

## Overview and Research Interests

(As of February 27, 2023) I'm a senior researcher in the PROSE team at Microsoft, where I work on developing state-of-the-art program synthesis technologies to make writing and using software easier and more enjoyable. Prior to this, I graduated from MIT with a PhD in Computer Science, advised by Martin Rinard.

## Education

### Academic Qualifications

- 2016-2021 **PhD in Computer Science**  
*Massachusetts Institute of Technology, Cambridge, MA.*
- 2013-2016 **Masters in Computer Science**  
*New York University: Courant Institute of Mathematical Sciences, NY, NY.*  
*GPA: 3.89, MS Research/Thesis Fellowship Award Fall 2015, funding work on A2Q (an order-aware optimizing query compiler for AQuery)*
- 2007-2011 **Bachelor of Arts in Economics and Minor in German Studies**  
*University of Pennsylvania, Philadelphia, PA.*  
*GPA: 3.93, Phi Beta Kappa, Summa Cum Laude, Dean's List (08, 09, 10)*

## Industry Work Experience

- June 2022 to date **Senior Researcher PROSE Team**, Microsoft, Remote.  
◦ Working on program synthesis technologies for a variety of developer, data scientist, and end-user applications. A lot of my work focuses on developing and applying large language models to programming tasks, such as program repair and natural language to code synthesis. Some of our recent work on custom LLMs has been featured in [news outlets](#).
- July 2021 to June 2022 **Researcher PROSE Team**, Microsoft, Remote.
- Summer 2020 **Intern Facebook AI Research**, Facebook, Remote (due to COVID-19).  
◦ Worked with the SysML team on a novel tensor compiler, writing C++ for JIT compilation, benchmarking against Halide/TVM
- Fall 2018 **Part-Time Research Collaborator Big Code Team**, Facebook, Remote.  
◦ Applying deep learning to identify and highlight core code functionality
- Summer 2018 **Intern Software Engineering**, Facebook, Boston.  
◦ Applying deep learning to code search Worked with the Big Code team on applications of neural networks to code search
- Summer 2015 **Intern Data Science**, Cloudera, San Francisco.
- 2011 – 2014 **Full-Time Securitized Credit Research Associate Non-Agency Mortgages and US Housing**, Morgan Stanley, New York.
- Summer 2010 **Richard B. Fisher Scholar Fixed Income Generalist Sales and Fixed Income Credit Strategy**, Morgan Stanley, New York.

Summer 2009 **Douglas Paul Scholar** *Investment Banking and Alternative Investments*, Morgan Stanley, New York.

---

## Academic Work Experience

Fall and Spring 2021 **Advanced Undergraduate Research Class** TA, MIT.

2015 – 2016 **Graduate Course in Compiler Construction** Grader, NYU.

Fall 2014 **Graduate Course in Programming Languages** Teaching Assistant, NYU.

---

## Publications

- [1] Jose Cambronero, John Feser, Micah Smith, and Samuel Madden. Query optimization for dynamic imputation. *PVLDB*, 10(11):1310–1321, 2017.
- [2] José P. Cambronero, Jiasi Shen, Jürgen Cito, Elena Glassman, and Martin Rinard. Characterizing Developer Use of Automatically Generated Patches. In *VL/HCC (Short Paper)*, 2019.
- [3] José P. Cambronero, Hongyu Li, Seohyun Kim, Koushik Sen, and Satish Chandra. When Deep Learning Met Code Search. In *FSE (Industry Track)*, 2019.
- [4] José P. Cambronero, Thurston H.Y. Dang, Nikos Vasilakis, Jiasi Shen, Jerry Wu, and Martin Rinard. Active Learning for Software Engineering. In *SPLASH Onward!*, 2019.
- [5] José P. Cambronero and Martin Rinard. AL: Autogenerating Supervised Learning Programs. In *SPLASH OOPSLA*, 2019.
- [6] José P. Cambronero, Jürgen Cito, and Martin Rinard. AMS: Generating AutoML Search Spaces from Weak Specifications. In *ESEC/FSE*, 2020.
- [7] Limor Appelbaum, José P. Cambronero, and et al. Development and Validation of a Pancreatic Cancer Risk Model for the General Population Using Electronic Health Records: An Observational Study. In *European Journal of Cancer*, 2020.
- [8] Fatjon Zogaj, José Pablo Cambronero, Martin C Rinard, and Jürgen Cito. Doing more with less: characterizing dataset downsampling for automl. *Proceedings of the VLDB Endowment*, 14(11):2059–2072, 2021.
- [9] José P. Cambronero, Micah Smith, Jürgen Cito, and Martin Rinard. Learning Repair Rules for Machine Learning Pipelines from AutoML Search Traces. In *Under submission*, 2020.
- [10] José P. Cambronero, Raul Castro Fernandez, and Martin Rinard. wranglesearch: Mining Data Wrangling Functions from Python Programs. In *Under submission*, 2021.
- [11] Rohan Bavishi, Harshit Joshi, José Cambronero, Anna Fariha, Sumit Gulwani, Vu Le, Ivan Radiček, and Ashish Tiwari. Neurosymbolic repair for low-code formula languages. *Proc. ACM Program. Lang.*, 6(OOPSLA2), oct 2022.
- [12] José Cambronero, Sumit Gulwani, Vu Le, Daniel Perelman, Arjun Radhakrishna, Clint Simon, and Ashish Tiwari. Flashfill++: Scaling programming by example by cutting to the chase. *Proc. ACM Program. Lang.*, 7(POPL), jan 2023.
- [13] Harshit Joshi, José Cambronero, Sumit Gulwani, Vu Le, Ivan Radicek, and Gust Verbruggen. Repair is nearly generation: Multilingual program repair with llms. *arXiv preprint arXiv:2208.11640 (AAAI 2023)*, 2022.
- [14] Bram Wasti, José Pablo Cambronero, Benoit Steiner, Hugh Leather, and Aleksandar Zlateski. Loopstack: a lightweight tensor algebra compiler stack. *arXiv preprint arXiv:2205.00618*, 2022.

- [15] Harshit Joshi, Abishai Ebenezer, José Cambronero, Sumit Gulwani, Aditya Kanade, Vu Le, Ivan Radiček, and Gust Verbruggen. Flame: A small language model for spreadsheet formulas. *arXiv preprint arXiv:2301.13779 (under submission)*, 2023.
- [16] Mukul Singh, José Cambronero, Sumit Gulwani, Vu Le, Carina Negreanu, Mohammad Raza, and Gust Verbruggen. Cornet: A neurosymbolic approach to learning conditional table formatting rules by example. *arXiv preprint arXiv:2208.06032 (under submission)*, 2022.
- [17] Jialu Zhang, José Cambronero, Sumit Gulwani, Vu Le, Ruzica Piskac, Gustavo Soares, and Gust Verbruggen. Repairing bugs in python assignments using large language models. *arXiv preprint arXiv:2209.14876 (under submission)*, 2022.

## Language skills

- **Programming Languages:** Proficient in: Python, Javascript/Typescript, R, C#.
- **Natural Languages:** Native fluency in English and Spanish. Working proficiency in German.

## Service

- **Artifact Evaluation Committee OOPSLA 2020, CAV 2020, PPOPP 2018**

## Mentoring/Advising

- Harshit Joshi (Microsoft): research fellow advisor
- Jialu Zhang (Yale/Microsoft): summer 2022 internship advisor
- Jennifer McCleary (MIT) MEng Thesis: pancreatic cancer risk modeling (Fall 2019 - January 2020)
- Alex Berg (MIT) Undergraduate research: pancreatic cancer risk modeling (Summer 2020)
- Thomas Xiong (MIT) MEng Thesis: pancreatic cancer risk modeling (Fall 2020 - Spring 2021)
- Lori Zhang (MIT) Undergraduate research: pancreatic cancer risk modeling (Summer 2020 - Spring 2021)
- Harshit Joshi (Microsoft): Research fellow in the PROSE team on automated program repair (Fall 2021 to date)
- Mukul Singh (Microsoft): Research fellow in the PROSE team working on NL-to-Code (Spring 2022 to date)
- Abishai Ebenezer (Microsoft): Research fellow in the PROSE team working on automated program repair (Fall 2022 to date)