

Bitcode

The Pitch



What's the problem?

Data acquisition, for AI, AI companies, AI developers

The most valuable input to AI is the hardest thing to buy.

- License it → nine-month deals that collapse (OpenAI × NYT).
- Scrape it → get sued. Anthropic just paid **\$1.5B**.
- Need a pipeline? → buy the company. Meta paid **\$14.3B**.
- Own dead code? → it rots — sold for **~5%** of value, if it sells.

No prices. No liquidity. No market. The fuel of the AI era is still bought by hand.

Does it matter?

~\$100B of mattering today, growing at ~25+% Y/Y

One of the biggest markets forming — and no one owns it.

- **\$4B → \$23B** AI training data, now to 2034.
- **~\$8–12B** spent acquiring data this year — AI's fastest-growing line.
- **\$37B** enterprise 2025 gen-AI spend; human data now rivals compute.
- **~2028** — the public web of **human data runs dry** (Epoch AI); demand compounds, supply doesn't.

Code is the highest-value data — with no liquid market. Zero.

Colossal demand. Trapped supply. No exchange in between.

Bitcode, Experiential

A market you operate in minutes — not a procurement war.

You don't negotiate a data deal. You Deposit, Read, Settle.

- **Reviewable AssetPacks, not raw repos** — sellers list measured intelligence; buyers pick the fit.
- **Source-safe previews** — see the value before you pay; the crown jewels stay locked until rights clear.
- **Need-first buying** — describe the data, get candidates ranked by fit. No 40TB dumps to sift.
- **Dignity for trapped code** — dead and private code becomes inventory that earns, not code that rots.
- **Settle in clicks** — BTC for money, BTD for the knowledge you bought. No contracts, no NDAs, no sales calls.

Colossal demand. Trapped supply. Bitcode is the exchange in between — felt as Deposit, Read, Settle.

Bitcode, Technological

The stack that turns technical source into priced inventory.

Not just another model. A market for measured code — measurement, inventory, settlement.

- **The AssetPack is the commodity** — a measured, source-safe patch. Priced knowledge, not a corpus.
- **Measurement is law** — absolute scores on deposit, Need-fit scores on read, BTD as knowledge volume. That's how code gets a price.
- **A real exchange** — supply (Depository), demand (Needs), clearing (BTC settlement, assets rights + BTD settlement, assets delivery). The layer this market has never had.
- **Source-safe by construction** — fail-closed disclosure across web, API, and agents. Trust is a system property, not an NDA.
- **Two industrial pipelines** — deposit synthesis and read synthesis; hosted, observable, built to scale.

Code is the highest-value data with no liquid market. Bitcode is the measurement layer, the inventory, and the exchange.

Cryptography: Bitcode Tokens (\$BTD, AssetPacks)

The rails that turn measured knowledge into scarce, ownable rights.

Not a meme coin — ledger objects backed by measurement.

- **BTC — the money.** Buyers settle reads in Bitcoin; finality before anything ships.
- **\$BTD — the volume.** Scarce (21M cap), data-backed; minted from measured Need-fit, after payment
- **AssetPack — the commodity.** A measured pack (patch + measurements + metadata), held as on-chain co-ownership: the depositor keeps it, the buyer is added.
- **Volume is protocol; price is market.** The protocol measures how much knowledge fits the Need; the market prices the token.
- **Source-safe, fail-closed.** Measurements are visible, IP stays locked, and money + rights clear before delivery.
- **Fully open-source.**

The AI era's fuel becomes a ledger object — volume you can hold, packs you can own, paid in Bitcoin.

Bitcoin is the cash rail. Bitcode is the knowledge rail.

Who am I?

“I want to be the very best, like no one ever was.”

My Pokémon are data. I find them — it takes a good eye and a feel for what the market will want next. I catch them on Bitcode, because trading them anywhere else earns you less, are harder to buy, and are insecure. And I train them — measuring them, making them much more valuable to the market. That's the whole game. It's the one I was built to play. My first company, Engi Holdings, Inc., failed — so I'm not starting from scratch. I'm starting forged. I've learned to build and lead a team, to run with operational discipline, to truly hear a customer, and to push ambitious technology forward. The lessons that break many founders, I already carry. I'm the best person in the world to build this — where art meets science. Decades of full-stack software engineering, decades in the creative arts, real entrepreneurial scars, and a commitment and foresight to this category I've yet to see in anyone else — down to the deepest truth-layers of the economics of knowledge. This is my life's work, and I intend to win it.

The plan?

TOTAL VICTORY

Now

Launch • ~30 days

MVP: measurement, the exchange, tokenization.
An excellent product on infrastructure that holds.
First suppliers, first buyers, first revenue.
The two-sided market opens.

The flywheel turns.

Sooner

Team • Lovers

The founding team, assembled.
The first large cohort of users.
Not people who try Bitcode —
people who love it.

They don't leave.

Soon

Market Gets Smart

Measurement deepens toward real demand.
The market reshapes around what's most needed.
Plugins inside the tools devs already use.
Native chat, over MCP.

The market compounds.

Later

Game Over

An internal model on the full Bitcode Depot
Enormous engineering gains, compounding.
Code becomes all technical knowledge — STEM.
\$BTD: 3rd-largest crypto • AssetPacks: #1 NFT.

The category is ours.

Money has Bitcoin.

Knowledge has Bitcode.

\$BTD: 3rd-Largest Crypto
AssetPacks: #1 NFT

Appendix

Every claim in this deck, mapped to its source — drawn from two internal briefs: the Bitcode Market Analysis and the Supply & Demand study. Every load-bearing figure was cross-checked against multiple independent sources.

- A1 Problem — the four broken paths
- A2 Market sizing — the numbers behind \$100B
- A3 The ~\$100B “mattering” stack
- A4 Why now — the thesis
- A5 Competitive white space
- A6 Take-rates & today’s friction
- A7 TAM / SAM / SOM
- A8 Product mechanics & method
- A9 Full bibliography



A1 • Problem — the four broken paths

Slide: “Data acquisition is the hardest thing to buy.”

Claim on the slide	Figure / fact	Source
License it → deals collapse	OpenAI–NYT ~9-month talks, then litigation; AFP–Mistral “after months”	Press Gazette; Quartz
Scrape it → get sued	Anthropic \$1.5B settlement (~\$3,000/book, ~500k works; Bartz v. Anthropic)	NPR; Authors Guild
Need a pipeline → buy the company	Meta \$14.3B for 49% of Scale AI (Jun 2025)	Time; Forbes
Own dead code → it rots	Kodak: up to \$4.5B → \$94M (~4%) after ~11 mo; Nortel raised 450% of estimate	IEEE Spectrum
No prices / liquidity / market	~52 known licensing deals; 38 sellers → 12 buyers; marketplaces <2% of data-trade volume	Neudata; Data Boutique

A2 • Market sizing

Slide: “~\$100B of mattering today, growing ~25%+ Y/Y.”

Market (definition)	Today → out-year	CAGR	Source
AI training dataset (narrow)	\$3.2B → \$16.3B (2033)	22.6%	Grand View Research
AI training dataset (narrow)	\$3.59B → \$23.2B (2034)	22.9%	Fortune Business Insights
AI training dataset (conservative)	\$3.35B → \$14.9B (2035)	16.1%	Precedence Research
Data collection & labeling	\$3.77B → \$17.1B (2030)	28.4%	Grand View Research
Dataset licensing for AI training	\$4.8B → \$22.6B (2034)	18.8%	MarketIntel
Custom-LLM training platforms	\$2.82B → \$8.63B (2030)	25.1%	Research and Markets
Enterprise LLM	\$5.65B → \$41.6B (2033)	28.3%	Grand View Research
Enterprise gen-AI spend (2025)	\$37B (\$18B infrastructure)	—	Menlo Ventures

A3 • The ~\$100B “mattering” stack

How the headline number is built — and where it’s soft.

Bucket (today)	Figure	Source
Enterprise gen-AI spend	\$37B (2025)	Menlo Ventures
Data-acquisition wallet	~\$8–12B (datasets \$3.9B + labeling \$3.8B + expert-data vendors \$4–6B)	GVR; Sacra
Model-build / fine-tuning	~\$6–9B (custom-LLM \$3.5B + enterprise-LLM \$5.65B)	R&M; GVR
Strategic data M&A	\$14.3B (Meta ↔ Scale AI)	Time; Forbes
Data slice of frontier capex	~\$12–20B (3–5% of Alphabet \$175–185B, Meta \$66–72B, MSFT ~\$100B+)	Company capex guidance
≈ Total “mattering” today	~\$70–90B → rounds to ~\$100B; growth ~25%+	This stack

A4 • Why now — the thesis

Scarcity, value, and legal risk — each independently sourced.

Claim	Evidence	Source
Data runs dry ~2028	Public human-generated text exhausted 2026–2032 (median ~2028); indexed web ~500T words	Epoch AI
Code is the highest-value data	In pre-training: +8.2% reasoning, +4.2% world knowledge, +6.6% win-rate, ~12× on code	Cohere, “To Code, or Not To Code?”; VentureBeat
Human data now rivals compute	Data-labeling spend grew +88× vs compute +1.3× (2023–24)	Time; Forbes; D. Kang (UIUC)
Legal risk repriced “free” data	GitHub Copilot suit live on license claims (Doe v. GitHub); Microsoft MAI = “clean, licensed data”	githubcopilotlitigation.com; Microsoft AI

A5 • Competitive white space

No token-backed, quality-measured, live-priced market exists for code IP.

Player	Measures quality?	Live pricing?	Rights / notes
Bitcode	Yes — absolute + dynamic	Yes — live bidding	Unlimited resale; native to code
Story / DATA Foundation	No (provenance only)	No (mint fee + royalty)	Royalty-encumbered; general IP
Ocean Protocol	Retired (2024)	Partial (AMM)	Access licenses; not code
Vana	Weak (staking)	No	DataDAO tokens; not code
Sahara AI	Weak (attribution)	No	Non-transferable shares
Numerai / Bittensor	Strong (stake / validators)	No (closed system)	Predictions/outputs, not IP
Scale / Surge / Mercor	Internal QA only	No (bespoke)	Work-for-hire; manufactured
IPXI / IPwe	No	No	Failed IP exchanges (2015 / 2024)

A6 • Take-rates & today's friction

What comparable venues charge — and why the incumbents don't clear the market.

Venue (comp)	Take rate	Note
Databricks Marketplace	0%	Direct provider ↔ consumer (Delta Sharing)
AWS Data Exchange	3%	Seller keeps 97%
Ocean Protocol	~0.3%	On-chain datatokens (config-dependent)
Hugging Face Hub	No data take	Monetizes subscriptions / compute
Scale AI	~50% markup	Vertically integrated data engine
Mercor	~30–40%	Curated expert-data (contractors keep 60–70%)
App Store / Steam (ref)	~30%	Reproducible digital-goods ceiling

A7 • TAM / SAM / SOM

The sizing model — scenario estimates with explicit assumptions.

Layer	Today	Early 2030s	Basis
TAM — data-acquisition wallet	~\$8–12B	~\$25–30B+	Datasets + labeling + licensing + expert data
SAM — code-specific, rights-cleared	~\$1.5–3B	~\$6–12B+	Code = 15–25% of the wallet
SOM — 3% passive take	~\$4.5M	~\$45M	On ~\$150M → \$1.5B intermediated GMV
SOM — 30% curated take	~\$45M	~\$450M	On ~\$150M → \$1.5B intermediated GMV

A8 • Product mechanics & method

What the solution claims rest on.

Element	What it is	Source
The model	Deposit · Read · Settle; measurement → price → BTC settlement → \$BTD + AssetPack rights → delivery	Bitcode V48 SPEC / Whitepaper
AssetPack	A synthesized, measured patch (patch + measurements + metadata) — never raw source	Bitcode SPEC
\$BTD	Fungible knowledge-volume token; minted from Need-fit measurement after payment; 21M cap	Bitcode SPEC
Measurement anchor	Objective, test-pass-style scoring in the SWE-bench lineage	SWE-bench; arXiv 2509.16941
Research method	5-angle fan-out; adversarial cross-verification against independent sources; 14 of 15 figures confirmed	This appendix

A9 • Bibliography

Full source list across both briefs.

- Grand View Research — AI Training Dataset; Enterprise LLM; Data Collection & Labeling; Data Marketplaces
- Fortune Business Insights — AI Training Dataset Market
- Precedence Research — AI Training Dataset Market
- Business Research Insights — AI Training Dataset (broad)
- MarketIntel — Dataset Licensing for AI Training
- Research and Markets — Custom LLM Training Platforms Outlook (2026)
- Menlo Ventures — 2025 State of Generative AI in the Enterprise
- Quartz — The price of AI training data (\$5M–\$250M)
- Variety / WSJ — News Corp–OpenAI licensing deal
- TechCrunch — Reddit \$203M licensing disclosure
- Time / Forbes — Meta \$14.3B for 49% of Scale AI
- Press Gazette — publisher AI deals & lawsuits (OpenAI–NYT)
- Neudata — AI data-licensing market (52 deals)
- Epoch AI — Will we run out of data?; Frontier model count (arXiv 2504.16138)
- Cohere / VentureBeat — code in pre-training improves non-coding tasks
- NPR / Authors Guild — Anthropic \$1.5B copyright settlement
- GitHub Copilot Litigation (Doe v. GitHub); The Register — DMCA dismissal
- GitHub Octoverse 2025; Software Heritage (2026)
- a16z — How 100 Enterprise CIOs Are Building & Buying Gen AI (2025)
- IEEE Spectrum — The Lowballing of Kodak’s Patent Portfolio (2014)
- Fox Business / Wikipedia — Nortel patent auction (2011)
- CoinDesk — Vana VRC-20; Messari / Story Docs — Story & DATA Foundation
- Numerai / Chainwire; Contreras (2016) — IPXI; IAM (2024) — IPwe
- Databricks, AWS Data Exchange, Ocean Protocol — marketplace take-rate docs
- Data Boutique — data marketplaces broken (<2%); arXiv 2502.04484 — HF licensing
- Sacra — vendor profiles (Scale, Surge, Mercor, Turing, Handshake, Invisible)
- SWE-bench / arXiv 2509.16941 — patch-generation benchmark
- Bitcode SPEC / Whitepaper — github.com/advancedengineeredsoftware/Bitcode