PRODUCT INTRODUCTION

Introducing Baranos AI

Baranos AI makes LLM inference deterministic and verifiable, with verification and settlement on Fogo. Applications can check the computation behind an AI decision before using its result to settle a market, move funds or trigger an automated action.

With centralized LLMs and private APIs, an application may see the request and response while knowing little about the model version, evidence or execution settings behind them. A centralized provider, or an attacker who gains access, can change those details without the application knowing, even when the answer controls real funds.

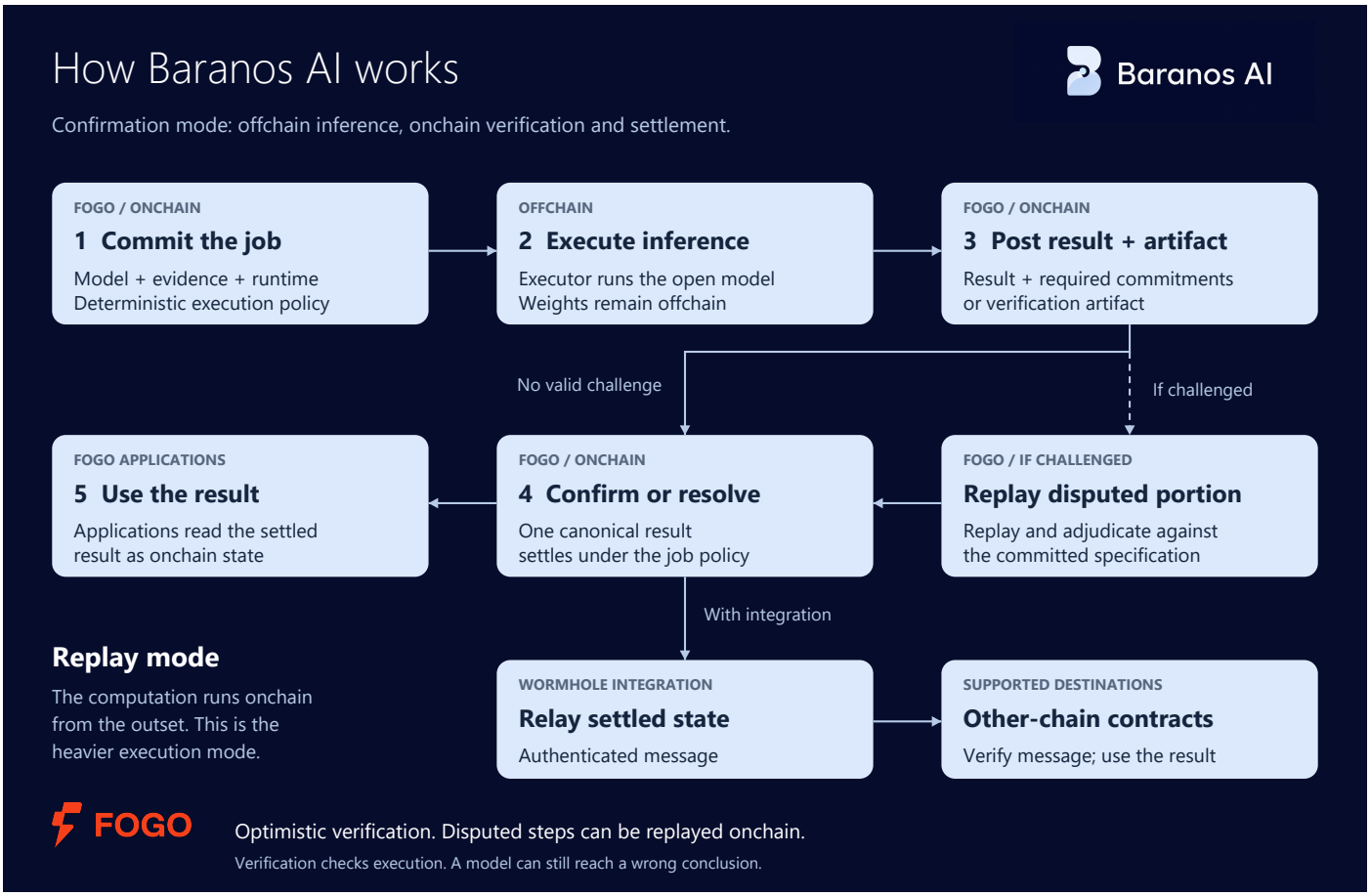
Baranos addresses this through decentralized verification. Each Baranos job commits to a model, its inputs and evidence, and the runtime and execution rules. The same committed job has one canonical result. Independent participants can reproduce the computation and challenge an output that does not follow that specification.

Confirmation mode is the normal operating path. An executor runs inference offchain and posts the result with the required commitments or verification artifact on Fogo. The result can be confirmed under the job's challenge policy if no valid challenge is raised. If challenged, the disputed portion is replayed and adjudicated onchain against the committed specification.

Every component can execute onchain, so an executor's work remains subject to onchain replay even when inference runs offchain. Replay mode runs the computation onchain from the outset, at a higher computational cost. This is an optimistic verification system: correctness is enforced through commitments, challenges and replay, without requiring a zero-knowledge proof for every inference.

Baranos uses open-weight models, whose weights can be fixed and made available for independent execution. In Confirmation mode, the weights remain offchain and a Merkle root commits to them. The team is starting with a 27-billion-parameter model and estimates that this verification approach could support models up to 10 trillion parameters; that upper bound remains an engineering estimate. Initial uses include analysing a wallet's trading history or interpreting a defined set of public discussions. These tasks can use capable open models without requiring the most advanced proprietary systems.

Execution and settlement



Prediction markets give traders a direct way to express a view on an event. A company can miss earnings while its shares rise, leaving a trader right about the event and wrong on the trade. An event market can instead tie its payout to the specified earnings outcome.

Baranos gives builders a resolution process they can use across such markets, including subjective questions. Participants agree on the interpretation criteria and evidence policy before trading, then commit the evidence used for resolution. Disputed execution is checked onchain against those terms, offering an alternative to opaque rulings or token-weighted outcome votes. Verification establishes that the agreed computation ran correctly. A model can still misinterpret the evidence or reach a wrong conclusion.

Fogo's fast blocks and low transaction fees help make frequent verification and disputed-step replay practical. Baranos targets minute-scale results and dispute resolution, with timing dependent on the workload and challenge policy.

Once settled, the result becomes onchain state. Through a Wormhole messaging integration, contracts on other supported chains could receive and use it. Fogo settles the result; Wormhole carries an authenticated message to a receiving contract. Builders could use Baranos for resolution where their users already trade, while defining their own market questions and evidence policies.