

515.508.1003
philip@kiely.xyz
San Francisco, CA

PHILIP J. KIELY
<https://philipkiely.com>

WORK EXPERIENCE

BASETEN

January 2022–Present

Founding Developer Advocate. Own developer relations, technical content, and conference speaking from pre-revenue through Series D. 100+ technical blog posts, dozens of conference talks and counting. Define DevRel function and hire team. Build integrations for popular tools like LangChain. Work closely with GTM on sales enablement and lead generation. Represent Baseten at conferences as a speaker or at the expo. Previously owned documentation and published over 100,000 words of documentation in two years.

Author of *Inference Engineering*, a 250-page technical book on inference covering model architectures, hardware options, software stacks, model performance optimization techniques, modality-specific inference considerations, and multi-cloud infrastructure (published 2026).

Selected speaking:

- [AI Engineer World's Fair 2025](#)
- [NVIDIA GTC 2025](#)
- [PyTorch Conference 2025](#)

Selected podcasts:

- [Software Huddle \(2024\)](#)
- [Software Engineering Radio EP 697](#)
- [Weaviate Podcast EP 105](#)

Selected writing:

- <https://www.baseten.com/blog/sota-performance-for-gpt-oss-120b-on-nvidia-gpus/>
- <https://www.baseten.com/blog/how-we-built-bei-high-throughput-embedding-inference/>
- <https://www.baseten.com/blog/how-multi-node-inference-works-llms-deepseek-r1/>
- <https://www.baseten.co/blog/llm-transformer-inference-guide/>

FREELANCE

April 2019–December 2021

Independent. Author of [Writing for Software Developers](#) (over 1000 copies sold). Freelance technical writer published by Digital Ocean, Twilio, Smashing Magazine, CSS Tricks, and more.

GUMROAD

September 2020–January 2021

Marketing. Operations across content, support, and growth plus projects including Gumroad Memberships launch, creating Gumroad University, and individual correspondence with top creators to compile feedback, recognize success, and grow their businesses. Led team of eight.

MARKFORGED

Summer 2019, Summer 2020

Software Engineer (2020). Software Engineering Intern (2019). Full-stack AngularJS and Electron developer working on 3D printer fleet management software.

Launched the company's first open-source software project ([GOMP](#)) and new engineering blog.

EDUCATION

GRINNELL COLLEGE

B.A. with Honors in Computer Science

Class of 2020
Grinnell, IA