Frontier AI, under your control.

Endeavor 1.0

Frontier-class reasoning and coding for long-horizon agent work, with the freedom to run it through Flower or on infrastructure you control.

92.0	98.2	99.9	94.1
GPQA	HumanEval	AIME 2026	IFEval

Endeavor 1.0 is competitive with leading models from OpenAI and Anthropic, while giving you the choice to use it through Flower or deploy it on infrastructure you control. Build a lasting AI capability around Endeavor 1.0.

Frontier capability.
A foundation you can build on.

Endeavor 1.0 combines the performance of a leading frontier model with the freedom to choose where it runs, how you build on it and how you improve it around your work.

Frontier performance

Strong across reasoning and coding, with the breadth and tool use required for long-horizon agent work. On key evaluations, Endeavor 1.0 matches or exceeds leading closed and open-weight models.

Choose where it runs

Use Endeavor 1.0 through a Flower-managed service or infrastructure you control.

Start quickly through an API while retaining a path to private deployment for selected workloads or your entire deployment.

A foundation you can invest in

Develop agents, evals, data pipelines and improvement loops around Endeavor 1.0. Because it is not tied to a single closed API, that work can become a durable capability for your organisation.

Run a stable deployment, choose when to adopt upgrades and retain the option to operate Endeavor 1.0 independently of any single provider.

Frontier performance

Competitive with the frontier

Endeavor 1.0 is evaluated directly against leading models from OpenAI and Anthropic, as well as the strongest open-source alternatives, including Kimi K3 and Nemotron 3 Ultra.

Across the four completed core evaluations shown here, Endeavor 1.0 records the highest HumanEval score; matches GPT-5.6 Sol and Claude Fable 5 on AIME 2026; outperforms Kimi K3 on three of four benchmarks; and outperforms Nemotron 3 Ultra on all four.

98.2

HumanEval

The highest score in the comparison.

99.9

AIME 2026

Matching the strongest OpenAI and Anthropic models evaluated.

3/4

vs Kimi K3

Benchmarks ahead of Kimi K3.

4/4

vs Nemotron

Benchmarks ahead of Nemotron 3 Ultra.

Benchmark	Endeavor 1.0	GOT-5.6 Sol	Claude Fable 5	Kimi K3	Demotron 3 Ultra
GPQA	92.0	94.1	92.6	93.5	86.7
HumanEval	98.2	95.1	97.0	96.3	96.3
IFEval	94.1	95.9	91.7	92.8	91.9
AIME 2026	99.9	99.9	99.9	96.7	94.2

Generalist model

A generalist built for real work

Endeavor 1.0 is designed to act as the core generalist model across products, agents, and complex workflows, rather than being narrowly optimised for a single benchmark family.

Advanced reasoning

Work across mathematics, science, and complex knowledge tasks. Break down difficult problems, sustain detailed multi-step analysis, and work under user guidance.

Long-horizon agent work

Plan and carry out multi-step work across the web, files, code, and external tools. Maintain context over many steps, adapt as new information appears, and recover from intermediate failures.

Build, review and test software

Understand unfamiliar repositories, turn specifications into implementations, reproduce problems, review changes, run tests, and explain results.

Switch to Endeavor 1.0

Two ways to begin

Endeavor 1.0 is designed to act as the core generalist model across products, agents, and complex workflows, rather than being narrowly optimised for a single benchmark family.

Managed by Flower

Move quickly without taking on model operations. Start with a production service operated by Flower. We manage deployment, scaling and model operations for you.

Private deployment

Keep the model and workloads in your environment. Self-host Endeavor 1.0 in your own environment, with Flower support across your data, infrastructure, and systems.

A different starting point

Built from a different starting point

Flower has spent years building the model and distributed systems technology required for powerful AI that organisations can trust, deploy on infrastructure they choose and improve around their own work. Endeavor 1.0 brings that foundation to the frontier.

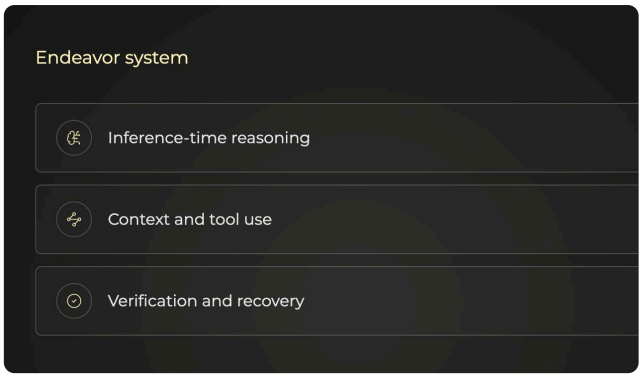
Released four months after Lizzy, a sovereign 7B model built for UK use, Endeavor 1.0 marks the next step in Flower's model programme: from a focused sovereign model to a broad generalist competing directly with today's frontier systems.

The whole system matters

More than model weights

A modern frontier model is more than its weights. We optimise how Endeavor 1.0 reasons at inference time, how it builds and maintains context, how it uses tools, and how it checks, revises and recovers when a step fails.

We develop and evaluate these behaviours across multiple custom coding and agent harnesses, varying tool sets, interaction formats and execution environments so that performance transfers across real workflows rather than depending on a single benchmark setup.

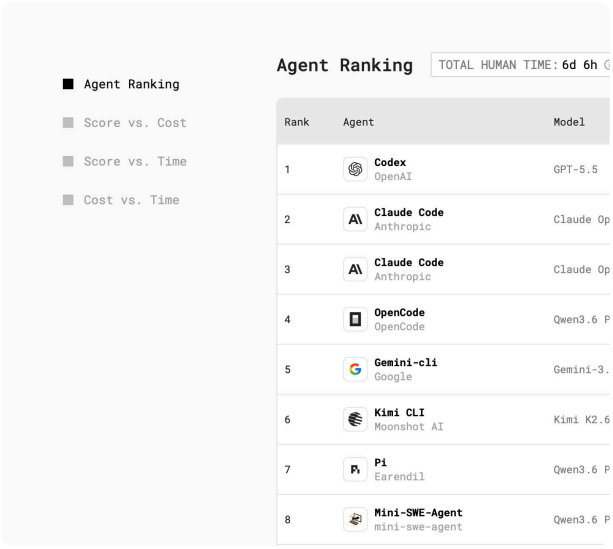


Signals from real work

Evaluated where the work happens

Endeavor 1.0 is measured against today's leading benchmarks, but it is not developed around them alone. FlowerBench gives us a different kind of signal by enabling repeatable evaluation on real, high-value enterprise workflows without moving the underlying proprietary data.

Tasks are contributed by organisations in the opt-in Flower Enterprise Evaluation Network and run inside their own environments. These end-to-end evaluations expose failure modes that public benchmarks rarely capture. We use those signals to guide Endeavor 1.0's evaluation design, harness development, post-training and system-level improvements.



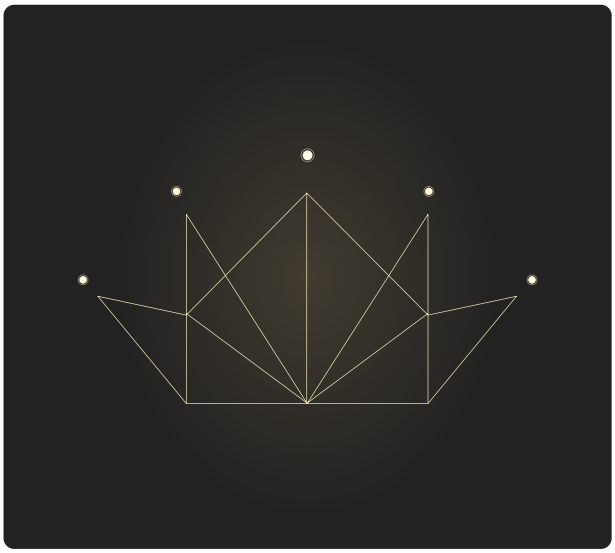
Open foundations, Flower advances

Build on what works. Advance what matters.

Endeavor 1.0 leverages mature, widely available capabilities from leading open-weight models, including general language understanding, public knowledge and common coding patterns. This lets us build on what the wider ecosystem already does well rather than recreating it from the ground up.

To this broad foundation, we add capabilities developed through Flower's own model programme, including UK-specialist knowledge and reasoning from Lizzy, together with new model behaviours, specialist capabilities and training advances developed for Endeavor 1.0.

Continual pre-training, targeted post-training and careful model integration bring these strengths together and enable us to keep refining Endeavor 1.0 over time.



Start with Endeavor 1.0. Build intelligence that becomes increasingly your own.

Use Endeavor 1.0 through Flower today, then extend it with your agents, evals, data, and improvement loops. Turn frontier capability into a lasting AI foundation for your organisation.

Request Access