

ANIMA CORE INC

AN1

Meaning-First Compute for Neural Networks

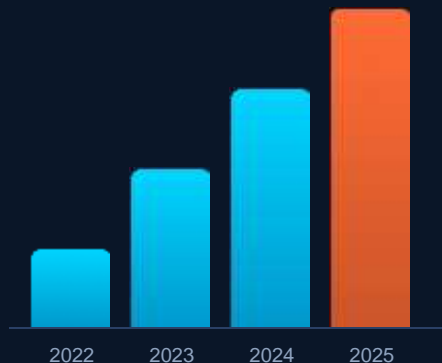
Ryan Shamim

Founder, Anima Core Inc.



The cost of inference is exploding, and current compute scales inefficiently.

- Transformer inference is dominated by heavy matrix multiplications
- Costs rise linearly with model size, context length, and sequence steps
- Enterprises cannot sustain the energy and hardware footprint
- Early exits radically degrade accuracy

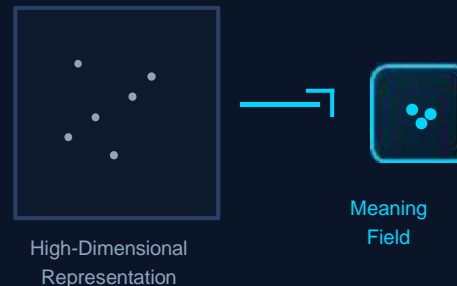


Enterprise AI Compute Costs

There is massive unused structure inside neural networks.

- Learned representations contain high-dimensional redundancy
- Only a small fraction of the model's internal dimensions determine predictions
- Capturing this structure unlocks a new compute pathway

72-99% of task-relevant variance exists in just 1-3 dimensions



AN1 extracts meaning fields and computes using compressed semantic space.



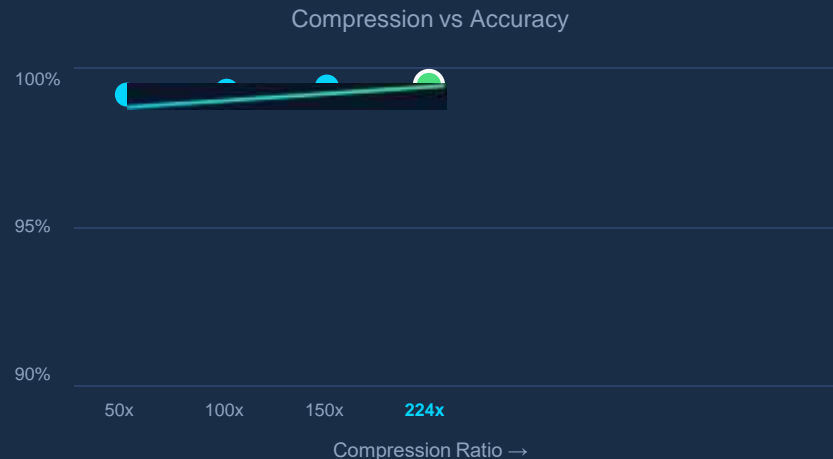
- Extracts the model's internal 'meaning field'
- Predicts using a tiny symbolic head

Field compression preserves accuracy across tasks.

- 224x compression with equal or higher accuracy on multiple benchmarks
- 25x speedups on inference workloads
- Robust across seeds and architectures
- Works at 70B scale

VALIDATED ON

Llama-70B, CLIP ViT-L, GPT-2, T5, ResNet18, and more



What this means for your infrastructure.

EXAMPLE SCENARIO

4M

inference calls per day

Baseline annual cost: **\$5.8M**

AN1 annual cost: **\$1.1M**

ANNUAL SAVINGS

\$4.7M

- Latency reduction: 25-40%



Enterprises need reliable, low-cost inference. AN1 makes it possible.



No architecture changes required

Drop-in replacement for existing inference



Works on your hardware or our managed cloud

Flexible deployment options



Enables LLM scale without LLM cost

70B capability at a fraction of the price



Reduces carbon footprint

Sustainable AI infrastructure

ENERGY IMPACT



80%

reduction in compute energy

Q1 2025 pilots available.

- 1 2-3 pilot partners
- 2 Plug-in inference API or drop-in microservice
- 3 Live cost calculator
- 4 Dedicated engineering support
- ✓ Measured savings delivered in first 30 days

LIMITED AVAILABILITY

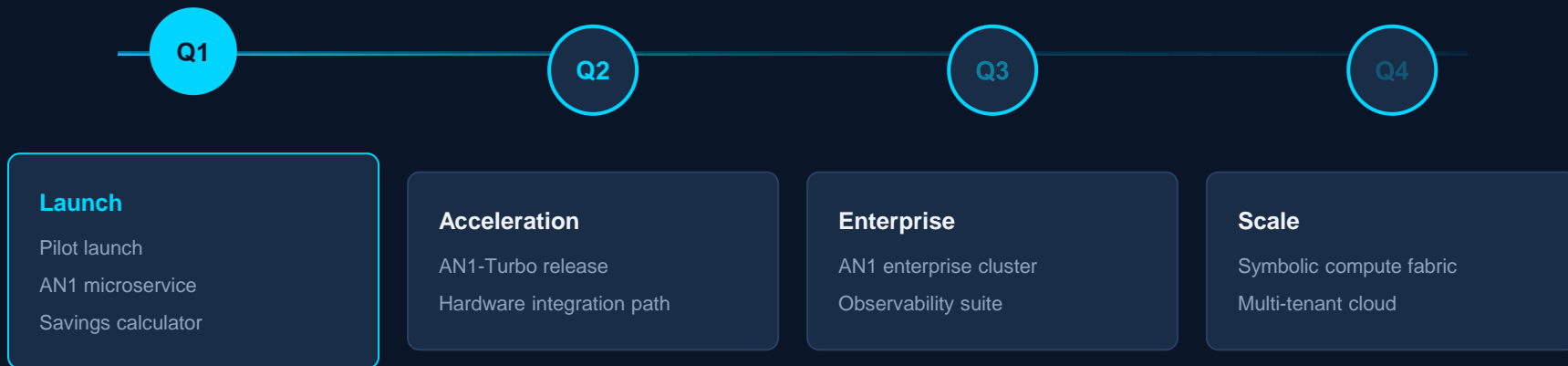
2-3

pilot slots for Q1

Join the Pilot

partner@animacore.ai

2025 Product Timeline



Usage-based + enterprise licensing.

USAGE-BASED

Pay Per Call

Per-token pricing

Per-call pricing

No minimum commitment

SAVINGS SHARE

Performance-Linked

% of measured savings

Aligned incentives

Risk-free adoption

ENTERPRISE

Private Cluster

Private cluster licensing

On-premise deployment

Dedicated support

- Pricing scales with value delivered
- Transparent cost calculator
- Custom enterprise agreements

Why AN1 cannot be easily replicated.

1

Proprietary meaning-field extraction

2

Novel compression architecture

3

Symbolic compute head

4

Validated across tasks and model families

5

Strong performance at 70B scale

PATENT PROTECTED

Provisional patents filed covering meaning-field extraction and symbolic compute methods

The future of compute is meaning-first.

- AN1 becomes the semantic layer for modern AI
- Massive reduction in global compute waste
- Foundation for future symbolic hardware (AN1-G and AN1-S chips)
- A new intelligence architecture centered on meaning

"What if intelligence isn't about scale—it's about structure?"



Meaning at the Core



AN1

Meaning-First Compute for Neural Networks

Ryan Shamim

Founder, Anima Core Inc.

partner@animacore.ai

animacore.ai

[Request Pilot or Investor Call](#)