

JON A. CHUN

Co-Creator, Human-Centered AI Curriculum | Founding Co-Director, AI CoLab
Kenyon College
jonchun@outlook.com

AI research scientist specializing in LLM security evaluation, multi-agent systems, and applied ML. Hands-on implementation from GPT-2 (2019) through contemporary LLMs. PI for Schmidt Sciences; PI representing the 25,000-member MLA at the US AI Safety Institute (now NIST CAISI). Published in ICML, ACL EMNLP, Frontiers in Computer Science, and Narrative, with 10 papers under review at top-tier 2026 venues. Created SentimentArcs, an open-source methodology adopted globally. Prior industry: co-founded SafeWeb (\$26M acquisition by Symantec; first In-Q-Tel security investment); two US patents. B.S. EECS UC Berkeley, M.S. EECS UT Austin (Highest Honors), AHA Research Fellow. Fluent in English, Japanese, and Spanish.

EDUCATION

B.S., Electrical Engineering & Computer Sciences (EECS), UC Berkeley (Honors; Tau Beta Pi, Eta Kappa Nu)
M.S., EECS / Biomedical Engineering, UT Austin (Highest Honors). Focus: Cognitive Science
AHA Research Fellow, University of Iowa College of Medicine. Published research on gene therapy and EHR systems
Advanced Technical Japanese, MIT Japan Program (NSF / US Navy Scholarship)

TECHNICAL EXPERTISE

Full-stack AI development from GPT-2 through contemporary LLMs (GPT-4, Claude, Llama, Mistral). Python, TensorFlow, PyTorch, scikit-learn, HuggingFace Transformers. LLM red/blue team testing, threat modeling, vulnerability detection, XAI, uncertainty quantification. Created SentimentArcs (open-source, adopted globally). Multi-agent architecture design, game theory simulations, automated agent frameworks, multimodal AI. Earlier: C/Assembly/UNIX (Lawrence Berkeley Labs); enterprise security architecture; two US patents.

SELECTED PUBLICATIONS

Computational Methods & Multi-Agent Systems

“AgenticSimLaw: A Juvenile Courtroom Multi-Agent Debate Simulation for Explainable High-Stakes Tabular Decision Making” (with K. Elkins and Y. S. Lee). Accepted, *LaMAS* 2026.

“Syntactic Framing Fragility: An Audit of Robustness in LLM Ethical Decisions” (with K. Elkins). *arXiv*, December 2025.

“AIStorySimilarity: Quantifying Story Similarity Using Narrative.” *ACL EMNLP/CoNLL*, Miami, 2024.

“MultiSentimentArcs: Affective AI and Multimodal Diachronic Sentiment Analysis.” *Frontiers in Computer Science*, 2024.

“eXplainable AI with GPT4 for Story Analysis and Generation” (with K. Elkins). *International Journal of Digital Humanities*, Vol. 5, Issue 2–3, 2023.

Governance, Security & AI Policy

“Risks and Opportunities of Open-Source Generative AI.” *Proceedings of ICML* 2024. Oral presentation (top 2%).

“If Open Source is to Win it Must Go Public.” Spotlight talk, CODEML, *ICML* 2025.

“Informed AI Regulation: Comparing the Ethical Frameworks of Leading LLM Chatbots Using an Ethics-Based Audit” (with K. Elkins). *arXiv*, 2024.

“Comparative Global AI Regulation: EU, China, and the US” (with C. Schroeder de Witt & K. Elkins). 2024.

Humanities & AI

“Can GPT-3 Pass a Writer’s Turing Test?” (with K. Elkins). *Journal of Cultural Analytics* 5(2), 2020.

“What the Rise of AI Means for Narrative Studies” (with K. Elkins). *Narrative* 30, no. 1, 2022.

“The Crisis of Artificial Intelligence: A New DH Curriculum for Human-Centered AI.” *Int. J. Humanity and Arts Computing* 17, 2023.
10 additional papers under review at ICML, FAccT, UAI, CogSci, TMLR, DMLR, and Journal of Cyber Policy (2026).

PATENTS

“Intelligent Secure Data Manipulation Apparatus and Method.” US Patent 7,730,528 B2 (2010). Symantec Corp.

“Method and Apparatus for Encrypted Communications to a Secure Server.” US Patent 8,065,520 (2011). Symantec Corp.

APPLIED TECHNOLOGY & INDUSTRY

SafeWeb, Inc. — Co-Founder, President & CEO (2000–2003). Built world’s largest online privacy service (2B transactions in 8 months). \$11M VC; first security investment from In-Q-Tel (CIA venture fund). Sold to Symantec for \$26M at 35× trailing revenue.

Symantec Corp. — Director, Development (2003–2004). Post-acquisition integration; launched Clientless VPN Gateway in one quarter, under budget. Zero attrition.

BWG Strategy LLC — AI Expert & Industry Analyst (2023–present). Featured speaker on OpenAI/Microsoft, open-source regulation, Google Gemini, IBM Watson.

Earlier: UC Berkeley COE, Entrepreneur in Residence (2005–2007) · Latin eVentures, CTO (\$8M raised) · Matrix Absence Mgt (NYSE: DFG), Director of IT · Lawrence Berkeley Labs, Research Programmer · SEMATECH, Japanese Semiconductor Analyst · MMS Int’l / McGraw-Hill Tokyo

GRANTS & AWARDS

Schmidt Sciences HAVI Award (2025) — One of 23 worldwide. PI & AI Architect, “Archival Intelligence: Rescuing New Orleans Heritage.”

Notre Dame–IBM Tech Ethics Lab Grant (2024) — Competitive international award. Co-PI, “How Well Can GenAI Predict Human Behavior?”

In-Q-Tel (CIA Venture Fund) — First security investment. SafeWeb, Inc.

POLICY, GOVERNANCE & ADVISORY ROLES

PI, US AI Safety Institute (now NIST CAISI) — Representing the 25,000-member MLA; LLM red/blue team testing

AI Alliance Essential AI Competencies Guide (2024) — Contributor (Meta, IBM, Intel, US Chamber of Commerce)

Executive Board Member, The Helix Center (2022–present)

TEACHING & RESEARCH IMPACT

Co-created the first human-centered AI curriculum (2016). 400+ mentored student research projects downloaded 95,000+ times from 4,000+ institutions worldwide. Flagship course became one of the largest on campus despite counting toward no major. DEI: 61% women, 13% Black, 11% Latine, >90% non-STEM, 0% drop rate. Courses: Programming Humanity, AI for the Humanities, Cultural Analytics, Frontiers in Generative AI: Autonomous Agent Networks.

LANGUAGES

Japanese, Spanish (advanced; JLPT, US Foreign Service Exam) | French (professional working) | Chinese (conversational)