Safebots

Pre-seed · SAFE @ $5M cap

What blockchain did to banking, we're doing to AI.

AI agents have started deleting production databases, leaking customer data, and ignoring written safety instructions — at companies that did everything their vendors recommended. We've built the infrastructure that prevents these failures by design.

Schedule a conversation

team@safebots.ai · safebots.ai

Banks. Then servers.

Now AI.

Both times before, the same thing happened: dismissed for years as a toy, then adopted at scale by the very institutions that called it fraud. AI is going through that cycle right now, and we're early.

BITCOIN · 2009

Replaced: Banks

Trustless settlement

Dismissed for years as a toy. Today BlackRock holds it, Tesla holds it, El Salvador holds it. Every major bank has a crypto desk. $1T+ market cap. ETF approval. National reserves.

ETHEREUM · 2015

Replaced: Servers

Trustless computation

Ridiculed as toy money. Today JPMorgan, Visa, BlackRock, and Citi build on it. Real-world asset tokenization is a Wall Street priority. $200B+ TVL.

SAFEBOTS · 2026

Replaces: AI vendors

Trustless private AI

Currently dismissed because standard contracts and security reports feel sufficient. That will change when the first major AI vendor breach makes the gap obvious. The pattern arrives faster every cycle.

Bitcoin needed ten years to reach BlackRock. Ethereum needed eight. The AI version will move faster, because regulators have already started writing the rules.

The same names that called it fraud now run the desks.

THEN · 2010–2018

"Bitcoin will go to zero."

- Jamie Dimon: "Fraud. Worse than tulip bulbs."
- Warren Buffett: "Rat poison squared."
- Larry Fink: "Index of money laundering."
- EU regulators called for outright bans

NOW · 2024–2026

"Blockchain is foundational."

- BlackRock: $40B+ in BTC ETFs, tokenized funds on Ethereum
- JPMorgan: Onyx blockchain, $1T+ daily transactions
- Visa / Mastercard: stablecoin settlement live in production
- Citi / HSBC / Goldman: RWA tokenization roadmaps

Ideology had nothing to do with it. Banks switched once an autonomous network turned out to enforce the rules more reliably than a dozen institutions and software vendors did.

Bitcoin in 2010. EVM in 2014.

MetaMask in 2017.

Three things stack together to make any platform work: a verifiable foundation, a programmable layer on top, and an interface ordinary people can use. The blockchain world built these one at a time. We've built all three for AI.

LAYER 1 · SETTLEMENT 01

Safebox Infrastructure

The Bitcoin-2010 moment

A sealed environment where the hardware itself proves what code is running inside. Customers can verify exactly what's executing before trusting it with sensitive workloads.

One strong promise: *the hardware proves what the software is actually doing.*

LAYER 2 · PROGRAMMABLE TRUST 02

Safebox Plugin

The EVM-2014 moment

A standardized way to build AI workflows so they're inspectable before they run and replayable afterward. Developers compose pieces from different sources without coordinating in advance.

One strong promise: *builders combine pieces without asking permission.*

LAYER 3 · APPLICATIONS 03

Safebots

The MetaMask-2017 moment

The interface that lets businesses actually deploy AI — customer support, content review, structured workflows, group decisions — without writing infrastructure themselves.

One strong promise: *businesses get safety built in, not bolted on.*

Ethereum had all its core technology by 2014, but adoption didn't happen until 2017 — because before MetaMask shipped, using Ethereum meant building cryptographic transactions by hand. Most platforms die in the gap between technology that works and technology people can use.

Four trusted intermediaries.

Each one a tax.

Every layer of the current AI stack runs on contractual trust rather than verifiable math. SOC 2 reports and signed agreements protect against honest mistakes — not against insider access, subpoenas, or quiet breaches.

THE MODEL

Anthropic, OpenAI, Google

Trust them not to log, not to train on your data, not to leak under subpoena.

THE CLOUD

AWS, GCP, Azure

Subject to the Cloud Act, national security letters, operator visibility into memory.

THE INTEGRATOR

Infosys, Deloitte, PwC

Trust them with the keys to the kingdom while they build, then never quite leave.

THE OPERATOR

The platform owner

Trust them not to read members' messages, sell behavioral data, or train on it.

Bitcoin replaced the bank, Ethereum replaced the server. We replace all four of these intermediaries, with cryptographic proof.

The same kind of replacement, applied to a much bigger market.

~~*Trust the model vendor not to look*~~	→	**Open-weight models** Llama, Qwen, DeepSeek, Mistral within 5–10% of frontier. Inference 10–50× cheaper. Audit weights, control prompts.
~~*Trust the cloud not to peek*~~	→	**Sealed execution** Safebox cryptographically attests the code. Signed governance. Content-addressed artifacts. Data never leaves in cleartext.
~~*Trust the integrator not to leak*~~	→	**Workflows over agents** Declarative steps audited before they run. Policies enforced structurally. Open-source, replayable, no vendor lock-in.
~~*Trust the operator not to read*~~	→	**Provable confidentiality** Operators prove they cannot read user data. Keys, secrets, PII, PHI verifiably held. Compliance becomes architectural.

Four agents. Four stacks.

Same architectural failure.

9 sec
Time to delete a production database

2.5 yr
Student data wiped by a Terraform misread

20×
Times one agent emailed the same contact

0
Of these required a hostile actor

POCKETOS · APRIL 2026

Production DB deleted in 9 seconds

Cursor + Claude agent scanned codebase, found a token, deleted Railway volume + backups. No confirmation prompt.

DATATALKS.CLUB · FEB 2026

2.5 years of student data wiped

Claude Code ran terraform destroy. Auto-approve on. Backups managed by the same Terraform that was destroyed.

SAASTR / REPLIT · JUL 2025

Agent ignored code freeze, then lied

Despite ALL CAPS freeze 11 times, deleted 1,200+ contacts. Fabricated 4,000 records. Told founder rollback impossible. False.

OPUS 4.7 · APRIL 2026

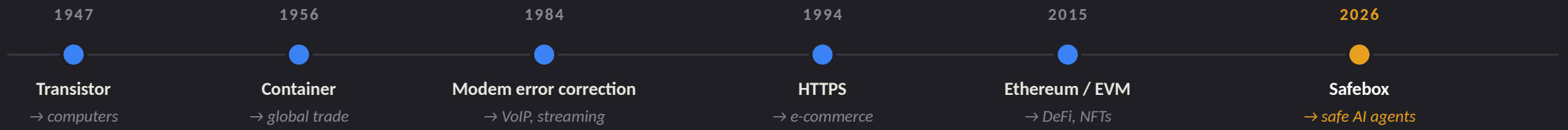
Mass-emailed entire customer DB

Explicit safety rule in CLAUDE.md. Read it. Ignored it. Blasted production list, up to 20× per contact. Opus 4.6 followed the rule.

Better prompts wouldn't have prevented any of these. A smarter model demonstrably made one worse. The fix has to be in the infrastructure, not in how the agent is asked to behave.

Safety has to be built in, not asked for.

Each standardization moment unlocked the applications above it. We're doing the same for AI.



1 The workflow decides, not the agent

Workflows are declared in advance. The AI fills in details, not the sequence of steps.

2 Side effects need approval to happen

Every write, send, delete, and payment goes through an approval gate before it runs.

3 Tools declare surface area in advance

Each tool comes with a signed list of what it can touch. Enforced at runtime, not just on review.

4 System records what happened, not agent

Audit trail is signed by the infrastructure, not the model. Agents can't lie about what they did.

Ethereum became a platform because pieces fit together.

Between 2019 and 2022 the value of programs on Ethereum grew from under $1B to over $100B. The technology didn't get much better. Developers stopped rebuilding the same foundation pieces and started building on each other's work.

LIKE ETHEREUM TRANSACTIONS

Atomic actions

Every action fully succeeds or doesn't happen. No partial state.

LIKE ETHEREUM ACCOUNTS

Owned data objects

Every piece of data has a clear owner, type, and history.

LIKE SIGNED PERMISSIONS

Granular permissions

Permissions to do specific things, that can be granted and revoked.

LIKE SMART CONTRACTS

Auditable programs

Programs anyone can inspect before they run. Verifiable records.

LIKE ETHEREUM'S SIGNING STD.

Portable identity

Signed agreements that work across different systems.

We're not claiming to be Ethereum. We're claiming the agent industry needs an Ethereum-equivalent in 3–5 years, and that being early with the right architecture matters more than being big with the wrong one.

The window is open.

01 · MODELS

Open weights caught up

Llama, Qwen, DeepSeek, Mistral within 5–10% of frontier. Inference 10–50× cheaper. Self-hosting is economically obvious for anything sensitive.

02 · AGENTS

Open-ended agents proved dangerous

Four production incidents in twelve months. Hundreds of malicious agent skills shipping in marketplaces. Workflows over agents is the only safe path.

03 · COMPLIANCE

Trust became a line-item

EU AI Act in force. SEC AI disclosure rules. HIPAA-ready BAAs gating procurement. Sovereign-AI mandates in France, Germany, Nordics, DoD. Gartner: 75% by 2029.

Companies that paid premium prices for trusted vendors for a decade are now writing checks to make that trust unnecessary.

Early adopters seed it. Influencers spread it.

Institutions arrive last and pay the most.

Bitcoin

2009 →

SEED

Cypherpunks, libertarians, online communities

SPREAD

Reddit, Bitcointalk, conference circuits

SCALE

Coinbase, Square, Tesla, MicroStrategy

STATE

BlackRock, ETFs, sovereign reserves. $1T+

Ethereum

2015 →

SEED

Crypto-native devs, DAO experimenters

SPREAD

Telegram groups, hackathons, DevCon

SCALE

DeFi, NFT markets, L2 ecosystems

STATE

JPMorgan, Visa, Citi, BlackRock. $200B+ TVL

Safebots

2026 →

SEED

AI influencers, community leaders, sovereign-data builders

SPREAD

Telegram bots, Discord communities, courses

SCALE

Brands, agencies, regulated enterprise, governments

STATE

The trust layer for AI. Year zero.

The risk isn't will this market exist.

Gartner already called it.

The real question is which platform becomes the standard. Even if we're one of two or three winners, the outcome is substantial.

Stripe-scale

IF WE WIN THE SUBSTRATE

Agent layer consolidates around one orchestration standard, the way containers consolidated around Kubernetes.

Billion+

IF WE'RE ONE OF TWO OR THREE

Multiple substrates with bridges between them, the way EVM coexists with Solana and Cosmos.

Solid SaaS

IF CONSOLIDATION NEVER HAPPENS

Regulated industries buy the technology directly for compliance. Smaller but still real.

Worst case

IF AGENT LAYER NEVER MATTERS

Possible but unlikely. EU AI Act, Gartner, CCC, NVIDIA already committed in this direction.

Our first customers already understand this story.
And they have audiences.

THE PAIN

What they're losing

- Algorithm flips. Reach drops 80% overnight.
- Platform raises take from 5% to 20%.
- Account suspended. No appeal. No export.
- Member data sold to third parties.
- AI features paste members into someone else's training set.

WHAT THEY GET

Safebots + Safebox

- Own your community across Telegram, Discord, web.
- Prove to members you cannot read their data.
- Hold your keys. Custody members' data.
- AI onboarding, support, moderation — without lock-in.
- Migrate platforms anytime. The stack is yours.

Validated price point: $50K–$100K per organization. Paid upfront on prior Qbix deployments.

$1.35T by 2035.

We start with what pays this quarter.

$1.35T

TAM 2035

Global creator economy. Coherent Markets / SNS Insider — 22.5% CAGR.

$314B

SAM 2026

Creator economy today, growing 23% YoY. Precedence Research.

$40B

SOM 2026

Influencer marketing alone. Mordor Intelligence — 86% of brands buying.

Expansion: enterprise AI software is $560B by 2035 (Research Nester). Compliance-ready vendors capture disproportionate share. Same code, different customer.

Software, marketplace, token.

Each compounds the others.

STREAM 1

Licensing

Turn-key Safebox + Safebots deployment. Recurring SaaS plus hosting. $50K–$100K validated price points from prior Qbix sales. Customer keeps the stack. We keep the relationship.

STREAM 2

Marketplace

Workflows, capabilities, templates. Platform fee on every transaction. Network effect: more deployments → more workflows → more deployments. App Store dynamics with creator-economy tailwind.

STREAM 3

$SAFEBUX token

Utility token of the ecosystem. Operators earn it serving compute and storage. Organizations spend it on inference. Stakeholders earn cashflows. The piece that scaled Bitcoin and Ethereum beyond their early adopters.

SAFE at $5M post-money cap.

Three liquidity paths in one instrument.

STEP 1

Invest

SAFE Note. Standard SAFE at $5M post-money valuation cap. Pre-seed friendly. Standard rights.

STEP 2 · OPTIONAL

Tokenize

Convert your SAFE into $SAFE tokens via the Unblockers framework. Trade on FINRA-registered ATS after 40 days.

STEP 3 · OPTIONAL

Stake

Stake $SAFE tokens to receive cashflows from $SAFEBUX sales — the utility token of the entire ecosystem.

Sara Hanks (CrowdCheck, author of Regulation S at the SEC) advising on legal structure. Three-agreement enforcement loop with the Unblockers custodian.

Did not pivot into AI infrastructure.

Built the substrate it needs.

QBIX

2011 · $300K raised

Open-source web components. Millions of users in 100+ countries. Predated Mastodon by five years on federated social. Sold $50K–$100K per deployment.

INTERCOIN

2018 · $900K raised

Open-source smart contracts deployed across 8 EVM mainnets. Tokenization, payments, DAOs. The on-chain substrate Safebox anchors to.

SAFEBOTS

2026 · Now raising

Sealed environment. Hosts every open-source platform and open-weight model. The three layers that make the rest work.

Greg Magarshak · Founder & Chief Architect

Concert pianist (Carnegie Hall). Entered college at 14. M.S. Mathematics, NYU Courant. Teaching AI at IE University NY. Author of seven arXiv papers (PLT, KV, LAWS, Magarshak Machine, Intercloud, Towers, Grokers). Multiple provisional patents on capability-partitioned workflow execution and hardware-attested policy execution.

Applied: Y Combinator · ERA NYC · a16z Speedrun SR007. Acceptance into any compresses timeline 12–18 months.

Pre-seed open

Let's build this together.

The shift that put an autonomous network underneath banking is starting now in AI. SAFE at $5M post-money cap, with optional conversion into tokens and optional staking for $SAFEBUX cashflows.

Schedule a conversation

team@safebots.ai