CURRICULUM VITAE
Peter Alexander Jansen, PhD
Associate Professor, College of Information, University of Arizona.
pajansen@arizona.edu
www.cognitiveai.org

Research Areas

Recent: Automated scientific discovery, automated coding agents, virtual/simulated/text-based environments for agents, agent architectures for long-horizon multi-step problems. *Less-recent:* Explanation-centered inference, automated reasoning
Common contributions: I am an AI/Natural Language Processing researcher. I self-identify as a research methods researcher, and frequently build new instrumentation, methods, representations, resources, and tools that enable doing new kinds of research, or asking new or finer-grained research questions. I also do a substantial amount of modeling and experimentation contributions, increasingly in more applied domains (like scientific discovery).

Chronology of Education

2010 **Ph.D., Psychology & Neuroscience, McMaster University, Canada**
Dissertation: A self-organizing computational neural network architecture with applications to sensorimotor grounded linguistic grammar acquisition.
Major Fields: Cognitive Modelling, Knowledge Representation

2005 **B.I.S., University of Waterloo, Canada**
Major Fields: Computer Science, Cognitive Science, Physics

Chronology of Employment

2023- **Visiting Research Scientist**, Allen Institute for Artificial Intelligence (Ai2)
Sabbatical (2023-2024), Continuing (2024-)
2023- **Associate Professor**, School of Information, University of Arizona
Courtesy Appointment in Department of Computer Science, and Department of Linguistics.
2016 **Assistant Professor**, School of Information, University of Arizona
Courtesy Appointment in Department of Linguistics.
2015 **Research Professor**, Department of Linguistics, University of Arizona
2013 **Postdoctoral Research Associate**, School of Information, University of Arizona
2012 **Senior Artificial Intelligence Engineer**, Scanadu Inc, NASA Ames Research Campus
2010 **Postdoctoral Research Associate**, Electrical Engineering, University of Arizona

Publications/Creative Activity

In my primary fields of artificial intelligence and computational linguistics, conference publications are generally ranked higher than journal articles. These are full papers that go through the normal peer review process, as in a journal.

Work in Progress

51. **Jansen., P.**, Clark, P., Downey, D., Weld, D. Generating Literature-Driven Scientific Theories at Scale. *Arxiv Preprint.*
50. **Jansen., P.**, Hassan, S., Narasimha, P. CodeDistiller: Automatically Generating Code Libraries for Scientific Coding Agents. *Arxiv Preprint.*
49. Vasu, R., **Jansen., P.**, Siangliulue, P., Sarasua, C., Bernstein, A., Clark, P., Dalvi, B. HARPA: A Testability-Driven Literature-Grounded Framework for Research Ideation. *Arxiv Preprint.*

Refereed conference and workshop articles, published or accepted in final form

48. Bragg., J., D'Arcy, M., et al. (2026). AstaBench: Rigorous Benchmarking of AI Agents with a Scientific Research Suite. (2026). *Proceedings of the International Conference on Learning Representations (ICLR).*
47. **Jansen., P.**, Tafjord, O., Radensky, M., Siangliulue, P., Hope, T., Dalvi, B., Majumder, B., Weld, D., Clark, P. (2025). CodeScientist: End-to-end Semi-Automated Scientific Discovery with Code-based Experimentation. In *Findings of the Association for Computational Linguistics (ACL Findings).*
46. **Jansen, P.**, Hassan, S., Wang, R. (2025). Matter-of-Fact: A Benchmark for Verifying the Feasibility of Literature-Supoprted Claims in Materials Science. In *Proceedings of Empirical Methods in Natural Language Processing (EMNLP).*
45. Weir, N., Dalvi, B., Weller, O., Tafjord, O. Hornstein, S. Sabol, A., **Jansen, P.**, Van Durme, B., Clark, P. (2025). From Models to Microtheories: Distilling a Model's Topical Knowledge for Grounded Question Answering. In *Proceedings of the International Conference on Learning Representations (ICLR).*
44. **Jansen., P.**, Côté, M.A., Khot, T., Bransom, E., Dalvi, B., Majumder, B., Tafjord, O., Clark, P. (2024). DISCOVERYWORLD: A Virtual Environment for Developing and Evaluating Automated Scientific Discovery Agents. In *Proceedings of the Advances in Neural Information Processing Systems 37 (NeurIPS) Datasets & Benchmark Track (Spotlight).*
43. Wang, R., Todd, G., Xiao, Z., Yuan, E., Côté, M.A., and **Jansen, P.** (2024). Can Language Models Serve as Text-Based World Simulators? In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL).*
42. Zhang, L., **Jansen, P.**, Zhang, T., Clark, P., Callison-Burch, C., Tandon, N. (2024). PDDLEGO: Iterative Planning in Textual Environments. In *Proceedings of the 13th Joint Conference on Lexical and Computational Semantics (*SEM).*
41. Weir, N., Sanders, K., Weller, O., Sharma, S., Jiang, D., Jiang, Z., Mishra, B., Tafjord, O., **Jansen, P.**, Clark, P., and Van Durme, B. (2024). Enhancing systematic decompositional natural language inference using informal logic. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP).*
40. Majumder, B., Dalvi, B., **Jansen, P.**, Tafjord, O., Tandon, N., Zhang, L., Callison-Burch, C., and Clark, P. (2024). CLIN: A Continually Learning Language Agent for Rapid Task

Adaptation and Generalization. In *Proceedings of the Conference on Language Modeling (COLM).*

39. Wang, R., Todd, G., Yuan, E., Xiao, Z., Côté, M.A., and **Jansen, P.** (2023). ByteSized32: A Corpus and Challenge Task for Generating Task-Specific World Models Expressed as Text Games. In *Proceedings of Empirical Methods in Natural Language Processing (EMNLP).*
38. **Jansen, P.** (2023). From Words to Wires: Generating Functioning Electronic Devices from Natural Language Descriptions. In *Findings of Empirical Methods in Natural Language Processing (EMNLP Findings).*
37. Wang, R., and **Jansen, P.** (2023). Self-Supervised Behavior Cloned Transformers are Path Crawlers for Text Games. In *Findings of Empirical Methods in Natural Language Processing (EMNLP Findings).*
36. **Jansen, P.**, Cote, M.A. (2023). TextWorldExpress: Simulating Text Games at One Million Steps Per Second. In *Proceedings of the European Chapter of the Association for Computational Linguistics (EACL) System Demonstrations.*
35. Wang, R., **Jansen, P.**, Cote, M.A., Ammanabrolu, P. (2023). Behavior Cloned Transformers are Neurosymbolic Reasoners. In *Proceedings of the European Chapter of the Association for Computational Linguistics (EACL).*
34. Wang, R., **Jansen, P.**, Cote, M.A., Ammanabrolu, P. (2022). ScienceWorld: Is your Agent Smarter than a 5th grader. In *Proceedings of Empirical Methods in Natural Language Processing (EMNLP).*
33. **Jansen, P.** (2022) A Systematic Survey of Text Worlds as Embodied Natural Language Environments. In *Proceedings of the* 3^{rd} *WordPlay Workshop (Wordplay).*
32. **Jansen, P.**, Boyd-Graber, J. (2022). Picard understanding Darmok: A Dataset and Model for Metaphor-Rich Translation in a Constructed Language. In *Proceedings of the* 3^{rd} *Workshop on Processing Figurative Language (FigLang).*
31. Xie, Z., Kwak, A.S., George, E., Dozal, L.W., Van, H., Jah, M., Furfaro, R. and **Jansen, P.** (2022). Extracting Space Situational Awareness Events from News Text. In *Proceedings of the Language Resources and Evaluation Conference (LREC).*
30. **Jansen, P.**, Smith, K., Moreno, D., Ortiz., H. (2021). On the Challenges of Evaluating Compositional Explanations in Multi-Hop Inference: Relevance, Completeness, and Expert Ratings. In *Proceedings of Empirical Methods in Natural Language Processing (EMNLP).*
29. Dalvi., B., **Jansen, P.**, Tafjord, O., Xie, Z., Smith., H., Pipatanangkura, L., Clark, P. (2021). Explaining Answers with Entailment Trees. In *Proceedings of Empirical Methods in Natural Language Processing (EMNLP).*
28. **Jansen, P.** (2020). Visually-Grounded Planning without Vision: Language Models Infer Detailed Plans from High-level Instructions. In *Findings of Empirical Methods in Natural Language Processing (EMNLP Findings).*
27. **Jansen, P.** (2020). CoSaTa: A Constraint Satisfaction Solver and Interpreted Language for Semi-Structured Tables of Sentences. In *Proceedings of Empirical Methods in Natural Language Processing (EMNLP) System Demonstrations.*
26. Xu, D., **Jansen, P.**, Martin, J., Xie, Z., Yadav, V., Madabushi, H. T., Tafjord O., and Clark, P. (2020). Multi-class Hierarchical Question Classification for Multiple Choice Science Exams. In *Proceedings of the Language Resource and Evaluation Conference (LREC).*
25. Smith, S., Zhang, Z., Culnan, J., and **Jansen, P.** (2020). ScienceExamCER: A High-Density Fine-Grained Science-Domain Corpus for Common Entity Recognition. In *Pro-*

ceedings of the Language Resource and Evaluation Conference (LREC).

24. Xie., Z., Thiem, S., Martin, J., Wainwright, E., Marmorstein., S., and **Jansen, P.** (2020). WorldTree V2: A Corpus of Science-Domain Structured Explanations and Inference Patterns supporting Multi-Hop Inference. In *Proceedings of the Language Resource and Evaluation Conference (LREC).*
23. Khot, T., Clark, P., Guerquin, M, **Jansen, P.**, and Sabharwal., A. (2020). QASC: A Dataset for Question Answering via Sentence Composition. In *Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI).*
22. Thiem, S., and **Jansen, P.** (2019). Extracting Common Inference Patters from Semi-Structured Explanations. In *Proceedings of the Workshop on Commonsense Inference in Natural Language (COIN).*
21. **Jansen, P.** (2018). Multi-hop Inference for Sentence-level TextGraphs: How Challenging is Meaningfully Combining Information for Science Question Answering? In *Proceedings of the 12th Workshop on TextGraphs (TextGraphs).*
20. **Jansen, P.**, Wainwright, E., Marmorstein, S., and Morrison, C. (2018). WorldTree: A Corpus of Explanation Graphs for Elementary Science Questions supporting Multi-hop Inference. In *Proceedings of the Language Resource and Evaluation Conference (LREC).*
19. Kwon, H., Trivedi, H., **Jansen, P.**, Surdeanu, M., and Balasubramanian, N. (2018). Controlling Information Aggregation for Complex Question Answering. In *Proceedings of the European Conference on Information Retrieval (ECIR).*
18. **Jansen, P.** (2017). A Study of Automatically Acquiring Explanatory Inference Patterns from Corpora of Explanations: Lessons from Elementary Science Exams. *Proceedings of the Workshop on Automated Knowledge Base Construction (AKBC).*
17. Sharp, B., Surdeanu, M., **Jansen, P.**, Valenzuela-Escarcega, M. A., Clark, P., and Hammond, M. (2017). Tell Me Why: Using Question Answering as Distant Supervision for Answer Justification. *Proceedings of the Conference on Natural Language Learning (CoNLL).*
16. **Jansen, P.**, Balasubramanian, N., Surdeanu, M., and Clark, P. (2016). What's in an Explanation? Characterizing Knowledge and Inference Requirements for Elementary Science Exams. In *Proceedings of the Conference on Computational Linguistics (COLING).*
15. Sharp, B., Surdeanu, M., **Jansen, P.**, Clark, P., and Hammond, M. (2016). Creating Causal Embeddings for Question Answering with Minimal Supervision. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP).*
14. Sharp, B. **Jansen, P.**, Surdeanu, M., and Clark, P. (2015). Spinning Straw into Gold: Using Free Text to Train Monolingual Alignment Models for Non-factoid Question Answering. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics-Human Language Technologies (NAACL HLT).*
13. **Jansen, P.**, Surdeanu, M., and Clark, P. (2014). Discourse Complements Lexical Semantics for Non-factoid Answer Reranking. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL).*
12. Forbes, A., Surdeanu, M., **Jansen, P.**, and Carrington, J. (2013) Transmitting Narrative: An Interactive Shift-Summarization Tool for Improving Nurse Communication. In *Proceedings of the IEEE Workshop on Interactive Visual Text Analytics.*
11. **Jansen, P.**, Dunlop, M. J., Golish, D. R., and Gehm, M. E. (2012). Adaptive feature-specific spectral imaging. In *Proceedings of the SPIE Defense Security and Sensing Symposium (SPIE DSS).*

Refereed journal articles, published or accepted in final form

10. **Jansen, P.**, Sharp, B., Surdeanu, M., and Clark, P. (2017). Framing Question Answering as Building and Ranking Answer Justifications. *Computational Linguistics (CL), 43, 407-449.*
9. Fried, D., **Jansen, P.**, Hahn-Powell, G., Surdeanu, M., and Clark, P. (2015). Higher-order Lexical Semantic Models for Non-factoid Answer Reranking. *Transactions of the Association of Computational Linguists (TACL), 3, 197-210.*
8. Golish, D., Vera, E., Kelly, K., Gong, Q., **Jansen, P.**, Hughes, J., Kittle, D., Brady, D., and Gehm, M. (2012). Development of a scalable image formation pipeline for multiscale gigapixel photography. *Optics Express (OE), 20, 22048-22062.*
* 7. **Jansen, P.**, and Watter, S. (2012). Strong systematicity through sensorimotor conceptual grounding: an unsupervised, developmental approach to connectionist sentence processing. *Connection Science (CS), 24, 25-55.*
* 6. **Jansen, P.**, Fiacconi, C., and Gibson, L. (2010). A computational vector-map model of neonate saccades: Modulating the externality effect through refraction periods. *Vision Research (VR), 50, 2551-2558.*
* 5. **Jansen, P.**, and Watter, S. (2008). SayWhen: An automated method for high-accuracy speech onset detection. *Behavior Research Methods (BRM), 40, 744-751.*

** denotes publications substantially based on work completed as a graduate student.*

Chapters in scholarly books and monographs

4. Walls, R., and **Jansen, P.** (2022) Information Organization, Storage, and Management. In Jah., M. (ed.) *Space Domain Awareness.* Colorado Springs: CEI, 219-242.

Non-refereed conference and workshop articles

3. Thayaparan, M., Valentino, M., **Jansen, P.**, Ustalov, D. (2021). TextGraphs 2021 Shared Task on Multi-Hop Inference for Explanation Regeneration. In Proceedings of the 15th Workshop on TextGraphs (TextGraphs).
2. **Jansen, P.**, Ustalov, D. (2020). TextGraphs 2020 Shared Task on Multi-Hop Inference for Explanation Regeneration. In Proceedings of the 14th Workshop on TextGraphs (TextGraphs).
1. **Jansen, P.**, Ustalov, D. (2019). TextGraphs 2019 Shared Task on Multi-Hop Inference for Explanation Regeneration. In Proceedings of the 13th Workshop on TextGraphs (TextGraphs).

Selected Computer Programs and Open Source Hardware

Open Source Hardware is like Open Source Software, but also includes designs for electronic hardware (electronics and mechanical designs), all released under similarly permissive licenses.

P3. **Open Source High-Speed Magnetic Camera.** A high-speed camera for magnetic fields. 2018. `https://www.sparkfun.com/products/14652/`.

P2. **Open Source Computed Tomography (CT) Scanner.** Two models of computed tomography (CT) scanners, intended for science pedagogy. 2014-2016. `https://hackaday.io/project/5946-openct2`.

P1. **The Tricorder Project.** Several models of handheld scientific scanners, similar to the Tricorder from Star Trek. 2014-2023. `https://hackaday.io/project/1395-open-source-science-tricorder`.

Media

Exhibits

2015 **German Museum of Technology**. Open Source Science Tricorder Project placed on permanent exhibit. Berlin, Germany. `http://www.sdtb.de`

Selected Media Coverage

Automated Scientific Discovery

M1. **New Scientist: Human scientists are still better than AI ones – for now (October 2024)**

COVID-19

M1. **UANews: UArizona Makers Race to Provide Personal Protective Equipment (April 2020)**
M2. **UA SBS: SBS Champions: iSchool Professors Peter Jansen and Win Burleson Make COVID-19 Equipment (May 2020)**

UA Teaching

M3. **UANews: Students in Engineering, iSchool Discover Sweet Success (June 2017)**
M4. **Wildcat News: Eureka: UA maker class challenges student engineers (June 2017)**

Open Source Hardware

M5. **Washington Post: Homemade tricorders and handheld health care (March 2012)**
M6. **TechCrunch: A Chicken In Every Pot And An Open-Source Tricorder In Every Home (March 2012)**
M7. **Ars Technica: Researcher publishes specs for real Linux-powered Star Trek tricorder (March 2012)**.
M8. **WIRED (UK): Researcher publishes specs for real Linux-powered Star Trek tricorder (March 2012)**
M9. **PBS Arizona (TV): Technology and Innovation: A Working Tricorder (March 2012)**
M10. **THE VERGE: Scientist designs and shares open-source plans for real-world Tricorders (March 2012)**
M11. **Engadget: Tricorder designs go open course: can detect magnetic fields, reveal Trekkies (March 2012)**
M12. **BoingBoing (Cory Doctorow): Open source "tricorders": handheld sensor packages for everyone (March 2012)**

M13. **International Business Times: Star Trek-like Tricorder Now a Reality: Why We're so Excited? (March 2012)**
M14. **Phys.org: Cognitive researcher designs and builds a real-world modular working tricorder (March 2012)**
M15. **Vice.com: Star Trekkin' IRL: The Iconic Tricorder Actually Exists (March 2012)**
M16. **Gizmodo: The World Gets One Step Closer To a Working Tricorder (March 2012)**
M17. **ZDNet Smart Planet: How to make your own tricorder (March 2012)**
M18. **MSNBC: Star Trek-like open-source tricorder sees magnetic fields and more (March 2012)**
M19. **Giyism: Nerds invent tricorder, of course it runs on Linux (March 2012)**
M20. **CBC: 'Tricorder' project seeks helping hands (April 2012)**
M21. **Reuters: Scientist beams up a real "Star Trek" tricorder (April 2012)**.
M22. **Forbes: Tricorder Update – Social Medicine is the Next Big Thing After Social Media (May 2012)**
M23. **NASA.GOV: From Star Trek to SCOUT: The Story of a Real-World Medical Tricorder (May 2012)**
M24. **Mythbusters Jamie & Adam's TESTED.COM: Maker Profile: Peter Jansen's Tricorder Project (February 2013)**
M25. **Slashdot.org: Tricorder Project Releases Prototype Open Source 3D Printable Spectrometer (September 2013)**
M26. **MAKE Magazine: Open-Source CT Scanner (April 2014)**
M27. **Bloomberg: Star Trek's Tricorders are Almost Here (June 2014)**
M28. **IEEE Spectrum: Make It So: Open Source, Arduino-Based Tricorder Nears Completion (October 2014)**
M29. **CNET Tomorrow Daily: An Arduino Tricorder (October 2014)**
M30. **UA News: Attention, Trekkies: Get your Tricorders here (November 2014)**
M31. **Smithsonian: A List of All the Times People Have Tried to Build a Working Tricorder (July 2015)**
M32. **Hackaday: Imaging Magnetism with a Hall Effect Camera (August 2017)**
M33. **Hackaday: High Speed Imaging of Magnetic Fields (February 2018)**
M34. **Hackaday: Coin-Operated Graphing Calculator Console (February 2022)**

Conferences/Scholarly Presentations

Invited Conferences

2022 **WordPlay Workshop** at North American Association for Computational Linguistics (NAACL)

Invited Colloquia

2018 **Allen Institute for Artificial Intelligence**. Distinguisted Lecture Series
2016 **University of Albany**. Department of Physics

Awarded Grants and Contracts

Federal

2025-2027 **Defense Advanced Research Projects Agency (DARPA)**
Title: Harnessing Artificial Intelligence and Language Modeling for Enhancing Innovation and Evaluating Research Claims
Amount: $7M (UA portion, $1M).
Role: University of Arizona PI (University of Pennsylvania lead; joint with University of Washington)

2018-2023 **National Science Foundation (NSF)**
Title: Explainable Natural Language Inference (Collaborative Research)
Amount: $254,464 (UA portion, total award $499,001).
Role: University of Arizona PI (Co-I: Surdeanu; Stonybrook PI: Balasubramanian)

Private Foundation

2022 **Allen Institute for Artificial Intelligence (AI2)**
Title: Compositional Explanations (continuation)
Amount: $50,000
Role: PI

2021 **Allen Institute for Artificial Intelligence (AI2)**
Title: Compositional Explanations
Amount: $50,000
Role: PI

2017 **Allen Institute for Artificial Intelligence (AI2)**
Title: Explanation-centered Structured Knowledge Base for Science Question Answering
Amount: $60,000
Role: PI

Honors and Awards

2021 **Teaching Award**, School of Information, University of Arizona (Department-Level)

Service/Outreach

National/International Service/Outreach

2025	AI & Scientific Discovery (AISD) Workshop 2025 Organizing Committee
2024	Structured Explanations Workshop 2024 Organizing Committee
2023	Structured Explanations Workshop 2023 Organizing Committee
2021	TextGraphs Workshop 2021 Shared Task and Organizing Committee
2020	National Science Foundation: Information and Intelligent Systems (IIS) Reviewer
2020	TextGraphs Workshop 2020 Shared Task and Organizing Committee
2019	TextGraphs Workshop 2019 Shared Task and Organizing Committee
2017	Public Talk, Penguicon, Detroit, MI
2016	Public Talk, North-east Star Trek Convention, Albany, NY