Carl N. Edwards

Champaign-Urbana, Illinois ● cne2@illinois.edu ● cnedwards.com ● ●

Education

University of Illinois Urbana-Champaign — PhD Candidate in Computer Science — Fall 2020 — Summer 2025
Focus: Artificial Intelligence, NLP for Molecular and Drug Discovery
Advisor: Professor Heng Ji
Committee: Heng Ji, Jiawei Han, Martin Burke, Kyunghyun Cho, ChengXiang Zhai, Gabriele Scalia
Thesis: Integrating Natural Language with Molecular Structure

University of Tennessee-Knoxville — Honors Computer Science BS — May 2020
Honors Math Double Major
STEM GPA: 4.0/4.0 — Overall GPA: 3.99/4.0
Summa Cum Laude — Chancellor's Honors Program
Mathematics Honors Program — Engineering Honors Program

University of Zürich — EuroScholars Research Study Abroad — Fall Semester 2018
Dynamic and Distributed Information Systems Group
Advisor: Abraham Bernstein
Project: Linking Knowledge Graphs and Images Using Embeddings

Experience

Genentech – Biology Research | AI Development (BRAID) — Senior AI Research Scientist — Fall 2025-

- I will be working on AI agents, LLM benchmarking, and multimodal LLM development.

Genentech – Biology Research | AI Development (BRAID) — Research Intern — Summer-Fall 2024

- Investigated the usage of natural language obtained from large language models (LLMs) for injecting biological domain knowledge into the virtual screening process. Proposed a large-scale property prediction benchmark for the task.
- Explored instruction tuning and reinforcement learning with molecular feedback for improving LLM capabilities in the chemical domain.
- Examined evaluation methodologies for zero-shot discovery of novel chemistries and mechanisms of action in a large-scale antibiotic discovery campaign.

University of Illinois Urbana-Champaign – NSF Molecule Maker Lab Institute — Research Assistant — Fall 2020-2025

- Engaged in research community building by preparing tutorial and workshop proposals on "Language + Molecules," securing acceptance at EACL 2024 in Malta and ACL 2024 in Bangkok, and serving as the lead organizer for the "Language + Molecules" workshop.
- Developed systems which ingest large quantities of scientific papers to improve understanding of drugs in order to propose variants with improved properties. Algorithms are designed for seamless integration with "Lego chemistry" for rapid synthesis. (experimental results in progress, details withheld).
- Researched integrating molecular and natural language information. Work proposes *Text2Mol* and *MolT5* models, enabling novel downstream tasks of cross-modal retrieval of molecules from natural language queries, generation of molecules from textual descriptions, and molecule captioning. Overall research direction is towards language-enabled functional control of molecule design for applications such as drug development, organic photovoltaics, and advanced material design.
- Conducted information extraction, information retrieval, and text mining research to enable the *AlphaSynthesis* platform, which allows AI-assisted synthesis planning for molecular discovery and manufacturing based on large-scale analysis of scientific literature.

Allen Institute for Artificial Intelligence (AI2) – Semantic Scholar Research Intern Summer 2022

- Developed an AI model, *SynerGPT*, which uses a novel pretraining strategy to enable in-context learning for few-shot drug synergy predictions for rare drugs and cell lines. The resulting paper was accepted at COLM.
- Investigated context optimization for *SynerGPT* with future applications in creating a standardized assay for patient tumor biopsies that can enable patient-specific drug synergy prediction.
- Developed a methodology for a novel approach to drug design conditioned on desirable synergy tuples.

Carnegie Mellon University – Auton Lab Robotics Institute Summer Scholar (NSF REU) Summer 2019

- Detected organizations in multimodal sex trafficking dataset consisting of over 40 million data points.
- Integrated multiple similarity measures using face detection, word embeddings, and regex-extracted features to detect organizations in tens of millions of escort advertisements scraped from online sources.

University of Zurich – Dynamic and Distributed Information Systems Group EuroScholar Fall 2018

- Crafted two linked datasets between knowledge graphs and images.
- Incorporated embedding models based on InceptionNet and TransE into a joint embedding model.

Oak Ridge National Laboratory – CISR Department of Energy SULI Intern Summer 2018

- Researched global optimization algorithms for subarrayed phase-only radar beam synthesis.

University of Tennessee – Material Research and Innovation Lab Undergraduate Research Assistant Fall 2016-2018

- Performed Brownian dynamics simulations of flowing polymer solutions.
- Created visualizations from resulting data and computed solution physical properties.
- Prepared manuscript for publication in Journal of Molecular Graphics and Modelling.

Oak Ridge National Laboratory — CISR Higher Education Research Experiences Intern Summer 2017

- Developed radar simulations using proprietary technology.
- Designed phased array radar beam width optimization algorithms (GA, particle swarm, simulated annealing).

Nanomechanics, Incorporated. Software Engineering Intern Summer 2016

- Programmed proprietary software to interpret high speed data using C++ .
- Developed programs for Linux, Raspberry Pi, and ALSA libraries.

Preprints

C. Edwards*, C. Han*, G. Lee, T. Nguyen, B. Jin, C.K. Prasad, S. Szymkuć, B.A. Grzybowski, Y. Diao, J. Han, G. Liu, H. Peng, M.D. Burke, H. Ji. "mCLM: A Function-Infused and Synthesis-Friendly Modular Chemical Language Model." *arXiv preprint*. *arXiv:2505.12565*. 2025.

Conference Publications †

X. Li*, L. Wang*, Y. Luo, **C. Edwards**, S. Gui, Y. Lin, H. Ji, S. Ji. "Geometry Informed Tokenization of Molecules for Language Model Generation." In *Proceedings of the Forty-second International Conference on Machine Learning* (ICML). 2025.

C. Edwards, Z. Lu, E. Hajiramezanali, T. Biancalani, H. Ji, G. Scalia. "MolCap-Arena: A Comprehensive Captioning Benchmark on Language-Enhanced Molecular Property Prediction" *ICLR Workshop on Machine Learning for Genomics Explorations (MLGenX)*. 2025.

K. Yan, X. Li, H. Ling, K. Ashen, **C. Edwards**, R. Arroyave, M. Zitnik, H. Ji, X. Qian, X. Qian, S. Ji. "Invariant Tokenization for Language Model Enabled Crystal Materials Generation." In *Proceedings of the Thirty-Eighth Conference on Neural Information Processing Systems*. 2024.

C. Edwards, Q. Wang, L. Zhou, and H. Ji. "L+M-24: Building a Dataset for Language + Molecules @ ACL 2024." *ACL Workshop on Language + Molecules*. 2024.

C. Edwards, A. Naik, T. Khot, M. Burke, H. Ji, and T. Hope. "SynerGPT: In-Context Learning for Personalized Drug Synergy Prediction and Drug Design." In *Proceedings of the First Conference on Language Modeling* (COLM). 2024.

T. Nguyen‡, T. Torres-Flores, C. Hwang, **C. Edwards**, Y. Diao, and H. Ji. "GLaD: Synergizing Molecular Graphs and Language Descriptors for Enhanced Power Conversion Efficiency Prediction in Organic Photovoltaic Devices." In *Proceedings of 33rd ACM International Conference on Information and Knowledge Management (CIKM 2024)*. 2024.

H. Sprueill, **C. Edwards**, M. Olarte, U. Sanyal, C. Johnston, H. Liu, H. Ji, and S. Choudhury. "ChemReasoner: Heuristic Search over a Large Language Model's Knowledge Space using Quantum-Chemical Feedback." In *Proceedings of the Forty-first International Conference on Machine Learning* (ICML). 2024.

L. Zhou‡, **C. Edwards**, and H. Ji. "What a Scientific Language Model Knows and Doesn't Know about Chemistry." *NeurIPS AI for Science*. 2023.

S. Li, C. Han, P. Yu, **C. Edwards**, M. Li, X. Wang, Y. Fung, C. Yu, J. R. Tetreault, E. H. Hovy, and H. Ji. "Defining a New NLP Playground." In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (EMNLP2023) Findings. 2023.

H. Sprueill, **C. Edwards**, M. Olarte, U. Sanyal, H. Ji, and S. Choudhury. "Monte Carlo Thought Search: Large Language Model Querying for Complex Scientific Reasoning in Catalyst Design." In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (EMNLP2023) Findings. 2023.

C. **Edwards** and H. Ji. "Semi-supervised New Event Type Induction and Description via Contrastive Loss-Enforced Batch Attention." In *Proceedings of The 17th Conference of the European Chapter of the Association for Computational Linguistics* (EACL2023). 2023.

C. Edwards*, T. Lai*, K. Ros, G. Honke, K. Cho, and H. Ji. "Translation between Molecules and Natural Language." In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing* (EMNLP2022). 2022.[1]

[1] (Oral presentation, Top 4.8% of submissions)

X. Du, […], **C. Edwards**, […] and H. Ji. "Resin-11: Schema-guided event prediction for 11 newsworthy scenarios." In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: System Demonstrations* (pp. 54-63). 2022.

C. Edwards, CX. Zhai, and H. Ji. "Text2Mol: Cross-modal Molecule Retrieval with Natural Language Queries." In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (EMNLP2021). 2021.

K. Ros*, **C. Edwards***, H. Ji, and CX. Zhai. "Team Skeletor at Touché 2021: Argument Retrieval and Visualization for Controversial Questions." *CEUR Workshop Proceedings*. Vol. 2936. CEUR-WS. 2021.[2]

[2] Ranked 2nd/22 teams for retrieval quality.

† In my field of computer science, main conference publications at top venues are considered equivalent to journal articles and are often the preferred venue of publication.

* indicates equal contribution
‡ = mentored student.

Journal Articles

X. Zhang*, L. Wang*, J. Helwig*, Y. Luo, C. Fu*, Y. Xie*, […], **C. Edwards**, […], A. Aspuru-Guzik, E. Bekkers, M. Bronstein, M. Zitnik, A. Anandkumar, S. Ermon, P. Liò, R. Yu, S. Günnemann, J. Leskovec, H. Ji, J. Sun, R. Barzilay, T. Jaakkola, C. W. Coley, X. Qian, X. Qian, T. Smidt, and S. Ji. "Artificial Intelligence for Science in Quantum, Atomistic, and Continuum Systems." *Foundations and Trends in Machine Learning*. 2025.

C. Edwards, M. H. Nafar Sefiddashti, B. J. Edwards, and B. Khomami. "In-plane and out-of-plane rotational motion of individual chain molecules in steady shear flow of polymer melts and solutions." *J. Mol. Graph. Model*., *81*, 184-196. 2018.

C. Edwards, A. Wertz, and A. Dubrawski. "Using Similarity Measures to Detect Organizations in Online Escort Advertisements." *Robotics Institute Summer Scholar' Working Papers Journal*, 7, 43-49. 2019.

Workshop Organization

Q. Wang, W. Yin, L. Huang, M. Fung, X. Du, **C. Edwards**, and T. Hope. "Towards a Knowledge-grounded Scientific Research Lifecycle." Located at the 39th Annual AAAI Conference on Artificial Intelligence (AAAI2025). Philidelphia, USA. 2025.

C. Edwards, H. Ji, T. Hope, M. Li, Q. Wang, and L. Zhou. "Language + Molecules." *Located at the 62nd Conference of the Association for Computational Linguistics* (ACL2024). Bangkok, Thailand. 2024.

Tutorials

Q. Wang**, C. Edwards**, H. Ji, and T. Hope. "Towards a Human-Computer Collaborative Scientific Paper Lifecycle: A Pilot Study and Hands-On Tutorial." *In Proceedings of The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation* (LREC-COLING2024). 2024.

C. Edwards, Q. Wang, and H. Ji. "Language + Molecules." *In Proceedings of The 18th Conference of the European Chapter of the Association for Computational Linguistics* (EACL2024). 2024.

Invited Talks

- "Molecule-Language Multimodality." ChemLLMathon webinar, Chemindix 2024 Conference, Saudi Aramco, November 2024.
- "NLP for Scientific Discovery in Chemistry and Biomedicine." Workshop on Cross-Disciplinary Challenges and Opportunities in AI Applications to Science and Engineering at Michigan Tech, November 2024.
- "NLP for Scientific Discovery in Chemistry and Biomedicine." Keynote Talk, AI4Science Birds of a Feather at ACL2024, August 2024.
- "Language-Empowered Chemistry." AI-Empowered Chemistry Course Guest Lecture, NSF Molecule Maker Lab Institute, May 2024.
- "Integrating Molecules and Language." Delta Project AI Training Series. National Center for Supercomputing Applications, February 2024.
- "Integrating Molecules and Language for Discovery and Search." National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, November 2023.
- "Language-Guided Scientific Discovery for Chemistry." First International Conference for the Center of the Transformation of Chemistry (CTC), Ringberg Conference, Max Planck Institute of Colloids and Interfaces, September 2023.
- "Language-Guided Scientific Discovery for Chemistry." UIUC Campus-Wide NLP Talk Series, September 2023.
- "Translation Between Molecules and Natural Language." NVIDIA, November 2022.

Presented Posters

- "Extreme-Scale Heterogeneous Inference with Large Language Models and Atomistic Graph Neural Networks for Catalyst Discovery." Sprueill H.W., C. Edwards, M.V. Olarte, U. Sanyal, K. Agarwal, H. Ji, and S. Choudhury. 03/18/2024. American Chemical Society Spring 2024 National Meeting, New Orleans, Louisiana.
- "Literature-Based Kinase Inhibitor Fragment Replacement via Frequent Pattern Mining for Blood Brain Barrier Penetration." Molecule Maker Lab Institute NSF Site Visit 2023.
- "Translation Between Molecules and Natural Language." MMLI NSF Site Visit 2022, University of Chicago AI+Science Summer School.
- "Semi-supervised New Event Type Induction." EACL2023.
- "Text2Mol: Cross-modal Molecular Retrieval with Natural Language Queries." EMNLP2021, MMLI Annual Retreat 2021
- "Using Similarity Measures to Detect Organizations in Online Escort Advertisements", RISS Poster Session 2019.
- "Beam Broadening of Subarrayed Radar Arrays Utilizing Various Global Optimization Techniques" ORNL Undergraduate Poster Session 2018.

Accepted Abstracts

- "Integrating generative AI with computational chemistry for catalyst design in biofuel/bioproduct applications." Sprueill H.W., C. Edwards, M.V. Olarte, U. Sanyal, H. Ji, and S. Choudhury. American Chemical Society Spring 2024 National Meeting, New Orleans, Louisiana. March 18, 2024.
- "Out-of-plane rotational motion in shear flow of polymer melts and solutions." M.H. Nafar Sefiddashti, C.N. Edwards, B.J. Edwards, and B. Khomami, The Society of Rheology 89th Annual Meeting, Denver, CO, October 8-12, 2017.

Other

- B. Daniel, **C. Edwards**, and A. Anderson. "Phase-Only Beam Broadening of Contiguous Uniform Subarrayed Arrays Utilizing Three Metaheuristic Global Optimization Techniques." *arXiv preprint arXiv:2009.06123.* 2020.

Teaching, Leadership, and Service

- **Organizing Committee:** Language + Molecules Workshop at ACL 2024, 2nd AI for Research at AAAI 2025
- **Area Chair:** EMNLP 2024
- **Reviewing:** ACL-IJCNLP**,** ACL, NAACL, EACL, EMNLP, NeurIPS 2023 AI4Science, GenBio, AAAI, ACL Rolling Review, ICML AI4Science, Bioinformatics
- **NSF Molecule Maker Laboratory Institute:**
  - Student and Postdoc Council Educational & Outreach Activities Chair 2021-2023, Social Committee Member 2023-present
  - Awarded Certificate of Public Engagement
- **Teaching Assistant Experience**:
  - CS 412 – Introduction to Data Mining (Spring 2021)
  - CS 125 – Introduction to Computer Science (Fall 2020)
- **Undergraduate Mentorship:**
  - **Summer 2023:** Mentored three undergraduates on projects related to language-enabled protein composition and design, language-molecule association rule mining with language models on scientific literature, and scientific language model factuality evaluation and updating for knowledge base construction. Projects led to SOTA text-molecule IR results, an AI4Science workshop acceptance and ACL workshop acceptance, and a Master's thesis topic.
- **Outreach:**
  - **November 2023:** Cena y Ciencias – Dual language outreach program presenting DIY Solar Cell activity
- **Presentations to High School Students:**
  - **June 2023:** "AI and NLP for Drug Discovery and My Path", MMLI Summer Camp
  - **April 2021, 2022:** Illinois CS Sail course on "Learning Word Representations"
  - **Summer 2019:** Presentation on "Detecting Sex Trafficking Organizations" with AI4All@CMU
- **Presentations to Middle School Students:**
  - **April 2023:** Presentation on "Intro to AI for Chemistry" for underrepresented middle school boys.
  - **March 2023:** Presentation on "Intro to AI for Chemistry" for underrepresented middle school girls.

Skills

- **Programming**: Proficient in Python, C++, Matlab, and Java; Experience with C, Fortran, LaTeX, HTML, and CSS. Experience with Windows and Linux.
  - Proficient with PyTorch, Lightning, TensorFlow, Keras, pandas, NumPy, SciPy, NLTK, spaCy, PyMOL, RDKit
- **Computational Methods**: Natural language processing (NLP), information retrieval, information extraction, text mining, natural language generation, large language models, representation learning (text, knowledge graph, molecule, image, etc.), deep learning, machine learning, knowledge graphs, multimodal data, computer vision, reinforcement learning, transfer learning, scientific computing, statistical decision theory, normalizing flows, diffusion models, cheminformatics, molecule generation, molecule property prediction, quantitative structure-activity relationship (QSAR) prediction, physics-informed neural networks, AI for science, computational drug discovery, protein generation, protein language models, graph neural networks, neural ordinary differential equations, data mining, pattern recognition.

News Coverage

- MolT5: Chinese News, Amazon, Social Media(upper bound of 518,019 Twitter followers)
- Undergraduate: Edwards is Top Fellowship/Scholarship Student in Computer Science
  - Math Department: Outstanding Math Honors Students Pursuing Advanced Degrees
  - Goldwater Scholarship Coverage: University-wide, Math Department

Awards

- Selected Participant for University of Chicago DSI AI & Science Summer School (2023)
- Saburo Muroga Endowed Fellowship, UIUC (2020) – awarded to outstanding computer science graduate students
- **Goldwater Scholarship** (2019) – "the most prestigious undergraduate scholarship given in the natural sciences, engineering and mathematics" in the United States (Wikipedia)
- UTK Outstanding Computer Science Junior (2018) – awarded to a single junior in computer science based on academic merit
- Min H. Kao Scholar ($7,500) (2018, 2019) – roughly six students selected from EE, CS, and CE majors based on academic merit
- Pi Mu Epsilon Math Society Award Membership (2019)
- Schmitt Memorial Scholarship (Math) (2019)
- Thomas & Kathryn Shelton Award (2017, 2018, 2019)
- Volunteer Scholarship (2016-2020)
- UT Provost Scholarship (2016-2020)
- National Merit Scholar (2016)
- National AP Scholar (2016)
- State of Tennessee Governor's School for Computational Physics Attendee (2015)

Relevant Coursework

- Knowledge-Driven Natural Language Generation
- Scientific Machine Learning
- Data Mining Principles
- Natural Language Processing
- Deep Generative and Dynamical Models
- Advanced Information Retrieval
- Text Mining: A New Paradigm
- Transfer Learning
- Biologically-Inspired Computation
- Reinforcement Learning
- Introduction to Pattern Recognition
- Deep Learning

Activities

- **Illini Dancesport:** 2022-2024, Executive Board Member
- **UTK Machine Learning Club:** 2017-2020, Executive Board Member
- **HackUTK**: 2016-2020, UTK cybersecurity organization, VolHacks Hackathon 2016, 2017, 2019
- **Tau Beta Pi**: 2018-2020, National Engineering Honors Society
- **Taekwondo**: 2007-2016, martial art, black belt

- **Classical Singing**: 2015-2020, 2023-2025 (UIUC Symphonic Choir)
- **FIRST Robotics**: 2011-2016**,** Programming leader and team co-leader