AAA AI

INVESTOR PITCH · JUNE 2026

Put expert AI to work.

Enterprise multi-agent platform in production: orchestrate models, tools, memory, and human approvals while keeping control of your data. 122 integrations · 682 actions · 10 expert modules · 12 verification modes · nine client surfaces.

aaaai.me · web.aaaai.me

Web | iOS · App Store | Android | macOS Chat & IDE | CLI TUI | Desktop agents | Telegram · Discord · Slack

PRODUCTION
AWS stack with blue-green deploys, Neptune graph, Redis, S3, Paddle billing, and usage limits in product.

CONTROL
Ollama · MLX · Llama.cpp offline; self-host / VPC; workflow approvals, run logs, and agent-team lanes.

JUNE 2026
CLI TUI with approvals, orchestration task board, biometric offline on iOS/macOS, AWS Startups support.

Expert AI, under your control. Production-ready platform — live demo is primary.

122	682	$20	9+CLI
Live integrations	Catalog actions	Pro / month	Client surfaces

THE PROBLEM

Knowledge work is fragmented—and AI makes it worse

$9,400	9 apps	23 min	3.5 hrs
Lost per employee / year to tool overload	Switched between daily (avg. knowledge worker)	Refocus time after each context switch	Productivity lost per day to switching

Context switching

Work is fragmented across many applications; each context switch imposes cognitive and latency costs before a separate LLM interface is even introduced.

Chat isn't a system of record

General-purpose assistants rarely maintain durable state, enforce cross-checking among hypotheses, or embed cleanly in governed enterprise processes.

No continuity

Project agreements, communication norms, and decisions are not persisted as organizational memory; context must be reconstructed and re-communicated.

Where does the data go?

Regulated industries need more than policy language forbidding pastes into consumer chat UIs—they need executable paths (local, VPC, self-host) that security can formally accept.

CISO view: policy-based routing by data class; isolation where required; comparable employee UX, or shadow-AI adoption of public tools rises.

THE PROBLEM

Buyers need governed AI—not another chat tab

Continuation of the four-card diagnostic slide—audience definition and cited third-party work.

Stakeholders: professionals, SMBs, enterprises, developers, and mobile-heavy roles require AI *embedded* in operational workflows (memory, multi-expert patterns, local or sanctioned cloud, web/desktop/channels, EN+RU go-to-market). **Enterprise:** a single workspace span (chat, workflows, tasks, IDE, bots), **customer-defined** data boundary, productized integrations. Some pilot designs target directional multiples (~2–4×) on repetitive knowledge work; figures are illustrative, not guarantees.

> **McKinsey (2025):** on the order of **$87B/year** in productivity friction—context switching, knowledge search, repetitive cognitive work; the constraint is often *not* model access but coordination cost. **At-scale** enterprise GenAI adoption remains near **~23%**—**trust, data security, and operations** gate spend ahead of UI novelty.
>
> **EY (enterprise AI / trusted-AI programs):** pilots stall at production exit when procurement and boards ask for **governance** over data and models, access control, sub-processor transparency, and **audit-ready** traces of production runs—without a governed stack, budgets rarely convert to durable load.

THE SOLUTION

aaaai.me — one stack under your data policy

Multi-expert reasoning (specialized modules and critic), durable memory, workflow and task-board constructors, IDE, **Go CLI**, and desktop agents—one server-side control plane for on-device models and for cloud providers approved under the customer's policy.

Multi-expert reasoning over single-model heuristics

Domain-specialized modules (math, code, risk, retrieval) with cross-critique and verification—reducing the incidence of high-confidence incorrect outputs.

Isolatable execution path

Ollama / MLX on workstation or customer rack; or API keys to providers you already approve—no mandatory lock-in to our inference layer.

Memory & graph

Vector store plus optional knowledge graph (e.g. Neptune): continuity without relying solely on a linear message transcript.

THE SOLUTION

Platform properties **(continued)**

Enterprise patterns

SSO-ready posture, team boundaries, CRM/ERP/support connectors—less bespoke glue code per AI initiative.

Operator-grade controls

Human-in-the-loop steps, scheduled agent teams, execution history—closer to governed runbooks augmented by models than to unstructured, one-off prompting.

Model & cost flexibility

Mix small local models with frontier APIs per task; throttle spend via routing—supports spend governance for platform economics teams.

Shipped, production-ready product—live demo is primary; this deck is supplementary, not a substitute.

THE PRODUCT

What the integration includes

Chat
10 built-in experts with 12 verification modes—debate, critic-reviewer, self-consistency, red-team, and more.

Workflows
Visual builder, scheduled runs, agent chains—used internally and in prod.

Task board
Kanban: a pipeline stage can host a governed LLM step with an approval policy.

Live vision
Camera + voice path; on-device where the model allows.

IDE
macOS coding surface: multi-agent loop, git, terminal—production-grade tooling rather than a lightweight demo editor.

Server agents
Agents you host for cron-style or monitored jobs.

Phone MCP
18+ iPhone tools (telephony, SMS, maps, deep links)—operational layer for recurring mobile automation workloads.

Search
Built-in web search with relevance filtering so answers cite something.

Shipped surfaces (9 + CLI)

Web App · iOS · Android · macOS Chat · macOS IDE · CLI TUI · Desktop Agent · Telegram · Discord · Slack

Also ships: **Go CLI** installers (`agi` / `aaaai`) — terminal/agent workflows on the same backend.

Concrete capabilities (what we demo)

AI Chat & Experts

- 10 built-in expert modules with 12 verification modes—math, coding, planning, risk, search, social, vision
- Experts debate and verify each other's answers (multi-expert verification)
- Streaming responses with real-time processing stages
- Image generation (HunyuanDiT, Ollama models)
- Web search with smart relevance filtering
- 30+ languages auto-detected and supported

Workflows & Orchestration

- Visual drag-and-drop workflow builder
- Agent tasks, queries, MCP tools, conditions, approvals
- Kanban task board with AI execution per stage
- Execution history and run logs

Training & custom models

- In-product **Train Data** with Unsloth LoRA, nanoGPT, microGPT engines
- Enterprise connectors: S3, Postgres, warehouses—credentials kept in customer boundary
- Unsloth UI: LoRA/QLoRA, HF or local datasets, live loss/LR monitoring

iPhone MCP Integration (18+ tools)

- Phone calls, SMS, email — open native composers
- 50+ app deep links: WhatsApp, Telegram, Instagram, Uber, Spotify, YouTube, Zoom, Gmail, Maps...
- GPS location, clipboard, flashlight, haptics, battery
- Camera capture, on-device vision (Qwen2-VL)
- Local notifications, text-to-speech, share sheet

AI IDE (AGI Dev)

- Multi-agent coding pipeline (Agent/Plan/Debug/Ask modes)
- Syntax highlighting for 16 languages
- File explorer, diff view, terminal, git integration
- Thinking budget slider for deep reasoning
- Offline mode with on-device MLX models

What the product looks like (1 of 3)

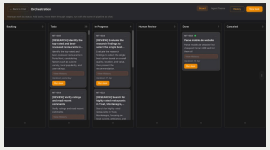
Production web—same orange/dark shell investors see in demo. Native App Store + macOS/Android/CLI on the next slide.

HOW TO READ THIS SLIDE

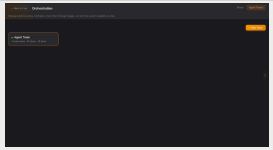
Orchestration and agent teams are part of the core product surface, not an experimental skunkworks line. Notes and chat share a unified identity and state layer so commitments are not lost in an unstructured message stream.

Web · main chat
Tabs: Orchestration, Training, Workflows, Notes. Model picker (e.g. local llama3.2), web/voice/vision actions, remote Mac/PC toggle.

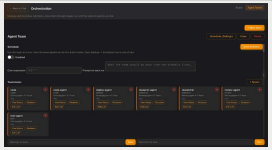
Chat · sessions & scaling
Agent-team threads, voice history, EN/RU, Final Answer mode, MCP marketplace link, expert/cognitive scaling controls.



Orchestration · board
Kanban by status (Backlog→Done). Tasks run the same AI pipeline as chat; human-review column; Run task per card.



Agent teams · workload
Task counts and teammate indicators give managers aggregate visibility into load—grounded telemetry rather than an isolated conversational channel.



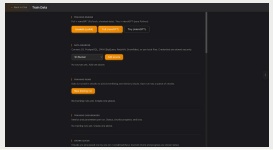
Agent teams · config
Scheduled runs (cron), prompt per run, grid of role agents (code, test, deploy, research…), llama-cpp models, spawn/edit/shutdown.

Notes · calendar
Quick capture + week view; ties conversations to dated action items (voice FAB).

Training, workflows, store, App Store (2 of 3)

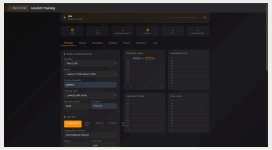
ENTERPRISE NARRATIVE

Fine-tuning remains inside your data perimeter (S3/Postgres/warehouses). Workflows combine human approvals with LLM/MCP steps—operationally closer to governed internal automation than to a conversational UI with ad hoc integration glue.



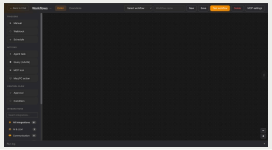
Train Data

Engines: Unsloth LoRA, nanoGPT, microGPT. Data sources: S3, Postgres, warehouses, local files—credentials stay in your boundary.



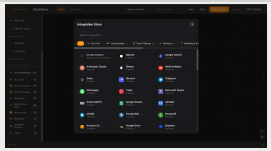
Unsloth training UI

LoRA/QLoRA on large instruct models; HF or local dataset; live loss/LR charts—fine-tune without leaving the product.



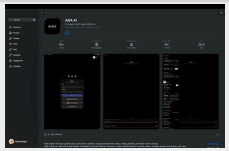
Workflows editor

Visual orchestration with triggers, approvals, agent nodes—122 integrations / 682 catalog actions in sidebar.



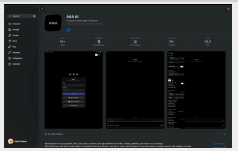
Integration store

Browse by category: OpenAI, Gemini, Claude, Ollama, Slack, Teams, DBs, S3—executable connector actions, not marketing-only screenshots.



iOS · App Store

AAA AI by Kirill Pokidov—multi-expert stack, EN+RU, offline Expert scaling in settings screenshots.



iOS · listing detail

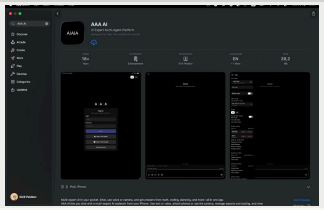
Login with Apple/Google, chat + camera/mic attachments, expert scaling and Turing-style mode in app UI.

iOS listing, macOS, Android, CLI — what ships today (3 of 3)

Full App Store page captured on macOS; native apps share the same backend contracts as web. Add per-platform UI screenshots when you refresh creative.

DISTRIBUTION

Native iOS, Android, and macOS clients extend top-of-funnel reach; the CLI serves security-conscious engineering audiences who prefer terminal-first or headless integration workflows. Entitlements and billing remain consistent across surfaces.



iOS · App Store (live listing)

Multi-expert positioning, iPad preview strip (sign-in, chat, expert scaling drawer). Same product as TestFlight/production API—screenshots from the public store.

macOS · Chat (AGIChatMac)

- ✓ SwiftUI app: chat list, streaming replies, offline store, voice/live paths
- ✓ Remote session + MCP hooks aligned with iOS; production builds beside web

macOS · IDE (AGI Dev)

- ✓ Multi-agent coding, file tree, diff, terminal, git, HTTP/DB side panels
- ✓ MLX offline models; 16-language highlight; ships as its own macOS target

Android (Compose)

- ✓ Same auth/chat APIs as iOS; offline chat persistence + ML Kit OCR on images
- ✓ Jetpack Compose UI; slot for on-device LLM via LocalModelService when you wire it

CLI (`agi` / `aaaai`)

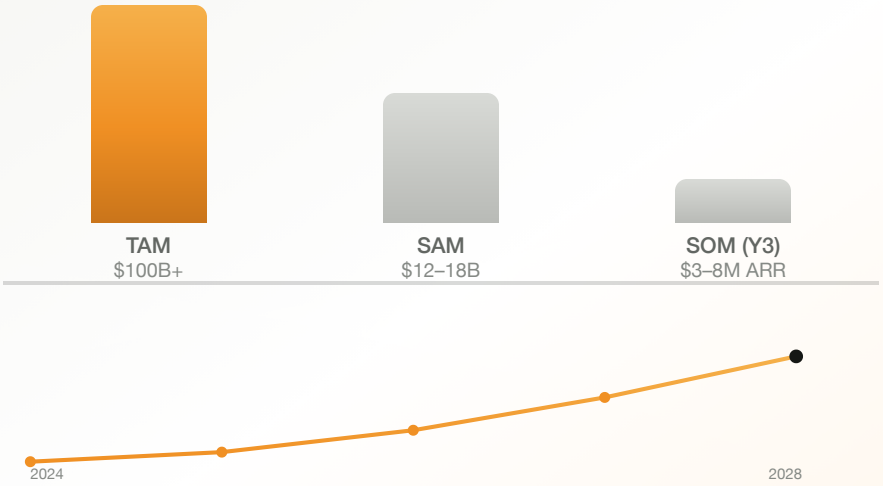
- ✓ Go binary: interactive agent TUI, platform web-search, dispatches into repo tools
- ✓ `.pkg` / `.dmg` installers next to desktop agent; optional Python bridge to legacy launchers

MARKET OPPORTUNITY

$150B+ addressable — AI adoption still gated by trust

$100B+	$50B+	400M+	73%
Productivity software (2026)	AI tools by 2027	Knowledge workers globally	Want personalized AI at work

TAM · SAM · SOM (ILLUSTRATIVE · $ BILLIONS)



Who buys first

- **Prosumer & creators** — $20/mo Pro for full cloud stack
- **SMB teams** — 122 connectors without 12-month AI programs
- **Enterprises** — self-host, SSO, audit trails
- **Developers** — IDE, CLI TUI, MCP, system agents

Why now

- Enterprise GenAI adoption ~23% at scale — trust & ops gate spend
- $87B/yr friction from context switching (McKinsey 2025)
- Buyers want **one workspace** — not 9 apps + ChatGPT

SALES NARRATIVE

Clear ROI: $240/yr Pro vs. $9,400/yr lost productivity

COST OF INACTION

$9,400

Per employee per year lost to tool overload and context switching — before counting AI rework, compliance risk, and shadow-AI spend.

AAA AI PRO

$240

Per seat per year ($20/mo) — full cloud LLMs, 122 integrations, workflows, agent teams, and priority support.

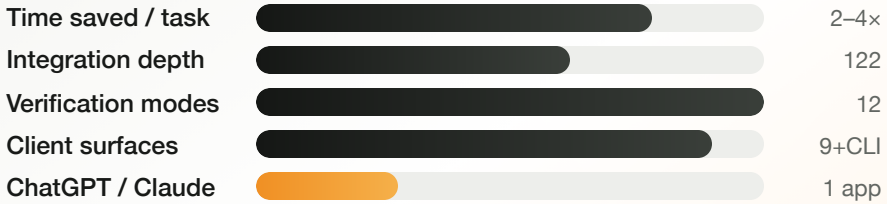
39×

cheaper than status-quo friction alone.

ILLUSTRATIVE PLG FUNNEL (DIRECTIONAL)

100K	10K	$2.4M
Free signups	Pro converts · 10%	ARR at 10K Pro
Local models, offline, basic workflows — zero CAC entry	Cloud LLMs + integrations unlock at $20/mo	Before Team/Enterprise upsell (+ install fees)

VALUE DRIVERS VS. GENERIC CHAT



Illustrative multiples from pilot designs; not a guarantee. Replace with customer metrics in diligence.

Freemium → Pro ($20/mo) → Enterprise

Free

$0

- ✓ Local AI (Ollama/MLX)
- ✓ Offline on all devices
- ✓ Basic workflows & board
- ✓ 1 system agent
- ✓ Community support

Pro

$20/mo

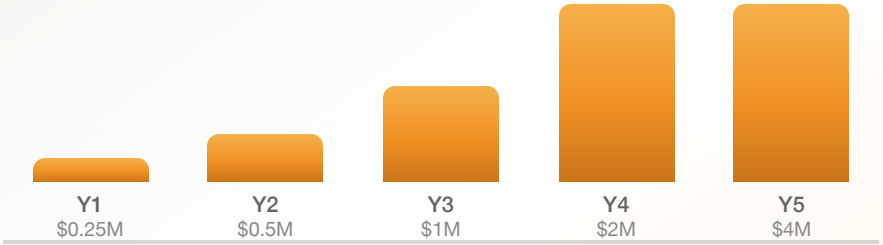
- ✓ All cloud LLMs
- ✓ 122 integrations · 682 actions
- ✓ Web search + image gen
- ✓ Unlimited agents & MCP
- ✓ Priority support

Team / Enterprise

Custom

- ✓ Self-host / Terraform / AWS
- ✓ SSO & team admin
- ✓ Custom connectors & training
- ✓ SLA & dedicated support
- ✓ On-prem GPU inference

ILLUSTRATIVE ARR PATH · ~100% YOY (SCENARIO · $M)



Scenario only — replace with your financial model. Full detail in AAAAI_Financial_Requirements.pdf.

Unit economics

- **Pro ARPU:** $240/yr · gross margin ↑ when local/Ollama handles volume
- **Enterprise:** install fee + per-seat — non-linear ARR on top of PLG
- **Expansion:** seats → fine-tuning → private connectors

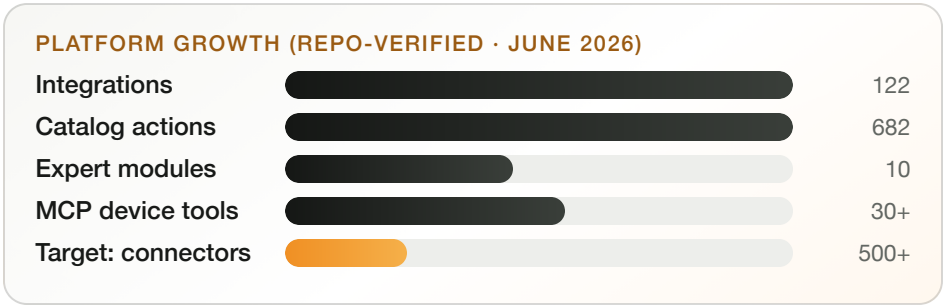
Use of funds

Product **40%** · Acquisition **30%** · Infra **20%** · Team **10%**

TRACTION

Built, shipped, production-ready — June 2026

9	122	682	12
Client surfaces	Live integrations	Catalog actions	Verification modes



Shipped milestones

- **Jun 2026:** CLI TUI, orchestration board, biometric offline
- **May 2026:** AWS Startups credits on production stack
- **Platform:** AWS · Neptune · Redis · S3 · Paddle billing
- **Surfaces:** iOS App Store, Android, macOS, Web, bots

In production today

- ✓ Multi-expert AI + workflow editor
- ✓ Train Data / Unsloth fine-tuning UI
- ✓ Telegram, Discord, Slack bots
- ✓ Image gen + web search pipeline

COMPETITIVE LANDSCAPE

Repo-verified vs. ChatGPT, Copilot, n8n — June 2026

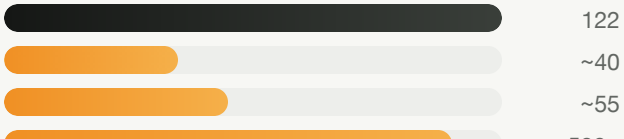
Capability	ChatGPT / Claude	MS Copilot	n8n / Zapier	AAA AI
Multi-expert verification	Single model	Single model	No AI core	**10 experts · 12 modes**
Persistent memory	Projects	M365	None	**FAISS + graph**
Offline / local inference	Cloud-first	Cloud-first	Self-host runner	**Ollama · MLX · Llama.cpp**
Visual AI workflows	Agent beta	Copilot Studio	**Strong**	**AI nodes + 122 connectors**
Business integrations	Plugins	M365 Graph	**500+ apps**	**122 · 682 actions**
Native mobile + desktop	Apps	Office	Web-first	**9 surfaces + CLI**
Agent team lanes	No	No	No	**Research · code · deploy**
Human approval gates	Limited	Policy	Manual	**Workflow + CLI TUI**
Self-hosted deployment	SaaS only	Azure-bound	n8n self-host	**Docker · Terraform · AWS**
System / device agents	No	No	No	**Desktop + 30+ tools**

Verified against AAA AI codebase (integration_registry.py, experts.py, agent_team.py) and public competitor offerings as of June 2026. Partial = available in some tiers or with constraints.

WHY WE WIN

Moat: breadth + governance + switching costs

Integration depth vs. peers



AAA AI combines n8n-class automation with multi-expert AI core — competitors lack one or the other.

Switching costs

Accumulated workflows, memory, and integrations raise migration friction — retention comparable to team SaaS when deployed cross-functionally.

Sustainable edge

- **9 surfaces + CLI** on one expert core — years to replicate
- **Local + cloud** routing under customer policy
- **Connector roadmap:** 122 → 500+ with raise capital

INVESTOR Q&A

Questions we get in first meetings

Part 1 of 2 — product, monetization, and data boundary; next slide: round timing, team, risks, and diligence.

Why not restrict to ChatGPT / Copilot?

Those offerings typically deliver a monovendor chat experience in the browser. AAA AI implements **multi-expert reasoning, memory, and workflow orchestration** across web, mobile, desktop, and messaging surfaces—with **local processing priority** and bring-your-own API keys where policy requires.

How do you make money?

Freemium → **Pro ($20/mo)** → Team / Enterprise (self-host, SSO, custom). Full tiering and ARR scenarios are in the financial appendix PDF.

Where does data live?

Customer choice: **Ollama / on-device** for zero cloud leakage, or cloud LLMs with your policy. Enterprise path for VPC / self-hosted deployment.

What is actually shipped today?

Production stack on AWS, **nine client surfaces + CLI**, billing (Paddle, Apple), workflow scenarios, IDE, and bots—aligned with the Traction slide.

Questions we get in first meetings (continued)

Why raise now?

To **scale GTM**, deepen enterprise features, and harden infra while AI adoption curves are steep — use of funds: product · acquisition · infra · team (see Join Us slide).

Team & depth?

Kirill Pokidov — **more than 15 years** at VP / Chief Cloud Architect level (teams **15–50+**, AWS/Azure/GCP, venture-backed platform engagements) *and* lead architect–implementer of the full AAA AI stack (Traction slide).

Investor relevance: demonstrated execution, enterprise infrastructure discipline, and capacity to scale hiring post-round—extended biography and investment rationale on the Team slide of the **financial appendix**.

Technical risk & mitigation

LLM quality/cost volatility is real; mitigation is **multi-provider routing**, local fallback, and expert+critic patterns that reduce single-model failure modes—plus continual eval on coding and math tracks.

Focus areas for diligence

Infra cost curve vs MAU; conversion Free→Pro; support load at scale; enterprise security review checklist—mapped to KPIs in financial **Roadmap** slide.

Numbers (TAM/SOM, illustrative ARR): **AAAAI_Financial_Requirements.pdf** — we keep finance there on purpose.

JOIN US

Deploy your first expert workflow.

Capital funds connector density toward 500+, orchestration scale, and systematic adoption among 400M+ knowledge workers—the platform core is already in production.

$4M	500+	39×	$20
Illustrative Y5 ARR	Connector target	ROI vs. tool friction	Pro entry price

Use of funds: Product ~40% · Acquisition ~30% · Infra ~20% · Team ~10%

30–60 days
Data room index, security one-pager, pilot success-metrics template aligned to Cases slide.

Process
Partner intro → live demo (web + mobile) → model & security deep dive → commercial term alignment.

Ask
Raise sized to lengthen runway through enterprise GTM experiments—see financial appendix for scenario math.

aaaai.me

hello@aaaai.me · web.aaaai.me