

Local AI that acts inside real software.

The local runtime layer for agentic software.

Run AI agents inside games, simulations, enterprise tools, robotics systems and mission-critical environments without cloud latency, recurring API costs or loss of control.

PRODUCTION-PROVEN

PATENT-PENDING

\$3M SEED

Cloud AI can answer. Pontex lets agents act locally.





Local models can think. Real software still needs a way to let them act.

Developers are forced to stitch together inference, memory, speech, retrieval, scheduling and tool execution just to connect AI safely to live software.

Smaller Models

Capable AI is moving onto the device.

Stronger Hardware

GPUs, CPUs and NPUs can run increasingly powerful AI locally.

Demand for Control

Lower latency, privacy, offline operation and reduced cloud dependency.

**Running a model locally is not enough.
Software still needs an execution layer.**

**The model layer is becoming local.
The software layer is becoming agentic.
Pontex sits at the intersection.**

Pontex is built for that shift.



Pontex embeds AI agents where decisions and actions happen.

AI

Models · Speech · Memory · Retrieval

**PONTEx RUNTIME**

Reason · Remember · Route · Validate · Execute

**SOFTWARE**

Games · Simulations · Enterprise · Robotics

**CONTROLLED ACTION**

Tools · Workflows · Commands · Environment Changes

LOCAL

REAL-TIME

OFFLINE-CAPABLE

HARDWARE-AGNOSTIC

HOST-CONTROLLED

Cloud AI can answer. Pontex lets agents act locally.



Not a prototype. Not a pitch.

Pontex is live in production.

PRODUCTION DEPLOYMENT

First production deployment inside the founding team's TCG World Unity HDRP environment.

Full local agent loop

Voice → Reasoning → Memory → Tool Call → Environment
Action → Response

LIVE DEMO AVAILABLE

FEB 2026

LIVE PRODUCTION

TCG World deployment

~0.3 SEC

VOICE LATENCY

measured local loop

~4 SEC

TOOL/ACTION

full action cycle

6

LLM FAMILIES

validated locally

4

BACKENDS

CUDA · DirectML · Vulkan · CPU

2

ENVIRONMENTS

Unity + desktop agent

Voice → Reasoning → Memory → Tool Call → Environment Action → Response

Local execution. No cloud inference required. CUDA · DirectML · Vulkan · CPU · Foundry Local compatible



Models will change. The architecture is the asset.

CORE RUNTIME IP

3 Core Patent Families

01 **Native Inference + Memory Bridge**
Local inference and zero-copy/native memory handoff

02 **Real-Time Execution**
Scheduling, hardware routing and deterministic replay

03 **Safe Agent Control**
Speech, RAG, structured output and controlled tool execution

Defensible infrastructure, not a commodity wrapper.

STRUCTURAL MOAT

Production hardening, cross-stack portability and host-native control create a defensible execution layer.

DEFENSIBLE BY DESIGN

PRODUCTION

PROVEN

MODEL + HW

AGNOSTIC

HOST-NATIVE

CONTROL

Partners integrate once, preserve model and hardware choice, and avoid rebuilding the execution layer internally.

COMPOUNDING ADVANTAGE

Built to remain the execution layer as models, hardware and host environments evolve.



Model serving is not agent execution.

APPLICATION

Games · Simulation · Enterprise · Robotics

**PONTEX**Host-Integrated Agent Execution
State · Memory · Scheduling · Safe Actions**MODEL SERVING**

Foundry Local · llama.cpp · Local Runtimes

**MODELS + HARDWARE**

Open Models · GPU · CPU · NPU · Edge

MODEL-AGNOSTIC

Choose the right model for the deployment.

HARDWARE-AGNOSTIC

Run across available local compute.

HOST-NATIVE

AI operates inside the actual software environment.

Pontex is the host-execution layer for local AI.



07 | BUSINESS MODEL

Pilot → License → Integrate → Deploy

1 · PILOT

Focused integration in the customer's environment

**2 · LICENSE**

Commercial runtime / SDK license

**3 · INTEGRATE**

Embedded runtime layer across products

**4 · DEPLOY**

Strategic infrastructure deployment

REVENUE MODEL**PRODUCT LICENSING**

SDK · per-product · per-title

ENTERPRISE DEPLOYMENT

licensing · support · implementation

STRATEGIC INTEGRATION

platform · field-of-use · long-term agreements

Pontex can be licensed as a product, integrated as a platform layer, or deployed as strategic infrastructure.



A massive market is moving toward local, agentic software.

\$47BGLOBAL EDGE AI
2026**\$446B**EDGE AI PROJECTED
2034**UP TO 40%**ENTERPRISE APPS
WITH TASK-SPECIFIC AI AGENTS BY 2026

Pontex targets the execution layer inside that shift.

GAMING + SIMULATION

Games · Engines · Digital worlds

ENTERPRISE SOFTWARE

Private agents · Workflows · Applications

DEFENSE + SECURE SYSTEMS

Simulation · Air-gapped systems · Mission environments

INDUSTRIAL + ROBOTICS

Automation · Digital twins · Physical AI

HARDWARE + OEM

PCs · Edge devices · Embedded platforms

SOVEREIGN AI

Private · Local · Controlled deployments

500+ STRATEGIC LICENSING TARGETS

Initial strategic universe across software platforms, game studios, OEMs, defense, industrial and sovereign AI.

One runtime architecture. Multiple high-value licensing markets.



We built the technology before we raised the capital.

**DAVID EVANS****Co-Founder & CEO**

Leads strategy, partnerships and commercialization from the production environment where Pontex was first proven.

**JUSTIN DEL GIUDICE****Co-Founder & COO/CFO**

Leads finance, operations, corporate execution and fundraising for Pontex.

**RICHARD SANTOS****Co-Founder & Technical Inventor**

Architected and shipped the Pontex runtime and leads next-generation runtime research. Inventor across the core runtime IP.

Founder-built. Product-first. Already executing.

Capital accelerates Pontex. It does not create it.



\$3M Seed · Post-Money SAFE · \$50M Valuation Cap

Scale a proven runtime into a category-defining AI infrastructure platform.

SCALE THE PLATFORM

- Commercial SDK/runtime
- Developer tooling
- Multi-platform integration
- Enterprise/security deployment

DRIVE DEPLOYMENT

- Design-partner integrations
- Production licenses
- Repeatable licensing motion
- Strategic platform relationships

EXTEND THE MOAT

- Core patent prosecution
- Runtime Architecture IP
- Production Hardening at Scale
- Next-generation local AI architecture

18-24 MONTH OUTCOMES

COMMERCIAL SDK

deployed externally

SALES MOTION

production licensing

MULTI-PLATFORM

runtime deployments

IP PORTFOLIO

expanded protection

ENTERPRISE READINESS

Security + deployment hardening

We proved we can build it. Now we are raising to scale it.

The companies that control local agent execution will control a critical layer of the next AI stack. Pontex is built for that layer.