National Laboratory Research Associate)

- Research in theoretical physics

University of California, Berkeley, CA **1998.1 - 2002.5**

(Graduate Student Instructor)

- Teaching discussion sections, reviews >100 students/class

Internships

Indiana University Cyclotron Facility (1996), Fermilab (1995) DAQ software

Patents

1. "Secure social connection via real-time biometrics and cognitive state comparison" (US10922365B2)
2. "System, method, and recording medium for communication and message comparison with encrypted light signals" (US10084758B2)
3. "Adaptive selection of message data properties for improving communication throughput and reliability" (US10305765B2)
4. "Systems and methods for generating document variants" (US11080341B2)
5. "Single Channel Multiple Access Communications System" (US11646894B2)
6. "Method, system, and computer program product for multi-layer analysis and detection of vulnerability of machine learning models to adversarial attacks" (WO2024220790A1)
7. "Performance determination of machine learning models based on decision boundary geometry" (WO2024173653A1)

   ... 5 additional patents pending

Public Speaking

(invited to give talks at the following)

Institute for Artificial Intelligence and Fundamental Interactions *(MIT)*, Institute for the Physics and Mathematics of the Universe *(Kashiwa, Japan)*, Kavli Institute for Theoretical Physics *(Beijing)*, International Conference of High Energy Physics*(Beijing, Moscow, Philadelphia)*, Euler Institute *(St. Petersburg, Russia)* (in Russian), Chinese Physical Society*(Taiyuan U., Wuhan U., Tsinghua U., Nanjing U.)* (in Chinese), INFORMS NY *(New York City)*

Skills

Technical Skills

- Software Engineering
  - Full stack product lifecycle (inception to delivery)

    * Languages: Python, C++, Rust, Go, Java, C#, Javascript, Perl, Fortran, Bash; GitHub CoPilot
    * IDEs: VS Code, Cursor, emacs, vim
    * Source Control + Code Review: GIT, SVN, Crucible, Bazel
    * Continuous Integration, Testing, and Diagnostics: Jenkins, pytest, flake8, Valgrind, GDB
    * Databases + Containerization: DB2, MySQL, Redis, Chroma, Pinecone, FAISS, Hadoop, Cassandra, Docker, Kubernetes
    * Web server + Multi-tier architecture: Apache HTTP, Node.js, PHP, CGI scripting
    * Security, Cryptography, Blockchain (Bitcoin and Cryptocurrency @ Princeton U. (Coursera))
- Machine Learning
    * Generative AI: Open-source LLM (Mistral, Llama, ...), Diffusion-based Image Generation, LLM Prompt Engineering, Vector DB, LangChain, Retrieval Augmented Generation, Agent-based systems, Hallucination detection; Azure AI Cognitive Services, Amazon Bedrock Framework
    * Techniques: Neural Network first-priciples design and implemenation (MLP, DNN, CNN, AE, RNN, LSTM, Transformer), Support Vector Machines, Gradient-boosted Decision Trees, Logistic Regression, Clustering (Kmeans, Gaussian, Hierarchical, DBSCAN), Reinforcement Learning, RLHF
    * Methods: Supervised and Unsupervised Learning, Transfer & Few Shot Learning, Data Augmentation
    * Optimization: SGD, Adam, simulated annealing, genetic algorithms, simplex
    * Tools: PyTorch, scikit-learn, TensorFlow, Trax, Octave, R
    * Machine Learning Certification @ Stanford University (Coursera)
- Natural Language Processing
    * Preprocessing: encodings (UTF8, BPE, Wordpiece), segmentation, tokenization, lemmatization
    * Structural: Grammar parse trees, POS tagging, morphological analysis
    * Sentiment Analysis, Text Summarization, Paraphrasing, Machine Translation
    * Advanced Modeling: BERT, ONMT, GPT, T5, LLMs
    * Named Entity Recognition and Classing
    * Chatbot design and implementation
    * Tools: sklearn, Natural Language ToolKit, WordNet, gensim, TextBlob, HuggingFace
    * Text Retrieval and Search Engines Certification @ U. Illinois Urbana-Champaign (Coursera)
    * NLP Specialization Certification (Probabilistic Models, Classification with Vector Spaces, Sequence Models, Attention Models) @ Stanford University (Coursera)
- Data Analysis
    * Techniques: Bayesian statistics, confidence, regression, multivariate anal-

ysis, graphs/networks, physical modeling, feature extraction
  - Visualization: matplotlib, jupyter, JavaScript D3.js, Tableau, WPF, HTML
  - Tools: NumPy, Pandas, R, Octave, Mathematica, Matlab
  - Mining Massive Datasets Certification @ Stanford University (Coursera)
- Communication
  - Documentation: LaTeX, Wiki Markup, AI-based (Perplexity, OpenAI Deep-Research, NotebookML, Gamma, ...)
  - Project Management: JIRA, Confluence, MS Office, Copilot
  - Presentation: Prezi, ppt, Lovable POC, Zoom, webinars, podcasts

Language Skills

Fluent in written and spoken:

- Chinese (Mandarin, both traditional/simplified, near-native level)
- Russian (professional level)
- Ukrainian (conversational level)
- English (native)

Other Interests

- Kungfu (certified Black Belt from Chengdu Sports University)
- Chinese Calligraphy (certified in Li style from Sichuan University)
- Go (certified 2-dan from Chengdu Weiqi Society; former East Bay Chinese School instructor)
- Acting & Modeling (played major role in TV series "King of San-Da" 2005)
- Western Fine Arts (pen/ink drawing, intaglio printmaking, oil painting)
- Scouting America (Assistant Scout Master for Troop 52, Lakeway, TX)

Selected Publications

1. "Harmonic LLMs are Trustworthy" (arxiv.org/abs/2404.19708) Published in XAI-IJCAI (sites.google.com/view/xai2024/proceedings)
2. "ONNXExplainer: an ONNX Based Generic Framework to Explain Neural Networks Using Shapley Values" (arxiv.org/abs/2309.16916) Published in XAI-IJCAI (sites.google.com/view/xai2024/proceedings)
3. "Harmonic Machine Learning Models are Robust" (arxiv.org/abs/2404.18825)
4. "Evaluating Quality of Answers for Retrieval-Augmented Generation: A Strong LLM Is All You Need" (arxiv.org/abs/2406.18064)
5. "Identifying AI content with Variety" (medium.com/@1054h34/identifying-ai-content-with-variety-128b4c728724)
6. "Identifying ChatGPT with Variety" (medium.com/@1054h34/identifying-chatgpt-with-variety-ce392b09e38e)

7. "A Secure and Comparable Text Encryption Algorithm" (arxiv.org/abs/1308.3294)

(additional professional peer-reviewed articles, available upon request, or on arXiv)