Nikos Bosse

Research engineer specializing in probabilistic forecasting and its evaluation. PhD on infectious disease forecasting and forecast evaluation, creator of the forecast-evaluation package scoringutils, now building and benchmarking LLM forecasting agents at FutureSearch. Top-ranked forecaster in the DARPA Replication Markets prediction market.

Berlin, Germany

+49 176 8016 5076
nikosbosse@gmail.com

github.com/nikosbosse
linkedin.com/in/nikosbosse

💼 EXPERIENCE

FutureSearch — Research Engineer
San Francisco, US / Remote. 04/2024 – present

- Build and evaluate LLM-based forecasting and web-research agents
- Published the first pipeline to generate and resolve high quality real-world forecasting questions at scale with LLM research agents
- Co-developed Bench to the Future, a hermetic pastcasting backtest suite, allowing for rapid evaluation and iteration on AI forecasting agents

Metaculus — Research Coordinator
Delaware, US / Remote. 12/2021 – 05/2024

Developed Metaculus' research strategy and coordinated with researchers and external partners.

DebateConsult — Rhetoric and Communications Trainer
Germany. 2016 – 2024

Rhetoric and communication training for clients including works-council chairs of leading German companies.

🎓 EDUCATION

London School of Hygiene & Tropical Medicine — PhD
London, UK. 09/2020 – 05/2024

Thesis on infectious-disease forecasting: evaluation of probabilistic forecasts (proper scoring rules) and of human-judgment forecasts.

Universität Göttingen — Medicine (MD)
Göttingen, Germany. 2015 – 2022

Universität Göttingen — MSc Applied Statistics (minor: Computer Science)
Göttingen, Germany. 2017 – 2020

GPA 4.0; Florenz Sartorius Award for the best graduate of the year. Coursework incl. nonparametric methods, generalized linear models, machine learning, deep learning, parallel computing.

Universität Tübingen — BSc International Business Administration
Tübingen, Germany. 2011 – 2015

GPA 3.9; Deutschlandstipendium scholarship. Coursework incl. statistics, finance, micro- and macroeconomics; thesis in empirical economics.

SKILLS

Probabilistic forecasting & forecast evaluation

Statistical modeling (GLMs, nonparametrics, time series)

Machine learning, LLM agents & evals

Public speaking & negotiation

AWARDS

Jacob-Henle Scholarship

ACX Grant
For the development of a Forecasting Wiki

Forecast Innovation Prize
Awarded for the creation of a novel forecasting app

Outstanding Ideas Award by the Universität Göttingen

PROGRAMMING

Python · R (CRAN package author) · Bash · Git

LANGUAGES

German, English (both C2) · French (C1) · Spanish (B2)

🖥️ FORECASTING & PROJECTS

Forecasting & Prediction Markets — Track record

- DARPA Replication Markets (prediction market on study replicability): ranked 1st in 3 of 7 rounds; 3rd, 5th and 11th in the others
- 4th of 1,047 participants in a prediction tournament on withkindred.com
- Participated in the Existential Risk Persuasion Tournament (Philip Tetlock, Forecasting Research Institute)

scoringutils ↗ — Creator and Project Lead

2019 – present

R package for evaluating probabilistic forecasts (proper scoring rules, model comparison, visualization), used, among others, by the US CDC, the European Centre for Disease Prevention and Control and Médecins Sans Frontières.

crowdforecastr — Creator and Project Lead

2020 – 2021

- Built an R Shiny app to elicit COVID-19 forecasts from human forecasters
- Ran a twelve-week forecasting tournament with Public Health England to inform COVID-19 decision-making in the UK

University Debating — 50+ national and international tournaments

- Vice World Champion (English as a foreign language), World Universities Debating Championship 2014; South German Champion 2014 & 2015

📄 SELECTED PUBLICATIONS

Google Scholar ↗ · h-index 34 · 17,778 citations

Evaluating Strategic Reasoning in Forecasting Agents ↗
Tom Liptay, Dan Schwarz, Rafael Poyiadzi, Jack Wildman, **Nikos I. Bosse**, arXiv, 2026

Automating Forecasting Question Generation and Resolution for AI Evaluation ↗
Nikos I. Bosse, [...], Dan Schwarz, arXiv, 2026

Deep Research Bench: Evaluating AI Web Research Agents ↗
Nikos I. Bosse, [...], Jack Wildman, arXiv, 2025

Bench to the Future: A Pastcasting Benchmark for Forecasting Agents ↗
Jack Wildman, **Nikos I. Bosse**, [...], Lawrence Phillips, arXiv, 2025

Scoring Epidemiological Forecasts on Transformed Scales ↗
Nikos I. Bosse, [...], Sebastian Funk, PLOS Computational Biology, 2023

Evaluating Forecasts with scoringutils in R ↗
Nikos I. Bosse, [...], Sam Abbott, arXiv, 2022