Alex (Oleksandr) Polozov

Staff Research Scientist, Google • polozov@google.com • +1 (425) 623-4121 • https://www.alexpolozov.com

I teach machines to write and analyze code, build language-driven systems to aid humans with procedural tasks, and research techniques for neuro-symbolic reasoning. Built and shipped tools for program synthesis from examples or natural language in multiple mass-market applications at Microsoft. Now working on the future of software at Google.

EDUCATION

Ph.D. in Computer Science   *2012-2017*

University of Washington, Seattle, USA

  Thesis: "A Framework for Mass-Market Inductive Program Synthesis"

  Advisors: Dr. Sumit Gulwani & Prof. Zoran Popović

B.Math. in System Analysis   *2008-2012*

National Technical University of Ukraine "Kyiv Polytechnic Institute"

  Thesis: "Structure and Term Prediction for Mathematical Text"

  Advisor: Prof. Yuriy Tymoshenko

EMPLOYMENT

Staff Research Scientist   *October 2021 – present*

Google, Mountain View, USA

X – the moonshot factory, Mountain View, USA

Principal Researcher   *August 2021 – September 2021*

Senior Researcher   *August 2017 – August 2021*

Microsoft Research, Redmond, USA

  Neural program synthesis from input-output examples and natural language, intersections of machine learning and formal methods, and neuro-symbolic reasoning in AI. Overseeing various product applications of neuro-symbolic reasoning for question answering, spreadsheet intelligence, software engineering, and data science.

Research & Software Engineer (part-time)   *October 2014 – April 2017*

Populus Group at Microsoft, Redmond, USA

  A founding member of the Program Synthesis using Examples (PROSE) R&D team at Microsoft.

  Developed the PROSE framework for automatic synthesis of domain-specific programs from incomplete specifications (e.g., input-output examples, constraints, demonstrations). It unified and generalized 12+ prior research publications, allowing one to develop an industrial-quality by-example technology in 10 weeks instead of 10 months. Its applications are deployed in multiple Microsoft products:

  - Excel FlashFill: string transformations by example;
  - Text and table extraction in PowerShell, Azure Data Factory, Azure Data Studio, PowerBI, PowerQuery;
  - Visual Studio IntelliCode: source code refactoring suggestions from repetitive edits.

Research Intern   *March 2014 – September 2014*

Microsoft Research, Redmond, USA

  Designed and developed a modular algorithmic framework for automatic synthesis of programs in domain-specific languages from inductive specifications. Generalizes five years of prior work in programming by examples done by the Sumit Gulwani's group and collaborators. This work became a foundation for the Microsoft PROSE team.

                                                                                            June 2013 – September 2013

  Designed a declarative language and an efficient interpreter for designing search strategies for microsegment queries based on linguistic predicates and semi-structured data on the Web.

                                                                                            June 2012 – September 2012

  Developed a language and an algorithm for data extraction from semi-structured images by example.

Software Development & Research Intern   *November 2011 – May 2012*

Yandex, Kyiv, Ukraine

  Developed a morphological engine with inflection prediction for Russian, Ukrainian, and English.

PUBLICATIONS

Books, book chapters, journal articles, monographs

[1] S. Chaudhuri, K. Ellis, O. Polozov, R. Singh, A. Solar-Lezama, and Y. Yue, "Neurosymbolic Programming", in *Foundations and Trends® in Programming Languages*, *7*(3), 158-243, 2021.

[2] S. Gulwani, O. Polozov, and R. Singh, "Program Synthesis", in *Foundations and Trends® in Programming Languages:* Vol. 4: No. 1-2, pp 1-119, 2017.

[3] O. Polozov. "A Framework for Mass-Market Inductive Program Synthesis." PhD dissertation, 2017.

Peer-reviewed conference publications

[4] A. Chowdhery, S. Narang, J. Devlin, M. Bosma, G. Mishra, A. Roberts, ... and N. Fiedel. "PaLM: Scaling Language Modeling with Pathways", 2022.

[5] G. Poesia., O. Polozov., V. Le, A. Tiwari, G. Soares, C. Meek, and S. Gulwani, "Synchromesh: Reliable code generation from pre-trained language models", in *10th International Conference on Learning Representations (ICLR),* 2022.

[6] T. Schuster, A. Kalyan, O. Polozov, and A. T. Kalai. "Programming Puzzles", in *35th Conference on Neural Information Processing Systems (NeurIPS), Datasets & Benchmarks Track*, 2021.

[7] C.-H. Lee, O. Polozov, and M. Richardson. "KaggleDBQA: Realistic Evaluation of Text-to-SQL Parsers", in *59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP),* 2021.

[8] X. Deng, A. H. Awadallah, C. Meek, O. Polozov, H. Sun, and M. Richardson. "Structure-Grounded Pretraining for Text-to-SQL", in the *Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT),* 2021.

[9] T. Yu, R. Zhang, O. Polozov, C. Meek, and A. H. Awadallah. "SCoRe: Pre-Training for Context Representation in Conversational Semantic Parsing", in *9th International Conference on Learning Representations (ICLR),* 2021.

[10] S. Amizadeh, H. Palangi, O. Polozov, Y. Huang, and K. Koishida. "Neuro-Symbolic Visual Reasoning: Disentangling "Visual" from "Reasoning"", in *37th International Conference on Machine Learning (ICML),* 2020.

[11] S. Srivastava, O. Polozov, N. Jojic, and C. Meek. "Learning Web-Based Procedures by Reasoning over Explanations and Demonstrations in Context", in *58th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2020.

[12] B. Wang, R. Shin, X. Liu, O. Polozov, and M. Richardson. "RAT-SQL: Relation-Aware Schema Encoding and Linking for Text-to-SQL Parsers", in *58th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2020.

[13] R. Shin, M. Allamanis, M. Brockschmidt, and O. Polozov. "Program Synthesis and Semantic Parsing with Learned Code Idioms", in *33rd Conference on Neural Information Processing Systems (NeurIPS),* 2019.

[14] M. Brockschmidt, M. Allamanis, A. L. Gaunt, and O. Polozov. "Generative Code Modeling with Graphs", in *7th International Conference on Learning Representations (ICLR),* 2019.

[15] S. Padhi, P. Jain, D. Perelman, O. Polozov, S. Gulwani, and T. Millstein, "FlashProfile: A Framework for Synthesizing Data Profiles", in *ACM SIGPLAN conference on Systems, Programming, Languages and Applications: Software for Humanity (SPLASH),* 2018.

[16] A. Kalyan, A. Mohta, O. Polozov, D. Batra, P. Jain, and S. Gulwani, "Neural-Guided Deductive Search for Real-Time Program Synthesis from Examples", in *6th International Conference on Learning Representations (ICLR),* 2018.

[17] R. Rolim, G. Soares, L. D'Antoni, O. Polozov, S. Gulwani, R. Gheyi, R. Suzuki, and B. Hartmann, "Learning syntactic program transformations from examples", in *39th International Conference on Software Engineering (ICSE),* 2017.

[18] M. Mayer, G. Soares, M. Grechkin, V. Le, M. Marron, O. Polozov, R. Singh, B. Zorn, and S. Gulwani, "User interaction models for disambiguation in programming by example", in *ACM Symposium on User Interface Software and Technology (UIST)*, 2015.

[19] O. Polozov and S. Gulwani, "FlashMeta: a framework for inductive program synthesis", in *ACM Conference on Object-Oriented Programming Systems, Languages, and Applications (OOPSLA)*, 2015.

[20] O. Polozov, E. O'Rourke, A. M. Smith, L. Zettlemoyer, S. Gulwani, and Z. Popović, "Personalized mathematical word problem generation", in *International Joint Conference on Artificial Intelligence (IJCAI)*, 2015.

[21] O. Polozov and S. Gulwani, "LaSEWeb: automating search strategies over semi-structured web data", in *ACM Conference on Knowledge Discovery and Data Mining (KDD)*, 2014.

Workshop publications

[22] V. Hellendoorn, P. Devanbu, O. Polozov, and M. Marron. "Learning to Infer Run-Time Invariants from Source Code", in *Computer-Assisted Programming (CAP) Workshop @ NeurIPS,* 2020.

[23] L. Yeung, Y. Bisk, and O. Polozov. "ALFRED Speaks: Automatic Instruction Generation for Egocentric Skill Learning", in *Embodied Vision, Actions & Language Workshop @ ECCV,* 2020.

[24] C. Wang, P.-S. Huang, O. Polozov, M. Brockschmidt, and R. Singh, "Execution-Guided Neural Program Decoding", in *2nd Workshop on Neural Abstract Machines & Program Induction (NAMPI)*, 2018.

[25] O. Polozov and S. Gulwani, "Program synthesis in the industrial world: inductive, incremental, interactive", in *5th Workshop on Program Synthesis (SYNT)*, 2016.

Preprints, technical reports, papers under review

[26] A. Ni, J. P. Inala, C. Wang, O. Polozov, C. Meek, D. Radev, and J. Gao. "Learning from Self-Sampled Correct and Partially-Correct Programs". *arXiv preprint arXiv:2205.14318,* 2022.

[27] V. J. Hellendoorn, P. T. Devanbu, O. Polozov, and M. Marron. "Are My Invariants Valid? A Learning Approach". *arXiv preprint arXiv: 1903.06089,* 2019.

[28] T. Shi, K. Tatwawadi, K. Chakrabarti, Y. Mao, O. Polozov, and W. Chen. "IncSQL: Training Incremental Text-to-SQL Parsers with Non-Deterministic Oracles". *arXiv preprint arXiv: 1809.05054,* 2018.

[29] C. Wang, K. Tatwawadi, M. Brockschmidt, P.-S. Huang, Y. Mao, O. Polozov, and R. Singh. "Robust Text-to-SQL Generation with Execution-Guided Decoding". *arXiv preprint arXiv:1807.03100,* 2018.

[30] V. Le, D. Perelman, O. Polozov, M. Raza, A. Udupa, and S. Gulwani, "Interactive Program Synthesis". *arXiv preprint arXiv:1703.03539,* 2017.

[31] O. Polozov, S. Gulwani, and S. Rajamani, "Structure and term prediction for mathematical text". Tech. Rep. MSR-TR-2012-7, 2012.

INVITED TALKS

- **Conversations with Data**
  Invited talk at the *IntEx-SemPar: Interactive and Executable Semantic Parsing Workshop at EMNLP 2020*.
- **Neuro-Symbolic Program Synthesis from Natural Language and Demonstrations**
  Invited talk at the *9th Workshop on Synthesis (SYNT), 2020.*
- **From Examples to Natural Language and Back**
  "State of the Art in Program Synthesis" Workshop, 2019.
- **Program Understanding, Synthesis, and Generalization with Graph Neural Networks**
  Invited talk, *Learning and Reasoning with Graph-Structured Representations Workshop at ICML 2019.*
- **Program Synthesis via Neural-Guided Deductive Search**
  Invited talk at the *Machine Learning + Programming Languages Workshop 2018.*
- **Bringing Program Synthesis to the Mass Markets**
  Approaches and Applications of Inductive Programming, Dagstuhl Seminar 2017.
- **Data Processing Using Input-Output Examples with Microsoft PROSE SDK**
  Creating Programming-By-Example Features in Arbitrary Domains with Microsoft PROSE SDK
  A series of tutorials hosted at the *Machine Learning & Data Science Conference (MLADS) 2017*.
- **PROSE: Inductive Program Synthesis for the Mass Markets**
  Invited talk & Hackathon for graduate students at *UC Berkeley, January 2017.*
- **Automated Program Synthesis**
  Invited talk at the *Human-Like Computing Machine Intelligence Workshop (MI20-HLC), October 2016.*

- **PROSE: Programming using Examples**
  Co-lectured with Sumit Gulwani. Invited tutorial at the *ACM SIGPLAN Conference on Programming Languages Design & Implementation (PLDI) 2016.*
- **Programming by Examples**
  Co-lectured with Sumit Gulwani at *Marktoberdorf Summer School 2015*.
- **Personalized Mathematical Word Problem Generation**
  Approaches and Applications of Inductive Programming, Dagstuhl Seminar 2013.

NOTABLE WRITING

- [Github Copilot – Technical Q&A](#) – Twitter, June 2021.
- [Conversations with data: Advancing the state of the art in language-driven data exploration](#) – Microsoft Research Blog, May 2021.
- [Program Synthesis in 2017-18](#) – Alex Polozov, July 2018.
- [Neural-Guided Deductive Search: A best of both worlds approach to program synthesis](#) – Microsoft Research Blog, April 2018.
- S. Gulwani, O. Polozov, and R. Singh, ["Program Synthesis",](#) in *Foundations and Trends® in Programming Languages:* Vol. 4: No. 1-2, pp 1-119, 2017.

ACADEMIC SERVICE, PC, REVIEWS

IJCAI-ECAI 2018, GPCE 2018, ML4PL 2018, ESEC/FSE SRC 2018, AAAI 2019, ICML 2019, IJCAI 2019, NeurIPS 2019, ACL Demos 2020, ICLR 2020, NeurIPS 2020, ICLR 2021, ICML 2021, ACL Demos 2021, NeurIPS 2021.

TEACHING EXPERIENCE

CSEP 590C: Domain-Specific Languages   *Spring 2016, Spring 2017*
University of Washington, Paul G. Allen School for Computer Science & Engineering

Together with Prof. Rastislav Bodik and Pavel Panchekha, co-designed and taught a graduate course on DSLs. Audience: professional software engineers with multiple years of industry experience. Course content includes foundations of compiler/interpreter development, a collection of well-known DSLs (D3.js, Mustache, Hadoop, React.js, Rx), program synthesis in PROSE, and lessons on DSL design.

PATENTS

[1] Zorn, B.G., Brockschmidt, M.M.J., Choudhury, P., Polozov, O., Singh, R. and Padhi, S., Microsoft Technology Licensing LLC, 2020. *Systems, methods, and computer-readable media for improved table identification using a neural network*. U.S. Patent Application 16/034,447.

[2] Polozov, O., Gulwani, S., Jain, P., Vijayakumar, A. K. and Mohta, A., Microsoft Technology Licensing LLC, 2019. *Neural-guided deductive search for program synthesis.* U.S. Patent Application 16/019,280.

[3] Gulwani, S., Jain, P., Perelman, D.A., Padhi, S. and Polozov, O., Microsoft Technology Licensing LLC, 2019. *Syntactic profiling of alphanumeric strings*. U.S. Patent Application 15/663,575.

[4] Gulwani, S., Zorn, B.G., Singh, R., Marron, M., Polozov, O., Le, V.M., Mayer, M., Soares, G.A. and Grechkin, M., Microsoft Technology Licensing LLC, 2017. *User interaction models for disambiguation in programming-by-example*. U.S. Patent 9,891,895.

[5] Gulwani, S., Polozov, O. and Tiwary, S.K., Microsoft Technology Licensing LLC, 2015. *Control of automated tasks executed over search engine results*. U.S. Patent 10,644,530.

[6] Gulwani, S. and Polozov, O., Microsoft Technology Licensing LLC, 2015. *Parsing and rendering structured images*. U.S. Patent 9,031,894.