COMPANY PROFILE

Somast

Somast is a research company building Somali language data, models and training.

At a glance

Name	Somast
Company type	Private limited liability company (LLC)
Name means	Somali Science and Technology
Founded	2023
Headquarters	Mogadishu, Somalia
Focus	Somali language data, models and education
Languages	Somali and English
Website	www.somast.so
Email	info@somast.so
Governing law	Federal Republic of Somalia
Registration	44052647694, Somali Government
Address	Mogadishu, Somalia
Phone	+252619708582

Numbers

20M+ people speak Somali
Estimate used across our materials (Oct 1, 2026)

3 open datasets
Published on our datasets page (Oct 1, 2026)

1.1B+ corpus tokens
SomNLP corpus from 17 sources (Sep 24, 2026)

5 bootcamp cohorts
Delivered since September 2025 (Oct 1, 2026)

200+ bootcamp alumni
People who took part in a cohort (Oct 1, 2026)

269+ open-source contributors
Unique GitHub contributors across our repositories (Oct 14, 2026)

Who we are

Somali is spoken by more than 20 million people, yet it is almost absent from the data that trains modern AI. Somast collects and publishes Somali text and speech, trains Somali models on it, and teaches people to build with it. The company was founded in 2023 and works from Mogadishu.

Our mission
AI infrastructure for the Somali language, built with Somali speakers.

Our values

Documented and shareable: Datasets, model details, and evaluation methods are documented so others can verify and build on our releases.
Somali-led: Built with Somali speakers, who shape the technology and its direction.
Rigorous & reproducible: Clear benchmarks, documented methods, and results others can reproduce and extend.

What we do

Data
Somali text and speech corpora with documented sources and licences.

Models
A Somali tokenizer, speech-to-text checkpoints and translation previews.

Education
Sponsored bootcamps and mentored learning tracks for Somali engineers and analysts.

Technology consulting
Software and advisory work for partner organisations, including websites and mobile apps.

History

2023 Somast founded
Started in Mogadishu to build the datasets, models and tools Somali needs for AI.

2023 Soplang v1 released
First release of Soplang, a programming language with Somali syntax.

2023 Open education started
Began teaching data, programming and machine learning to Somali learners on YouTube.

2024 First open corpora released
Published the first Somali text and speech datasets under open licences.

2024 Partnerships launched
Started working with organisations that use or support Somali language technology.

2025 Soplang v2 released
Second major release of Soplang, the Somali-syntax programming language.

2025 Talent programs launched
Launched bootcamps and fellowships that train local engineers in machine learning and NLP.

2026 Data and machine learning services
Custom data and machine learning work: we prepare data and build models for your needs.

Services

Build software
We design and build websites, web apps, mobile apps and internal tools for organisations in Somalia and beyond.

Data and AI
We turn your data into reports, dashboards and models, and build AI that reads, listens and answers in Somali.

Maintain and support
After launch we keep your software working, secure and up to date, and we improve it as your needs change.

Hosting, domain and email included
While you have a maintenance agreement with us, we provide your domain, hosting and business email at no extra charge.

Roadmap

2026 Release SomNLP-Corpus
: Full release of our Somali text corpus, with documentation and the licence of each source.

2026 Release Soplang v3
: Third major release of Soplang, the Somali-syntax programming language.

2027 Launch mentorship
: A mentorship program that connects learners with experienced practitioners.

Programmes

Git & GitHub Bootcamp · July 2026
Version control from first principles, commits, branching, merge conflicts, pull requests, and the collaboration workflow every other bootcamp assumes you already know.

DS & ML Bootcamp - Jun 2026 · June 2026
Hands-on Data Science and Machine Learning cohort with real datasets, model building, and capstone projects. Our largest cohort to date.

DS & ML Bootcamp - Feb 2026 · February 2026
Second Data Science and Machine Learning cohort. Covered statistics, sklearn, deep learning fundamentals, and real-world Somali datasets.

Python for Everyone · April 2026
Beginner-to-intermediate Python bootcamp open to all. Our most-starred repository, stars and forks from learners around the world.

DS & ML Bootcamp - Sep 2025 · September 2025
Our first full Data Science and Machine Learning cohort. Laid the foundation for all future programs with a curriculum now used by dozens of learners.

Leadership and team

Ismail Ainte · CEO
Leading Somast's strategy, operations, and growth.

Sharafdin Yusuf · Lead Engineer & Researcher
Leading the design and development of Somast' technology platforms and research.

Omar Tood · Lead Instructor & Researcher
Data pipelines, machine learning models, and technical training programs.

Salah Sharafdin · Software Developer
Building web applications, user-facing features, and internal tools.

Partners

Unkad: Research partner of Somast.
Dugsiiye: Sponsored every Somast bootcamp, all open to anyone.
Memvid: Somast was Memvid's lead open-source software engineer.
Soplang: A separate open-source project under Somast: a programming language with Somali syntax.
Docker: Sponsors Soplang, our open-source Somali-syntax programming language, through the Docker-Sponsored Open Source program.
Skilful: Somast is Skilful's technology partner, providing consulting and software development.

Casri Store: Somast is Casri Store's technology partner and consultant, and built its website and mobile app.
Youth Connect: Somast is Youth Connect's technology partner and consultant, and developed its website.

Funding

Somast is funded by three kinds of support.

- Sponsorship: Dugsiiye has sponsored every Somast bootcamp, all of them open to anyone.
- Client work: technology consulting and software development for partner organisations.
- Open-source sponsorship: Docker supports Soplang through the Docker-Sponsored Open Source program.

Contact

Press and general: info@somast.so
Partnerships: www.somast.so/en/become-partner
www.somast.so/en/company-profile