



WHITEPAPER · v1.0

Flate

The deterministic-first AI terminal for crypto

1. Executive Summary

Flate is "ChatGPT for crypto" – a research terminal that answers a natural-language question with a short written reply **and** live, interactive UI blocks: token cards, price charts, wallet breakdowns, news, and tables. You ask about a token, a wallet, the market, the news, or a core concept, and you get a structured, data-backed answer in one place.

What separates Flate from generic "AI for X" products is a single architectural decision: **numbers are computed from live data, not generated by a model.** Most assistants route every question through a language model – slow, costly, rate-limited, and prone to inventing figures. Flate instead resolves the majority of real crypto questions through **Quartz**, its deterministic engine, which reads directly from live data providers with **zero model spend**. A language model is used only when a question is genuinely open-ended, and even then its output is forced through a single validated contract before it reaches the screen.

Three pillars define the product:

- **Answers, not links.** A question returns a synthesized answer with the relevant cards and charts inline – not a list of tabs to dig through.
- **Deterministic-first.** Facts are computed; only judgement and explanation use a language model.
- **Structured output.** Every answer – fast-path or model-generated – conforms to one schema-validated contract, so the interface is always well-formed.

Flate is live in production as a complete MVP and is being actively extended. This document describes what exists today and the thesis behind it. A separate tokenomics paper will cover the token in full.

2. The Problem

Crypto research is fragmented, slow, and untrustworthy at exactly the moments it matters most.

Research is scattered across a dozen tabs. Answering one practical question – *is this token healthy, what is this wallet holding, why did the market move today* – means stitching together CoinGecko for prices, a block explorer for on-chain balances, an analytics suite for flows, a news feed, and a social timeline. The data exists, but it is dispersed and demands expert interpretation before it becomes an answer.

Generic AI hallucinates the numbers. General-purpose assistants are fluent but blind: they cannot see live on-chain state or current prices, and they will confidently state a market cap,

a balance, or a 24-hour change that is simply wrong. In a domain where a misplaced figure costs real money, plausible-but-fabricated is worse than useless.

Professional tools are priced and built for analysts. The serious on-chain platforms are powerful but expensive and designed for a professional analyst's workflow – not for someone who wants a direct, readable answer to a direct question.

The user at the centre of this is a crypto-native researcher or trader who already knows the space, but the framing is deliberately accessible to a curious newcomer too. Both want the same thing: a fast, exact, trustworthy answer – not another dashboard to learn.

3. The Solution

Flate is built on a **hybrid engine**: deterministic where possible, generative only where necessary.

Quartz – the deterministic fast path. A large library of intents is answered directly from live data with no language-model call. This covers the bulk of real crypto questions: prices and charts, market overviews, "why is X up or down today," tokenomics, conversions, gas, wallet analysis, and more. Because the answer is computed from data providers, it is instant, free of model cost, and exact.

The model fallback – only when needed. Open-ended or novel questions fall through to a language model that returns schema-validated structured output, which the frontend renders into the same rich UI blocks. The model is reached for reasoning and explanation, never for the underlying figures.

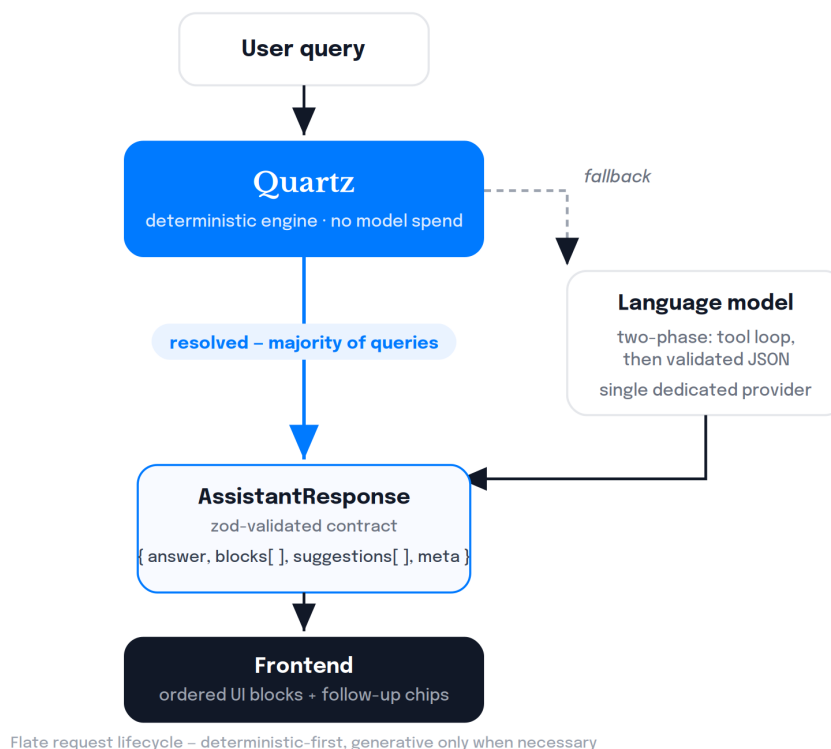
This split is the core of the product, and it is its defensible advantage. Quartz is Flate's primary intellectual property: a deterministic answer layer that makes most queries cost nothing, return instantly, and never invent a number. Competitors that pass every question through a model inherit that model's cost, latency, and hallucination – by construction.

A query, end to end

Take "Why is SOL up today?" Quartz recognizes the intent, fetches Solana's recent price action and the news driving it, and assembles the response: a one-paragraph explanation, a **token card** for SOL (price, 24-hour move, market cap, sparkline), and a **news list** of the relevant headlines – with three follow-up chips to continue. No model was called, the figures came straight from the data providers, and the answer arrived effectively instantly.

Smart-money flows and airdrop intelligence are delivered as a curated, weekly-refreshed dataset, clearly attributed as such. Real-time on-chain feeds for these are on the roadmap (Section 10).

4. Architecture



Flate request lifecycle – deterministic-first, generative only when necessary

Figure 1 – Flate request lifecycle: deterministic-first, generative only when necessary.

Every query enters one pipeline. Quartz attempts a deterministic resolution first; only if it returns nothing does the request reach the language model. Either way, the result is the same validated contract, and the frontend renders it identically.

Inside Quartz. Quartz detects the intent of a query, fetches the relevant live data, and assembles a structured response – deterministically, with no model in the loop. Its operating principles are transparent (intent → live data → structured JSON); the specific intent library and matching logic are Flate's core IP and are not published.

Two-phase model use. When a question is open-ended, the model runs in two phases. In phase one it may call tools to gather data – token info, news, wallet holdings, and so on. In phase two it emits the final structured JSON. Separating data-gathering from formatting keeps the output reliable.

A dedicated model provider. When the model is needed, Flate uses a single, dependable provider chosen for quality and consistency rather than a patchwork of free tiers. Cost-efficiency still comes from Quartz, not the model layer – because most traffic never reaches a model, model spend stays low even on a premium provider. The provider sits behind a thin adapter, so it can be swapped without touching the rest of the system.

One output contract. Every answer is an AssistantResponse: a markdown answer, an ordered list of typed blocks, follow-up suggestions, and meta (data sources, disclaimers,

wallet hints). It is validated with zod before rendering. This single contract is why the interface is always well-formed regardless of whether an answer came from Quartz or from a model.

Stack. Next.js (App Router) and React on Cloudflare Workers via OpenNext; Tailwind and shadcn/ui for the interface; Cloudflare D1 for chat history and KV for response caching; Reown AppKit with wagmi (EVM) and a Solana adapter for read-only wallet connection.

5. Data & Accuracy

The product rests on one commitment:

Numbers are computed from live data providers, never generated by a model.

Prices, market caps, balances, and changes are read from data sources and passed through untouched. The language model, when used, reasons and explains around those figures – it does not produce them. This is what makes Flate's output safe to act on in a way that a general assistant's is not.

Source attribution as a trust mechanism. Every answer carries its dataSources in meta, surfaced to the user. An answer always tells you where its data came from, so trust is verifiable rather than assumed.

Data providers.

Source	Used for
CoinGecko	Prices, charts (area/candles), market caps, global market, dominance, top movers, trending, tokenomics/ATH, historical price, project descriptions
Alchemy	EVM portfolios across Ethereum, Base, Arbitrum, Optimism, Polygon; live ETH gas
Helius	Solana portfolios
CryptoPanic + RSS	Crypto news, general and per-token
alternative.me	Fear & Greed Index
Flate curated dataset	Smart-money flows, airdrops, capital-flow / sector rotation, projects-without-a-token – weekly-refreshed, attributed

6. Product & Capabilities

Most of what follows is answered by Quartz with no model call. The interface renders every answer as ordered UI blocks – **12 block types** in total – so a reply is never a wall of text.

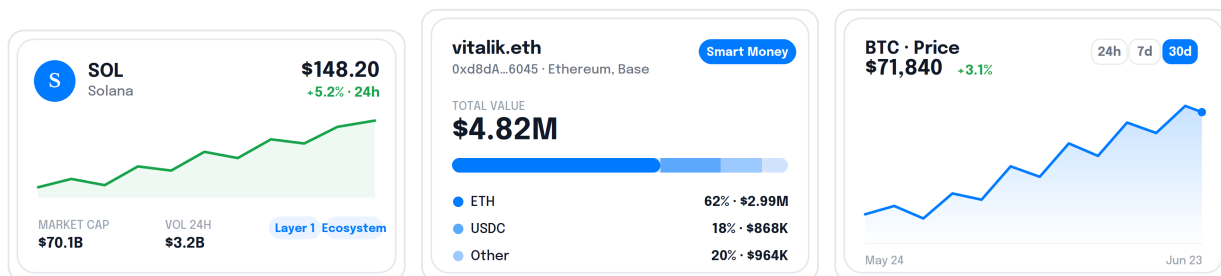
What you can ask

Markets & Tokens – *"What's SOL doing?" · "Why is the market down today?" · "Compare ETH and SOL."* Token price and info, area or candlestick charts (24h / 7d / 30d / 1y), side-by-side comparisons, a full market overview (total cap, dominance, Fear & Greed, top movers, news), trending coins, gainers and losers, tokenomics and all-time-high stats, value conversions, live ETH gas, and historical prices. You get the figures plus the cards and charts that contextualize them.

Wallets – *"Analyze this wallet." · "What does vitalik.eth hold?"* Portfolio analysis for any EVM or Solana address, ENS (.eth) or SNS (.sol) name, or a connected wallet – holdings, total value, and allocation breakdown. Connect read-only in one click, or simply paste an address.

Learn & Discover – *"What is a rug pull?" · "What is AVAX?"* A built-in glossary of **113 crypto concepts**, project primers, and curated smart-money, airdrop, capital-flow, and projects-without-a-token intelligence.

Example blocks



Token card

Wallet card

Price chart

The 12 block types

Block	Renders
text	Markdown prose
tokenCard	A single token: price, change, cap, volume, tags, sparkline, links
tokenList	A ranked or grouped set of tokens
walletCard	Address, total value, top holdings, allocation, labels
newsCard	One headline with source, summary, sentiment
newsList	A feed of related headlines
priceChart	Area or candlestick price over time
portfolioAllocation	Donut breakdown of a wallet's holdings
portfolioValueChart	Portfolio value over time
smartMoneyCard	A tracked wallet's accumulation/distribution
airdropCard	An airdrop: status, chain, requirements, links
table	Generic tabular data (e.g. capital flows)

Follow-up chips accompany every answer and are engineered to resolve back into Quartz, keeping the user in the instant, zero-cost lane. History is per-device with shareable thread URLs, requests can be cancelled mid-flight, disclaimers surface automatically, and the interface is fully responsive.

7. Positioning & Principles

Flate sits between the generic assistants (fluent but blind to live data) and the professional analytics suites (accurate but expensive and analyst-oriented).

	ChatGPT	Perplexity	Nansen / Arkham	Manual research	Flate
Data accuracy	Generated	Retrieved	Computed	Computed	Computed
Cost model	Per-query model	Per-query model	High subscription	Free, your time	~\$0 for most queries
Speed	Slow	Slow	Fast	Slow (manual)	Instant fast-path
Output	Text	Text + links	Dashboards	Raw data	Answer + live cards
Audience	Everyone	Everyone	Pro analysts	Pro	Everyone

Five principles govern the product:

1. **Answers, not links.** A synthesized, data-backed answer with the relevant cards inline – not sources to dig through.
2. **Deterministic where possible, generative where necessary.** Facts are computed; only judgement uses a model.
3. **Neutral and risk-aware.** Sharp and concise, no hype or shilling, with disclaimers on anything touching investing. Flate does not give personalized financial advice.
4. **Privacy-light by default.** No accounts required; identity is a per-device cookie; secrets are server-only.
5. **Cost-efficient by design.** Quartz keeps unit economics near zero for the bulk of traffic.

8. Security & Privacy

Read-only by design – no custody, ever. Wallet connection is strictly read-only. Flate never signs transactions, never holds private keys, and never takes custody of funds. It reads public on-chain data to answer questions; that is all. This is the single most important guarantee for a crypto-native user, and it is structural, not a policy.

Identity model. The MVP deliberately ships without accounts. Identity is a per-device cookie, history lives on Cloudflare D1, and all secrets are server-side and never exposed to the browser. This is a conscious scoping decision for the MVP, not an omission – optional accounts and cross-device sync are on the roadmap, layered on top of the existing cookie identity rather than replacing it.

9. Monetization & Token

Flate's business stands on its own, independent of any token. A **subscription Pro tier** – higher limits, alerts, and premium data – is the primary revenue path and does not depend on a token existing.

A **Flate token** is planned. At a high level it is expected to power premium access to the terminal's higher tiers and premium data, with **governance under exploration**. The token's full mechanics – supply, allocations, and utility – are deliberately out of scope here and will be covered in a dedicated tokenomics paper. Nothing in this document constitutes a token offering.

10. Current Scope & What's Next

Flate's MVP is complete and in production. Several things sit deliberately outside its current boundary – each a conscious choice, with a clear next step. This section is a direction, not a commitment; the full roadmap is a separate document.

- **Real-time on-chain smart-money & flows.** Today these run on a curated, weekly-refreshed dataset – reliable and attributed, but not live on-chain. Next: replace the dataset with real-time on-chain analytics.
- **Push, streaming & alerts.** Answers are request/response by design, which keeps the system simple and predictable. Next: watchlists and price/news/gas alerts, then live-updating cards.
- **Cross-device accounts.** The cookie identity keeps the product frictionless and privacy-light. Next: optional sign-in for history that follows the user across devices.
- **Gas beyond Ethereum.** Live gas is Ethereum-only for now. Next: additional chains for gas and portfolios.
- **Portfolio history over time.** The `portfolioValueChart` block exists but is not yet fed by a time-series source. Next: wire it to a balance-history feed, with profit-and-loss.

Further out, the platform direction includes a public API exposing the `AssistantResponse` engine, a mobile app reusing the same engine, and power features such as transaction simulation and on-chain "explain this contract."

11. Vision

Flate begins as a research terminal, but the ambition is larger: to become the **default interface for interacting with crypto** – the layer through which people ask questions, understand the market, and eventually act, all in plain language. The same engine that today

computes an exact answer can tomorrow power an API, drive agents, and extend from reading the chain toward safe, user-controlled action.

The thesis holds the whole way up. In a space drowning in dashboards and confident-but-wrong chatbots, the durable advantage is simple: **answers, not links – and numbers that are computed, not guessed.**

12. Disclaimers & Legal

This document is for informational purposes only. It is **not financial advice** and does not constitute a recommendation to buy, sell, or hold any asset; Flate does not provide personalized financial advice. Nothing here is an offer or solicitation to sell securities or a token, in any jurisdiction.

This whitepaper contains **forward-looking statements**. Roadmap items describe direction and intent, not guarantees; features, timing, and scope may change. Smart-money, airdrop, and related intelligence are provided as a curated dataset and are labelled as such within the product.