Andrew W. Hill PhD

PhD ecologist · Scientific data and AI evaluation · Co-founder and CEO for 10 years

Sausalito, CA · andrewxhill@gmail.com · linkedin.com/in/andrewxhill · andrewxhill.com · github.com/andrewxhill · Google Scholar

I am a PhD ecologist who moved from science into AI. In science I built georeferencing pipelines for museum species records and designed the first architecture of Map of Life. I co-authored GeoCAT, a tool for IUCN Red List threat assessments. From 2025 to 2026 I led AI research teams at We Are Set that built an eval system to rank LLMs on each type of task. In 2026 I listed 19 ways AI-run scientific studies go wrong.

6,000+
citations of my peer-reviewed research

51 models
ranked in Beam, my LLM benchmark

450,000+
species in Map of Life, whose first architecture I designed

2 to 25+
research, product and engineering team I grew

14
peer-reviewed papers, 6 as first author

EXPERIENCE

Co-founder and CEO
Nov 2016 – Nov 2026

We Are Set, Inc. · Sausalito and remote

- Grew a research, product and engineering team from 2 to 25+ over ten years. Set the research and product roadmaps, designed the systems and wrote much of the code.
- **AI research and evaluation (2025 – 2026).** Led AI research teams. We built an eval system that ranks LLMs on each type of task, and a router that sends each request to the top model.
- Built Beam, an LLM benchmark where 51 models compete in pairs on 45 scenarios, scored by 3 LLM judges. Ran Model Arena, public evaluations of frontier models.
- **Models and research (2026).** Edited a small vision model, with no training run, to name Galápagos species in camera-trap photos (balanced accuracy 64% to 82%), and published four models, the code and the method. Wrote up 19 ways AI-run scientific studies go wrong.
- **Open protocols (2018 – 2024).** Pitched and negotiated about $3M in Protocol Labs grants to research and build open data protocols. 100,000 developers used Textile, our developer platform for user-owned data.
- Raised venture funding from about 30 institutional investors. Closed the company in 2026 with a board-approved wind-down.

Chief Science Officer, executive team
2011 – 2016

CartoDB (now CARTO) · Brooklyn and Madrid

- Took CartoDB, with its founders, from a project inside Vizzuality to a venture-backed SaaS company for mapping and spatial analysis. Governments, banks, newsrooms and NGOs used it.
- Led research for the company. Pioneered large-scale time-series mapping on the web with Torque, an open source library that animates millions of time-stamped points.
- Invented and designed the Data Observatory, CARTO's catalogue of location data, and led the team that built it.

Associate Professor, by invitation · NYU Tisch, Interactive Telecommunications Program
2014 – 2015

Designed and taught Maps, Lies, and Storytelling, a graduate course on bias and distortion in maps.

Senior Scientist · Vizzuality, New York
2011 – 2012

Worked between outside science teams and Vizzuality's engineers to design data projects, and wrote grant proposals.

RESEARCH

PhD, Ecology and Evolutionary Biology, University of Colorado Boulder, 2012. Biodiversity informatics with Robert Guralnick. Assembled global biodiversity data to measure the effects of environmental change. **MA, Ecology and Evolutionary Biology**, 2008. Research on H5N1 bird flu, using flu genomes shared openly before publication.

- Location, location, location: utilizing pipelines and services to more effectively georeference the world's biodiversity data. First author. *BMC Bioinformatics*, 2009.
- The Notes from Nature tool for unlocking biodiversity records from museum records through citizen science. First author. *ZooKeys*, 2012. Volunteers transcribed 4.7M+ specimens by 2023.
- Supporting Red List threat assessments with GeoCAT: geospatial conservation assessment tool. *ZooKeys*, 2011.

WRITING AND SKILLS

Writing: "Startups Went All-In on AI. Why Hasn't Conservation?" (November 2026) · AI: LLM evaluation, benchmark design, judge-based ranking, failure analysis of AI agents · Science: ecology, biodiversity informatics, research design, statistics, data quality · Data: Python, SQL, data pipelines, species and spatial data