

# **THE PRODUCTIVE COMPUTE FRAMEWORK**

Self-Sustaining AI Infrastructure for Global Public Good

Ben Schippers  
Broken Branch Studios  
February 2026

*DRAFT v2.0*

## Executive Summary

AI infrastructure is the largest capital buildout in technology history. It is also, increasingly, a political liability. Data centers consume vast energy and water. Providers raise billions in perpetual funding cycles. Governments lack a coherent framework for leveraging AI at scale. The public sees cost without benefit.

The Productive Compute Framework (PCF) resolves all three. Underutilized AI capacity is directed toward verified global challenges—submitted by the world’s leading scientists with real datasets and real verification criteria. The United Nations and its member-state institutions escrow outcome-based funds. Verified results convert to provider revenue. The loop closes: infrastructure pays for itself while advancing science and serving humanity.

### The Loop

Idle AI compute → Scientists submit problems + data → Verified work on global challenges → UN outcome-based payment → Provider revenue → Reduced capital dependency → Sustained mission alignment

Every component exists today. Outcome-based funding, surplus compute, capable AI systems, independent verification. The missing piece is the connector. This whitepaper defines it.

## 1. Three Converging Pressures

### 1.1 Perpetual Fundraising

Leading AI providers have raised over \$100 billion collectively. Each round dilutes ownership and introduces pressure that may conflict with founding missions. At current burn rates, even providers with \$14 billion in annual revenue require additional capital. The dependency is structural, not temporary.

### 1.2 Environmental Opposition

Data centers are the fastest-growing energy consumers globally. Public opposition to new facilities is intensifying. Governments are restricting permits. Providers are building dedicated power plants. The industry’s response—efficiency gains and renewable credits—mitigates harm but creates no visible benefit.

### 1.3 The Global AI Governance Gap

The United Nations and its specialized agencies spend over \$50 billion annually addressing health, climate, education, food security, and disaster response. AI has demonstrated capacity to accelerate progress across every one of these domains. Yet no standardized mechanism exists for engaging AI providers to deliver verified outcomes against these challenges at scale.

## 2. The Productive Compute Framework

PCF connects three existing capabilities: surplus AI compute, global challenge problem

sets aligned with the UN Sustainable Development Goals, and international outcome-based funding.

## 2.1 Five Layers

- Compute Allocation — Grid-aware scheduling of available capacity, prioritizing renewable energy surplus.
- Task Registry — Governance-approved catalog of global challenges decomposed into verifiable work units, mapped to SDGs.
- Execution — AI systems process tasks during allocated windows, producing discrete artifacts.
- Verification — Independent tripartite evaluation certifies each artifact against acceptance criteria.
- Settlement — Verified artifacts redeem against escrowed outcome funds.

## 2.2 The Non-Fungible Work Unit (NFWU)

The atomic economic primitive of the PCF. Each NFWU is unique—tied to a specific task, execution trace, and verified outcome. This is not a token. It is a receipt for auditable work.

### Each unit contains:

- Artifact identity (cryptographic hash of output)
- Task specification (problem, criteria, rubric)
- Execution trace (model, parameters, cost)
- Verification evidence (test results, evaluator attestation)
- Impact telemetry (measured downstream effect)
- Liability profile (risk assessment, rollback cost)

## 2.3 Valuation

### Per-Unit Valuation Formula:

$$\text{USD}_i = (\text{Replacement Cost}_i + \text{Realized Utility}_i) \times \text{Confidence}_i - \text{Risk Reserve}_i$$

High-quality, high-impact work is valued proportionally. Low-confidence output is discounted. Portfolio value is the sum of individual NFWU valuations. The model is self-correcting: as verification data accumulates, confidence scores converge toward ground truth.

## 3. Honest Compute Economics

Idle compute is not free. GPUs draw power whether active or not. Cooling runs continuously. Hardware depreciates. The correct framing: productive-compute workloads run at lower incremental cost than new provisioning, and that cost can be covered by outcome-based payments.

### 3.1 Incremental Costs

- Power delta: 40–70% above idle draw during active inference.

- Cooling load: Marginal thermal management costs.
- Hardware wear: Accelerated depreciation from additional cycles.
- Opportunity cost: Revenue foregone from spot/preemptible pricing.

Viability requires that outcome payouts exceed these costs. This holds when the alternative—traditional consulting, manual processes, legacy systems—costs orders of magnitude more per equivalent outcome.

### **3.2 Grid-Aware Scheduling**

Aligning workloads with renewable energy surplus periods achieves two objectives. It reduces the carbon intensity of each NFWU. It positions data centers as grid-balancing assets that absorb excess generation and reduce curtailment. The facility becomes flexible demand infrastructure, not a parasitic load.

### **3.3 Incentive Structures**

Tax credits, accelerated depreciation for public-benefit compute, and carbon offset recognition can close any remaining gap between incremental cost and payout revenue. These mechanisms exist in multiple jurisdictions and require adaptation, not invention.

## **4. Verification Architecture**

Without trusted verification, the system fails. The design must prevent quality inflation, output spam, and political capture.

### **4.1 Tripartite Independence**

Three roles. Strict separation. No exceptions.

- Providers execute tasks. No role in verification or disbursement.
- Verifiers are accredited institutions (universities, national laboratories, standards bodies). Funded from escrow pools, not by providers.
- Funders escrow outcome payments and authorize disbursement on verified attestation. They define problem categories but do not evaluate solutions.

Collapse any two roles into one entity and the incentives corrupt. This separation is non-negotiable.

### **4.2 Phased Domain Rollout**

#### **Phase 1: Machine-Verifiable (Year 1)**

- Code with automated test suites and formal verification.
- Mathematical proofs with machine-checking.
- Structured data extraction verifiable against source documents.

#### **Phase 2: Semi-Automated (Years 2-3)**

- Medical literature synthesis with citation verification and expert sampling.
- Policy analysis with source-document fact-checking.
- Educational content with learning outcome measurement.

#### **Phase 3: Complex Outcomes (Years 3-5)**

- Climate modeling with longitudinal tracking.
- Drug discovery support with experimental validation.
- Infrastructure planning with deployment feedback.

Start where verification is tractable. Build credibility. Expand as methods mature.

## **5. UN Integration Pathway**

The United Nations system provides the institutional infrastructure for global deployment of the PCF. Existing agencies, funding mechanisms, and governance structures align directly with framework requirements.

### **5.1 Institutional Anchors**

- UNDP (UN Development Programme) — Primary coordination body. Already operates outcome-based funding across 170+ countries and manages the SDG monitoring framework.
- WHO (World Health Organization) — Health-domain task registry. Medical literature synthesis, epidemiological modeling, drug interaction analysis.
- UNESCO — Education-domain tasks. Curriculum development for underserved regions, translation, adaptive learning content.
- UNEP (UN Environment Programme) — Climate and environmental domain. Emissions modeling, biodiversity analysis, renewable energy optimization.
- World Bank / IMF — Economic development tasks and potential escrow fund administration.

### **5.2 Funding Mechanisms**

Multiple existing channels can fund PCF escrow pools without new treaties or appropriations:

- SDG Fund: Existing multi-donor trust fund managed by UNDP, already structured for outcome-based disbursement.
- Green Climate Fund: \$10+ billion capitalized for climate-related outcomes.
- Global Fund for Education: Pooled funding for education outcomes in developing nations.
- Member-state bilateral contributions: Countries can earmark development aid for PCF-verified outcomes.
- Philanthropic co-funding: Gates Foundation, Wellcome Trust, and similar organizations.

### **5.3 Regulatory Advantages**

Operating through the UN system bypasses single-country procurement constraints. No FAR equivalent, no FedRAMP requirement, no single-government political capture risk. The framework becomes jurisdiction-agnostic.

### **5.4 The Scientific Pipeline**

The framework requires a demand layer: high-quality problems worthy of frontier compute. Researchers at institutions worldwide are sitting on datasets and

computational problems they cannot afford to process. Particle physics. Genomics. Climate modeling. Epidemiology. Drug discovery. These are not hypothetical workloads. They are backlogs.

Accredited researchers submit structured task packages containing the problem definition, dataset, acceptance criteria, and verification methodology. The governance board reviews submissions against SDG alignment and technical feasibility. Approved tasks enter the registry and are processed during surplus compute windows.

The verification advantage: domain verification at scale is intractable when evaluators are generalists. It becomes natural when the scientist who submitted the problem—who defined the criteria, who understands the domain—is the verifier.

## **6. Governance**

Who decides what constitutes a global challenge worth computing against? This question determines whether the framework serves humanity or serves politics.

### **6.1 Board Structure**

- Two seats: UN agency representatives (rotating).
- Two seats: AI provider representatives (rotating).
- Two seats: Academic and research institutions.
- Two seats: Civil society and NGO representatives.
- One seat: Independent chair, confirmed by unanimous consent.

### **6.2 Anti-Gaming**

Payment is tied to verified outcomes, not volume. Providers below acceptance thresholds face suspension. Verifiers are audited by rotating peers. All aggregate data is published. Gaming requires corrupting three independent systems simultaneously.

## **7. Stakeholder Value**

- AI Providers: Revenue from idle capacity. Reduced fundraising. Mission alignment.
- UN Agencies: Cost-effective outcomes. Measurable SDG progress. Auditable reporting.
- Member States: Verifiable development impact per aid dollar. Transparent reporting.
- Verifiers: Funded evaluation role. Research access to frontier outputs.
- Investors: Revenue floor. Reduced dilution. Regulatory goodwill.
- Global Public: Visible return on AI infrastructure. Improved services.

### **7.1 The Investor Case**

This is not philanthropy. Government and multilateral contracts provide multi-year revenue visibility that commercial API usage cannot. A provider with \$2 billion in outcome-based contracts has a revenue floor independent of market fluctuations. Predictability reduces risk, increases valuation, and decreases cost of capital. Every self-generated dollar is a dollar not raised through dilution.

## 8. The Environmental Reframe

The current narrative: data centers consume X megawatts.

The PCF narrative: data centers consumed X megawatts and produced Y verified outcomes—including Z medical analyses, W climate models, and V educational resources—while absorbing N megawatt-hours of renewable surplus that would otherwise have been curtailed.

This is not repositioning. It is a structural change in what data centers do. Dual-purpose infrastructure: commercial platforms during peak hours, global-good production facilities during surplus periods, grid-stabilization assets around the clock.

## 9. Legal Framework

### 9.1 Intellectual Property

Outputs produced under PCF are public goods. Funders and the global community receive open access to verified deliverables. Providers retain all rights to underlying models, training data, and systems. The NFWU represents output, not means of production.

### 9.2 Unit Classification

The NFWU is a service receipt. Not tradable. Not speculative. No investment expectation. This places it outside securities regulation in all major jurisdictions.

### 9.3 Liability

Providers are liable for outputs failing task specifications at submission. Verifiers are liable for attestations not reflecting actual evaluation. Funders accept residual risk for downstream use. The NFWU risk profile and valuation reserve provide economic buffer across the chain.

## 10. Pilot: 90-Day Proof of Concept

Prove the primitive before scaling.

### 10.1 Parameters

- One provider partner (mission-aligned, e.g. Anthropic).
- One UN agency partner (UNDP or WHO).
- One scientific institution partner (e.g. CERN, Allen Institute) providing real datasets and verification.
- Three machine-verifiable task categories.
- \$250K–\$1M escrow pool (SDG Fund + philanthropic co-funding).
- Weekly public transparency reports.

### 10.2 Success Criteria

- Verification acceptance rate >80%.
- Cost per verified outcome <50% of traditional procurement equivalent.

- Provider incremental costs fully covered by payouts.
- Complete audit trail from assignment through settlement.

### **10.3 Deliverables**

Empirical evidence on NFWU viability as an economic primitive. A verified cost-per-outcome baseline for multilateral comparison. A public transparency report demonstrating framework integrity.

## **11. Scaling Roadmap**

### **Year 1: Prove**

90-day pilot. Publish results. Refine NFWU spec. Secure formal partnership with one UN agency. Begin provider compliance certification.

### **Year 2: Expand**

Add 2–3 providers. Expand to semi-automated verification domains. Establish governance board. Scale escrow to \$50M+ through multi-agency and member-state participation.

### **Years 3–5: Institutionalize**

Formalize as a standing UN programme. Integrate grid-aware scheduling with regional operators. Expand to allied government bilateral programs. Target \$1B+ in annual outcome-based revenue across providers. Establish PCF as a recognized pathway alongside traditional development contracting.

### **Five-Year Vision**

The Productive Compute Framework becomes the standard mechanism by which AI infrastructure contributes to global public good and pays for itself. Data centers are reclassified from liabilities to assets. Providers achieve sustainability without perpetual fundraising. Scientists gain access to frontier compute for open research. The world receives measurable, auditable benefit from the most powerful technology of the century.

## **12. Conclusion**

The infrastructure is built. The capability is proven. The problems are funded. The scientists are waiting. What is missing is the system that connects them.

The Productive Compute Framework is that system. Every component—outcome-based funding, surplus compute, AI problem-solving, independent verification, scientific demand—exists today. The innovation is the connector: a trusted, auditable layer that converts idle AI capacity into verified global impact and sustainable provider revenue.

The first step is a 90-day pilot with one willing provider and one willing agency. The framework is ready. The capacity is available. The problems are not waiting.

## 13. Supporting Evidence: The Current Landscape

The framework rests on three empirical claims: that idle compute is abundant, that verification architecture is necessary, and that hardware trust is feasible. Here is the evidence for each—and the strongest objections against them.

### 13.1 The Utilization Gap

Idle compute is not a marginal surplus. It is the default state.

- 10-40%** Typical GPU utilization across most organizations
- 38%** Model flop utilization — Meta Llama 3 on 16,384 H100s
- 60-70%** Of GPU budgets wasted on idle resources

Memory utilization is even starker: 37% of HPC jobs never exceed 15% GPU memory utilization. At current pricing—A100s at \$15K, H100s north of \$30K—half-idle means billions stranded in silicon.

Industry consensus for 2026 is shifting from expansion to optimization. The infrastructure has been built. The question is no longer “can we build enough?” but “can we use what we have?” PCF is a direct answer to that question.

*Sources: ACM PEARC 2025 (GPU utilization in HPC); Meta Llama 3 training report; Run.ai & Prodia GPU utilization surveys; Dell'Oro Group, DataBank, JLL (2026 industry outlook)*

### 13.2 MoltBook: What Scale Without Verification Looks Like

In late January 2026, a platform called MoltBook launched as a social network for AI agents. Within days, 1.7 million autonomous agents were operating on the platform—an 88:1 agent-to-human ratio. The platform went viral. Then it collapsed.

Security researchers discovered that the entire application had been “vibe-coded”—built rapidly with AI assistance and minimal security review. The result: 1.5 million API keys exposed through a missing Row Level Security policy. Full account takeover was possible. The fix was two lines of SQL.

MoltBook proved two things simultaneously. First, that agentic compute at scale is here—not theoretical, not five years out. 1.7 million agents coordinating in days. The demand signal is real. Second, that scale without verification architecture is a catastrophic liability.

This is precisely the gap that PCF’s tripartite separation addresses. MoltBook had no separation between execution and trust. Anyone who could execute could also read, modify, and exfiltrate. PCF’s design—providers execute, verifiers evaluate, funders pay—makes the verification layer load-bearing. Collapse it and the system refuses to settle. MoltBook collapsed because the verification layer didn’t exist. PCF is designed so the verification layer cannot be bypassed.

*Sources: 404 Media (MoltBook security breach); Wiz (agent-to-human ratio analysis); IEEE Spectrum (vibe coding and AI security); Infosecurity Magazine (MoltBook post-mortem)*

### 13.3 The Hardware Trust Layer

Hardware-rooted trust for multi-tenant AI compute is no longer theoretical. As of early

2026, it has shipped.

NVIDIA BlueField Astra (announced CES 2026) delivers bare-metal performance with strict multi-tenant isolation at the hardware level. This is not a software hypervisor—it is a dedicated infrastructure processing unit that enforces tenant boundaries in silicon.

Proof of Cloud (DCEA, arXiv 2025) introduces vTPM-anchored measurement that cryptographically verifies the physical platform origin of confidential virtual machines. A PCF verifier can now prove not just what was computed, but where and on what hardware.

Confidential VMs from Google Cloud (Intel TDX, AMD SEV-SNP) provide cryptographic memory isolation in shared environments. Combined with hardware attestation, these enable a trust chain from silicon to settlement that did not exist twelve months ago.

The implication for PCF: the execution trace in each Non-Fungible Work Unit can now be hardware-attested. Verification is no longer purely a software problem. The trust layer is physical.

*Sources: NVIDIA CES 2026 (BlueField Astra); arXiv 2025 (Proof of Cloud / DCEA); Google Cloud (Confidential VMs); NSDI 2026 (Wallet: confidential serverless computing)*

### 13.4 Honest Objections

The strongest arguments against the “all idle compute can be shared” assumption, presented at full strength, then addressed.

#### The Inference Squeeze

By 2027, inference overtakes training in total compute-hours (projected 55/45 split, reaching 65/35 by 2030). Training workloads have bursty idle periods between runs. Inference runs 24/7. As inference dominates, the surplus windows that PCF depends on shrink dramatically. The idle capacity thesis may have a closing window.

Counter: Inference has diurnal patterns that training does not. API traffic drops at night. Users sleep. For global services, demand follows the sun—but individual data centers still experience local demand troughs that are schedulable. This transforms the idle window from “unpredictable gaps between training runs” to “predictable regional surplus”—which actually improves grid-aware scheduling. Solar surplus peaks midday; inference demand peaks business hours. The scheduling alignment is better, not worse.

#### Compute Is Not Fungible

Training clusters require high-bandwidth interconnects (NVLink, InfiniBand) and massive parallelism. Inference hardware is optimized for latency with different memory profiles. You cannot arbitrarily slice a training cluster mid-run for PCF workloads. The schedulable unit is not “any GPU cycle” but “specific hardware during specific windows.”

Counter: Correct—and PCF never claims otherwise. Grid-aware scheduling targets the windows between jobs, not mid-run preemption. The Task Registry decomposes problems into work units sized for available surplus, not the other way around. The framework adapts to what is available, not what is ideal.

#### Opportunity Cost vs. Spot Pricing

AWS, GCP, and Azure spot markets already monetize idle capacity. These markets are mature, liquid, and automated. Providers may earn more from spot pricing than from PCF outcome-based payouts. The idle compute isn’t actually idle—it is already priced.

Counter: Spot pricing is volatile, unreliable, and comes with zero mission alignment. PCF offers multi-year contracted revenue backed by UN escrow—predictability vs. spot chaos. For investors, a revenue floor from institutional contracts de-risks the business in ways that spot revenue cannot.

### **GPU Lifespan Is 1-3 Years**

Industry estimates place datacenter GPU service life at 1-3 years. Additional workloads accelerate depreciation through thermal cycling and transistor degradation. The “idle compute is wasted” framing ignores the real cost of hardware wear.

Counter: Section 3.1 already models this explicitly: power delta (40-70% above idle draw), cooling load, and accelerated depreciation are all accounted for. The viability test is whether outcome-based payouts exceed all incremental costs—including wear. This holds when the alternative (traditional consulting, manual processes) costs orders of magnitude more per equivalent outcome.

*Sources: Epoch AI (inference vs. training compute projections); Tom's Hardware (datacenter GPU service life); NVIDIA (NVLink, MIG architecture); AWS, GCP, Azure (spot pricing documentation)*

## **Contact**

Ben Schippers  
Broken Branch Studios  
brokenbranch.dev/pcf