ILANA SHAPIRO

**CONTACT**

|  |  |
| --- | --- |
| Email: [ilshapiro@ucsd.edu](mailto:ilshapiro@ucsd.edu) | LinkedIn: [linkedin.com/in/ilana-shapiro-157447170](https://www.linkedin.com/in/ilana-shapiro-157447170/) |
| Website: <https://ilanashapiro.github.io> | GitHub: [github.com/ilanashapiro](https://github.com/ilanashapiro) |

**RESEARCH INTERESTS**

My interests lie in the integration of programming languages, automated reasoning, generative AI, and human-computer interaction. I aim to develop usable structural constraints on sequence models for the controllable generation of well-formed sequence data, and to apply formal reasoning methods to improve the automated reasoning tools that underlie these systems.

**EDUCATION**

## University of California, San Diego, La Jolla, CA *2023-present*

Ph.D. Computer Science

Programming Systems Group, advised by Professor Sorin Lerner

GPA: 3.96/4.0

## Pomona College, Claremont, CA *2018-2022*

B.A. Computer Science/Music (Flute) double major, minor in Mathematics

GPA: 4.0/4.0, Summa Cum Laude, Distinction in Senior Exercises

* ***Computer Science Thesis:*** "MusAssist: A Domain Specific Language for Music Notation"
  + Advised at Harvey Mudd College by Professor Ben Wiedermann
* ***Music Thesis and Recital:*** "Mieczysław Weinberg: Music Transcending Tragedy"
  + Advised by Professors Alfred Cramer, Joti Rockwell, and Eric Lindholm

**RESEARCH PROJECTS**

## Parallelizing Z3: Adaptive Cubing via Online Sampling of CDCL Conflict Traces[[code](https://github.com/Z3Prover/z3/blob/master/src/smt/smt_parallel.cpp)][[slides](https://docs.google.com/presentation/d/1fG-2aDKSI7zXmeVLD6aNk_5dQ2f9KXjodczDli39O_Q/edit)] *2025*

## At Microsoft Research, I worked with Nikolaj Bjorner on a novel parallelization algorithm for the SMT solver Z3. We developed an online cube-and-conquer approach: cubes are dynamically sampled during solving from CDCL conflict-variable heuristics and distributed to worker threads based on similarity. We are now exploring online parameter tuning.

## Training LLMs for Verified Programming [[code](https://github.com/ilanashapiro/lean-server)] *2025*

## At Microsoft Research, I helped train a 32B LLM specialized in program verification. I prepared and augmented Lean datasets for SFT and RL training, built a custom Dockerized Lean verification server, and evaluated model checkpoints.

## Synthesizing Composite Hierarchical Structure from Music Corpora[[paper](https://www.ijcai.org/proceedings/2025/1128.pdf)][[code](https://github.com/ilanashapiro/constraints_project)][[slides](https://docs.google.com/presentation/d/16hWNSMnztE_oKnXoRR_6XZm5xT8-Ewi_4FO_J33ekZE/edit?usp=sharing)][[talk](https://www.youtube.com/watch?v=7e-srxx-zrU)] *2025*

* Published at 34th International Joint Conference on Artificial Intelligence 2025. Combined stochastic and SMT techniques to frame and solve the nested NP-hard combinatorial optimization problem of music structure summarization as an extension of the Generalized Median Graph problem.

## Lexically Constrained Decoding of Transformers [[paper](https://ilanashapiro.github.io/files/gbs_paper.pdf)][[code](https://github.com/ilanashapiro/constrained_decoding_gbs_transformers)][[slides](https://docs.google.com/presentation/d/1XNCqX4Eab9vS3RscQFMXXLrzUaGAJc-_I9JRulo6qjc/edit?usp=sharing)] *2025*

* Adapted the constrained decoding algorithm Grid Beam Search (GBS) to impose lexical constraints on GPT2, and fine-tuned GPT2 on a corpus of Chekhov’s stories. GBS + fine-tuned GPT2 subjectively outperformed GBS + GPT2 alone.

## The Impact of GitHub Copilot on Test-First Development [[paper](http://ilanashapiro.github.io/files/tfd.pdf)][[code](https://github.com/broad-well/tdd-copilot-study/tree/main)] *2024*

## Conducted between-subjects pilot study to determine impact of Copilot on Test-First Development. Thematic analysis revealed that while Copilot enhanced coding speed, it resulted in superficial problem comprehension and decreased scope of the test suites.

**pgen-rs: LLM-Aided, Efficient, User-Friendly Genomic Data Wrangling** [[paper](http://ilanashapiro.github.io/files/pgen-rs.pdf)][[code](https://github.com/ilanashapiro/pgen-rs)][[slides](https://docs.google.com/presentation/d/1vJvrpacwQvY3_T-XpP-YkS1fOGqg6rAR5ZW3-SYXVUI/edit?usp=sharing)] *2024*

* Developed pgen-rs, a tool enabling end-users to write genomic data wrangling requirements in natural language and execute with Rust-based high-performance genomic data processor.

## ProCon: Continuous Enumeration for Just-In-Time Bottom-Up Synthesis [[paper](http://ilanashapiro.github.io/files/procon.pdf)][[code](https://github.com/rkthomps/ProCon)] *2024*

* Introduced continuous, rule-based enumeration for just-in-time bottom-up search in SyGuS problems, where programs are enumerated in order of continuous, nonrounded weights as determined by a probabilistic weighting function.

## MusAssist: A Domain Specific Language for Music [[TENOR paper](https://www.tenor-conference.org/proceedings/2023/11-TENOR_BOSTON_2023_paper_9804Shapiro.pdf)][[thesis2](https://cs.pomona.edu/classes/cs190/thesis_examples/Shapiro.22.pdf)][[code](https://github.com/ilanashapiro/MusAssist)][[demo](https://www.youtube.com/watch?v=Gok-O0Fjgd4)] *2022*

## Published at 8th International Conference on Technologies for Music Notation and Representation 2023. Created DSL bridging the abstraction gap between music theoretical structures and notation. Wrote Haskell-based compiler to MusicXML.

## Mieczysław Weinberg: Music Transcending Tragedy [[thesis1](https://scholarship.claremont.edu/cgi/viewcontent.cgi?article=1294&context=pomona_theses)][[recital](https://www.youtube.com/playlist?list=PLPwyBV1x-R8ZPdgk8EtO-Gh8l6Li0SH0j)] *2022*

* Published in Scholarship@Claremont. Wrote extensive musicology thesis examining narrative and memory in Weinberg’s *Kaddish* Symphony. Presented flute recital of my transcriptions of Weinberg’s cello works.

## Markov Chain Music Generation [[JHM paper](https://scholarship.claremont.edu/cgi/viewcontent.cgi?article=1848&context=jhm)][[code](https://github.com/ilanashapiro/Markov-Model-Music-Generationhttps:/github.com/ilanashapiro/Markov-Model-Music-Generation)] *2021*

* Published in *Journal of Humanistic Mathematics*. Created a novel system of Markov chains using inverse transform sampling, enabling end-users to rapidly generate musical sketches.

## Virtual Ensemble Assembly: Musicality in Separation [[WAC paper](https://doi.org/10.5281/zenodo.6860879)][[code](https://github.com/kpet123/Virtual-Ensemble-Assembly)] *2020*

* Published at Web Audio Conference 2022. Assisted on Prof. Christopher Raphael’s research at Indiana University Bloomington exploring synchronizing audio tracks without click tracks.

**PUBLICATIONS**

|  |  |  |
| --- | --- | --- |
| *\* equal contribution*  **Ilana Shapiro**, Ruanqianqian (Lisa) Huang, Zachary Novack, Cheng-i Wang, Hao-Wen Dong, Taylor Berg-Kirkpatrick, Shlomo Dubnov, and Sorin Lerner. “Synthesizing Composite Hierarchical Structure from Symbolic Music Corpora.” In *Proceedings of the 34th International Joint Conference on Artifical Intelligence (IJCAI ‘25),* Montreal, Canada, 2025.  **Ilana Shapiro**,Shubham Saha, Diya Lakhani, Shree Venkatesh, and Runqiu Xu. “Grid Beam Search for Constrained GPT-2 Decoding” Unpublished manuscript, 2025.  **Ilana Shapiro**, Michael Peng, and Andrew Lara. “The Impact of GitHub Copilot on Test-First Development.” Unpublished manuscript, 2024.  Cole Kurashige,\* Savitha Ravi,\* and **Ilana Shapiro**.**\*** “pgen-rs: LLM-Aided Efficient and User-Friendly Genomic Data Wrangling.” Unpublished manuscript, 2024.  Kyle Thompson, **Ilana Shapiro**, Ani Canumalla. “ProCon: Continuous Enumeration for Just-In-Time Bottom-Up Synthesis.” Unpublished manuscript, 2024.  **Ilana Shapiro.** "MusAssist: A Domain Specific Language for Music Notation." *Proceedings of the International Conference on Technologies for Music Notation and Representation (TENOR’23),* pp. 75-82, Northeastern University, Boston, MA, 2023.  Kaitlin Pet, **Ilana Shapiro**, and Christopher Raphael. “Virtual Ensemble Assembly: Musicality in Separation.” In *Web Audio Conference (WAC’22*), Cannes, France, 2022.  **Ilana Shapiro**. 2022. *MusAssist: A Domain Specific Language for Music Notation*. Bachelor’s thesis. Pomona College.  **Ilana Shapiro**. 2022. *Mieczysław Weinberg: Music Transcending Tragedy*. Bachelor’s thesis. Pomona College.  **Ilana Shapiro** and Mark Huber. “Markov Chains for Computer Music Generation.” In *Journal of Humanistic Mathematics, Volume 11 Issue 2 (July 2021), pp. 167-195*. | [[paper8](https://arxiv.org/abs/2502.15849)]  [[paper7](https://ilanashapiro.github.io/files/gbs_paper.pdf)]  [[paper6](https://ilanashapiro.github.io/files/tfd.pdf)]      [[paper5](https://ilanashapiro.github.io/files/pgen-rs.pdf)]  [[paper4](https://ilanashapiro.github.io/files/procon.pdf)]  [[paper3](https://www.tenor-conference.org/proceedings/2023/11-TENOR_BOSTON_2023_paper_9804Shapiro.pdf)]  [[paper2](https://doi.org/10.5281/zenodo.6860879)]  [[thesis2](https://cs.pomona.edu/classes/cs190/thesis_examples/Shapiro.22.pdf)]  [[thesis1](https://scholarship.claremont.edu/pomona_theses/265/)]  [[recital](https://www.youtube.com/playlist?list=PLPwyBV1x-R8ZPdgk8EtO-Gh8l6Li0SH0j)]    [[paper1](https://scholarship.claremont.edu/cgi/viewcontent.cgi?article=1848&context=jhm)] |  |

**INDUSTRY EXPERIENCE**

## Research Intern, Microsoft (Research in Software Engineering/RiSE Group) *Summer 2025*

## Researching SMT parallelization algorithms and natural language reasoning for verifiable code generation. Supervised by Nikolaj Bjorner.

## Freelance Software Engineer, Stainless *Feb. 2023-Oct. 2024*

## Make open-source contributions to codebases such as Stoplight Prism, node-tree-sitter,

## Microsoft Pyright, NPM Trends, and json-schema-benchmark.

## Software Engineer, Meta *Oct. 2022-Nov. 2022*

* Improve type safety of Python and Hack code in engineering bootcamp. Impacted by the 13% company layoff as a new hire.

## 3x Software Engineering Intern, Facebook *Summers of 2019, 2020, 2021*

* iOS/serverside fullstack intern on Facebook Events and Groups.

**Talks**

“Synthesizing Composite Hierarchical Structure from Symbolic Music Corpora.” *The 19th SoCal Programming Languages and Systems Workshop (SoCaL PLS)*, Feb. 2025.

“Deriving Structure from Music Corpora.” *Programming Systems Group, UC San Diego*, Apr. 2024.

“MusAssist: A Domain Specific Language for Music Notation.” *International Conference on Technologies for Music Notation and Representation*, May 2023.

**TEACHING EXPERIENCE**

**Teaching Assistant, CSE 130: Programming Languages, UCSD** (N=69)*Fall**2025*

**Teaching Assistant, CSE 130: Programming Languages, UCSD** (N=126)*Spring**2025*

**Teaching Assistant, CS 133: Database Systems, Pomona College** (N=20)*Spring**2020*

**ACADEMIC HONORS**

## NSF Graduate Research Fellowship, Honorable Mention *2025*

## The NSF GRFP recognizes and supports outstanding graduate students who are pursuing full-time research-based master's and doctoral degrees in STEM fields.

## The Phi Beta Kappa Award *2022*

* Endowed by the Pomona Chapter of Phi Beta Kappa, awarded to one senior selected for high quality of scholarship and promise of future distinction.

## The Rena Gurley Archibald High Scholarship Prize *2022*

* Awarded to the member(s) of the Pomona College graduating class ranking highest in scholarship.

## Distinction in Senior Exercise (Computer Science) *2022*

* Exceptional work on the senior exercise is awarded based on review by the entire faculty of the Computer Science Department at Pomona College.

## Distinction in Senior Exercise (Music) *2022*

* Exceptional work on the senior exercise is awarded based on review by the entire faculty of the Music Department at Pomona College.

## The Katherine J. Hagedorn Prize *2022*

* Awarded annually to the Pomona College student(s) demonstrating exceptional loyalty and dedication to their music studies.

**Phi Beta Kappa Induction (Junior Year) - Pomona College Chapter** *2021*

* 1 of 8 juniors awarded for "good moral character," distinguish in "breadth of culture," and "excellence of scholarship."

**The William F. Russell Prize** *2020*

* Awarded annually to the Pomona College prospective music major(s) showing substantial accomplishment and significant promise in their study of music.

**SKILLS**

**Programming Languages:** Python ● Haskell ● TypeScript ● C++ ● Java ● Objective-C ● SQL

**Tools/Frameworks:** LaTeX ● Git ● Functional and Object-Oriented Programming

**Domain Knowledge:** Programming Languages ● Automated Reasoning ● SAT/SMT Solvers ●

Stochastic/Combinatorial Optimization ● Neurosymbolic Generation ● User Studies

**OUTREACH AND MENTORSHIP**

**Presenter, Harmony Hacks @ CSU San Marcos** *Spring**2025*

* NSF-funded event to broaden participation of women in computing. I co-hosted a Q&A to inspire high school girls to pursue careers in CS.

**Mentorship Co-Chair, GradWIC UCSD** *Fall 2024-Spring 2025*

* Manage the UCSD Graduate Women in Computing Mentorship Program. Pair 174 mentees with mentors, personally mentor 2 students, host inclusive group activities.