HUSSH TECHNOLOGIES CORPORATION

🤫

The Fourth Utility
🤫 × NVIDIA — Deep Integration Technical Plan

Companion to the open letter at `hushh.ai/one/nvidia`. How sovereign supercomputing reaches every American household the way clean water, good electricity, internet, and content already do — metered per unit, billed to a subscription the family already trusts, with the 🤫 Private Agent One free for every American citizen.

Buy together. Build together. Sell together. Worked backwards from NVIDIA's own July 2026 roadmap, read from primary sources, cited throughout. An honest proposal — not a claimed deal, endorsement, or partnership.

THE ONE-LINE THESIS

Compute is the fourth household utility. The PCHP handshake is its meter. NVIDIA is its generator. 🤫 One is the reason a family turns it on.

WHAT'S INSIDE

Contents

How to read this document

This is the technical companion to the founder's open letter at `hushh.ai/one/nvidia` — the letter carries the heart; this carries the thick and thin. Every NVIDIA capability referenced here was read this week from NVIDIA's own published materials (blogs, newsroom, developer posts, product pages), each dated in §10. Every hussh claim is graded **IMPLEMENTED**, **IN DEVELOPMENT**, or **ROADMAP** — the same honesty discipline as our published technical specification: a design is judged by the honesty of its status section.

The Fourth Utility

An American household already runs on metered infrastructure it never thinks about: clean water, good electricity, fast internet — and on top of them, the subscriptions that carry content, gaming, sports, and wellbeing into the home. Nobody provisions a power plant to run a dishwasher. They flip a switch, and the meter counts.

Supercomputing is the next line on that bill. The always-on personal agent is the first workload ordinary families will own that genuinely needs it: continuous private inference on their mail, money, health, and home — plus burst capacity for the heavy moments. The correct consumer experience is the utility experience: **always available, metered per unit, billed on a statement they already pay, from a provider they already trust.**

> **The PCHP handshake is the meter. Every data access and every compute job presents a grant, emits a signed receipt, and lands in a hash-chained transparency log — the consent event, the metering point, and the billing point are one record.**

That collapse is the technical heart of this plan. Electricity needed a separate meter bolted to the house. Consent-first compute does not: the same cryptographic receipt that proves an access was authorized *is* the usage record that prices it. **IMPLEMENTED** — ed25519-signed receipts and the append-only, hash-chained log are built and tested today; §09 grades the rest of the pipeline honestly.

WHY THIS IS AN NVIDIA STORY

Utilities are generation plus distribution plus metering. NVIDIA has built generation at every scale a household will ever touch — a 1-petaFLOP desk box, a 20-petaFLOP station, rack systems setting world training records — and, in GeForce NOW, already operates a consumer rail that bills millions of households monthly for GPU time. hussh brings the missing two layers for the personal era: the consent-native meter, and the resident workload that gives a family a reason to turn the generator on.

The founder's north star, stated publicly in the open letter: a distributed edge-supercomputing grid owned not by a hyperscaler but by a network of garage and warehouse owners — compute for the world, owned by the many. §06 gives that grid its engineering rules.

02

The Subscription Rail

The commitment is already public, in the footer of every hussh page: 🤫 **Private Agent One is free for every American citizen.** The question this section answers is how the free agent's burst supercomputing gets paid for — without a new bill, a new card on file, or a new company to trust.

The flow, end to end

STEP	MECHANISM	STATUS
VERIFY	The person proves one existing subscription with a provider they already trust — Apple, Google, Microsoft, or NVIDIA — via platform entitlement attestation (StoreKit App Store Server API, Google Play subscription state, Microsoft Store entitlements, GeForce NOW membership). No password sharing; a signed entitlement token, verified server-side, anchored to the person's HusshID.	IN DEVELOPMENT
ACTIVATE	Agent One provisions free. On-device by default; the subscription is eligibility and billing anchor, never a data source. Defaults closed; the consent ceremony grants item by item.	IN DEVELOPMENT
METER	When a workload provably exceeds the device, 🤫 One Engine bursts it — provision, run, tear down, keys never persisted. Each burst job's PCHP receipt carries the unit count (tokens, joules, seconds). The receipt is the usage record.	RECEIPTS IMPLEMENTED · BURST ROADMAP
BILL	Usage settles as a per-unit line on the statement the family already pays — the partner platform as merchant of record, hussh as the metered service. Modeled on	ROADMAP · PARTNER-GATED

how in-app purchases and cloud-gaming minutes already clear on those rails.

THE GEFORCE NOW INSIGHT

NVIDIA does not need convincing that households will pay a monthly subscription for remote GPU time — **GeForce NOW is that business, operating today,** expanding city by city with RTX 5080-class servers. The proposal is an extension of a rail NVIDIA already runs, not a new category: the same household GPU relationship, extended from rendering frames to running the family's agents. One membership, two workloads, one bill.

Guardrails, stated up front

— Carrier and platform billing is regulated territory; the cramming-rule guardrail from our distribution canon applies — the platform partner is always merchant of record, line items are explicit, and consent to billing is itself a logged PCHP grant.

— Draft unit economics stay placeholders until real attach data exists ($6.90/mo box-class service lines and $0.69 reservations are published; per-unit burst rates are not). We will not print a rate card we have not measured.

— "Free for every American citizen" is a product commitment funded by the paid tiers and the metered network — §09's gates govern the order of operations.

03

Reading NVIDIA's Roadmap · five primary-source reads, worked backwards

Each read below is from NVIDIA's own materials published in July 2026 (full citations in §10), followed by the working-backwards implication for hussh.

READ 1 · INTELLIGENCE PER DOLLAR IS THE METRIC OF THE AGENTIC ERA

NVIDIA's Vera Rubin platform positioning (Jul 17) names continuous **post-training** — not pretraining, not one-shot inference — as the central workload of agentic AI: models given goals must keep adapting as tools and environments shift, so post-training loops back from production and never stops. The stated metric is **intelligence per dollar**, nested on cost per token; NeMo Gym and NeMo RL turn the loop into repeatable infrastructure; Rubin trains the largest models with a quarter of the GPUs of the Blackwell generation, and Vera CPUs deliver ~30% more RL-sandbox throughput than x86 alternatives.

Worked backwards: the household agent is the smallest post-training loop in the world — and the most private. A family's Puppy, running consent-gated RL on their own corrections, on their own silicon, is **personal post-training: intelligence per dollar at N = 1**, with the PCHP transparency log as the audit substrate for every rollout. Our PWS-1 efficiency score (tokens per joule × tokens per TCO-dollar-hour) is the same metric family, measured at the meter. Nemotron 3 Ultra's open weights — post-trainable on proprietary data, deployable locally — are exactly the class of model this loop wants. **ROADMAP** — gated on the on-device inference budget in §09.

READ 2 · THE PERSONAL LINE EXISTS, AND IT SHIPS A SECURE AGENT RUNTIME

DGX Spark (GB10, 1 petaFLOP, 128GB unified, ≤200B-parameter models) is marketed by NVIDIA as **a complete platform for local autonomous agents, built for always-on agent workloads**; DGX Station (GB300 Ultra, ~20 PF, 748GB coherent) extends it; and the DGX OS now ships streamlined **NemoClaw** installation — NVIDIA's security-and-privacy layer for always-on assistants across RTX PCs, Station, and Spark — alongside the OpenShell agent toolkit.

Worked backwards: our Ultra lineup is already built on this line — Ultra S is Spark, Ultra Max is Station, Ultra Custom is 1–4× RTX PRO 6000 Blackwell. The gap NVIDIA's runtime does not fill is ownership semantics: whose agent, whose grant, whose receipt, whose revocation. That is the layer we ship. §04 Seams 1–2.

READ 3 · THE OPEN SECURE AI ALLIANCE NAMES OUR EXACT LAYER AS THE GAP

On Jul 27 NVIDIA convened the Open Secure AI Alliance — with Red Hat, Microsoft, IBM, the Linux Foundation, OpenClaw, HPE and ~40 others — around one thesis: real AI safety depends on **the full agent stack — identity, permissions, harnesses, guardrails, logs and evaluation** — not just model weights, and those controls should be open so any defender can inspect and adapt them. NVIDIA contributed the open NOOA agent-harness framework; HPE contributes SPIFFE/SPIRE zero-trust workload identity.

Worked backwards: identity, permissions, and logs is a literal description of PCHP — grants, signed receipts, the hash-chained transparency log — implemented, tested, and deliberately kept open (monetize the network, never the spec; the posture of our open-source MCP release). **The concrete move: hussh applies to join the Alliance and contributes PCHP as the open consent-and-receipt layer of the agent stack,** with SPIFFE/SPIRE interop for agent identity and NOOA integration for harness-level enforcement. Our OpenClaw MCP server already sits in the registry; OpenClaw is an

inaugural Alliance member. **PCHP CORE IMPLEMENTED** · **ALLIANCE APPLICATION — THIS QUARTER**

READ 4 · RACK-SCALE ECONOMICS: DELIVERED FLOPS, AND SOFTWARE THAT COMPOUNDS

NVIDIA's GB300 NVL72 world record (Jul 21): 1,648 TFLOPs/GPU delivered on DeepSeek-V3 671B MoE pre-training — ~3× the prior GB200 NVL72 result at the same GPU count — on fifth-generation NVLink (1.8 TB/s per GPU, 130 TB/s non-blocking all-to-all), ConnectX-8 at 800 Gb/s per GPU, BlueField DPU isolation, with 97–98.5% per-GPU efficiency held from 256 to 1,024 GPUs. On unchanged hardware, software alone lifted throughput 1.5× in six months; success is measured in delivered FLOPs, not peak.

Worked backwards: two consequences. First, the burst tier a household or campus rents into keeps getting cheaper per unit of work on the same silicon — which is why our **matched-book iron rule** exists: capacity is bought back-to-back against committed demand only, never held naked (B200 rentals repriced +24.4% in March 2026 alone; an unhedged compute broker is a prop desk). Second, "delivered, not peak" is precisely the PWS-1 philosophy — measured on live production jobs, never vendor-claimed. An honest delivered-efficiency scoreboard is a scoreboard this platform wins.

READ 5 · PHYSICAL AI IS BEING MADE AGENT-READY, DOWN TO THE LIBRARIES

In a single July fortnight NVIDIA shipped: new Jetson Thor computers to advance mainstream robotics and edge AI (Jul 15); Omniverse libraries in the Agent Toolkit so agents build simulation-ready worlds (Jul 20); and PhysicsNeMo re-architected as agent-callable libraries plus CUDA-X solvers (cuISS, cuDSS, cuEST) so autonomous AI engineers reason with physics and run simulation as a tool (Jul 26) — with NemoClaw blueprints already in partner workflows (Synopsys).

Worked backwards: the household is a physical-AI site. Three concrete integrations, in honest order: **(a)** Jetson Thor-class devices join the home fleet as consent-gated *actuators* — a robot that touches the household's world acts only through scoped PCHP grants, its every action receipted like any data access. **(b)** A PhysicsNeMo/Omniverse digital twin of the home and garage — thermal, power, solar — is the engineering tool for siting Puppies where the letter says the grid grows: where power is cheap and the sun is abundant. **(c)** Small devices remain consent surfaces, not compute — a watch approves

and revokes; we do not ship spec lies. **ALL THREE ROADMAP** — sequenced after the critical path in §09.

> **Five reads, one pattern: NVIDIA is building the agentic era's generator, runtime, physics, and open safety stack — and publishing the metrics to judge it by. Everything hussh builds either meters that stack, gives it an owner, or gives a family a reason to buy it.**

04

The Integration Map · the full stack, then six seams

LAYER	NVIDIA SHIPS (THEIR MATERIALS)	HUSSH SHIPS (GRADED)
EXPERIENCE	—	Agent One: fetch · organize · guard; works with Siri and other AIs; free for every American citizen. **IN DEV**
CONSENT PLANE	OpenShell guardrails; NOOA harness research; Alliance identity work (SPIFFE/SPIRE)	PCHP: grants, ed25519 receipts, hash-chained transparency log, tombstone revocation; control plane carries capabilities and proofs, never content. **IMPLEMENTED · TESTED**
AGENT RUNTIME	NemoClaw (secure always-on assistants); OpenShell / Agent Toolkit; NIM microservices; Dynamo inference orchestration	🤫 One Engine: placement scheduler (on-device ↔ burst), provision-run-teardown, keys never persisted. **SCHEDULER IN DEV** · **BURST BOUNDARY DESIGNED, NOT WIRED**
LEARNING LOOP	NeMo RL · NeMo Gym; Nemotron 3 Ultra open weights (post-trainable locally)	Personal post-training on owned data, receipts on every rollout. **ROADMAP**
SILICON — OWNED	DGX Spark (GB10 · 1 PF · 128GB) · DGX Station (GB300 · ~20 PF ·	Ultra S / Ultra Max / Ultra Custom lineup, $0.69 reservations live; White-Glove Human-First install.

	748GB) · RTX PRO 6000 Blackwell (96GB, 1–4×) · Jetson Thor	**LINEUP PUBLISHED** · **LINUX AGENT RUNTIME = PORT**
SILICON — BURST	GB200/GB300 NVL72 → VR200 NVL144; NVLink 5; ConnectX-8; BlueField isolation; confidential computing on Hopper/Blackwell	Governed-surge tenant, customer-managed keys; receipts bound to attestation. **DESIGNED, NOT WIRED**
BILLING RAIL	GeForce NOW consumer membership rail (operating); marketplace & partner network (channel)	Subscription attestation → metered receipts → partner-MoR settlement (§02). **IN DEV** · **PARTNER-GATED**

One deliberate asymmetry, stated in the open: hussh is silicon-opinionated at the owned tiers and a neutral broker at the surge tier — our published spec names multiple cloud substrates, and our Watt Score is firewalled from resale by the church/state rule. That neutrality is not a hedge against NVIDIA; it is why the scoreboard means something when this platform tops it — measured, on delivered work.

Six seams, named and owned

SEAM 1 · One Engine ↔ DGX OS / NemoClaw — the Ultra runtime port

THEIRS DGX OS with NemoClaw for secure always-on assistants; OpenShell agent toolkit; NIM. OURS One Engine + Agent One, today Apple-native. JOINT The Linux agent-runtime port our spec already lists as roadmap, targeted at DGX OS first — Agent One resident on Spark/Station with NemoClaw as the security substrate and PCHP as the ownership layer above it. The single highest-leverage engineering workstream in this plan.

SEAM 2 · PCHP ↔ OpenShell / NOOA / confidential computing — consent above guardrails

THEIRS Guardrails, harness research, TEE attestation on Hopper/Blackwell. OURS Grants → receipts → log, implemented. JOINT A PCHP reference implementation for the NVIDIA agent stack: harness-level grant enforcement via NOOA hooks; burst receipts cryptographically bound to confidential-computing attestation, so "this job ran inside a TEE, on these fields, under this grant" is one verifiable record.

SEAM 3 · HusshID × spaceID ↔ SPIFFE/SPIRE — agent identity interop

THEIRS Alliance zero-trust workload identity (SPIFFE/SPIRE, via HPE). OURS One HusshID per human, one spaceID per node, every job a handshake. JOINT Map spaceID node identity onto SPIFFE verifiable identities so hussh nodes are first-class citizens of the open agent-security stack — the grid's machines cryptographically attestable to anyone's tooling, not just ours.

SEAM 4 · Puppy scheduler ↔ Dynamo — burst orchestration

THEIRS Dynamo inference orchestration; NVL72/144 rack systems; 800 Gb/s scale-out. OURS Placement as a scheduling decision — owned core, governed surge; tail-tolerant executor implemented and tested. JOINT One Engine's burst boundary wired to Dynamo-orchestrated capacity: decompose, route, attest, settle, tear down — keys never persisted, every hop a logged consent event with the unit count in the receipt (§02's meter).

SEAM 5 · Personal post-training ↔ NeMo RL / Gym / Nemotron — intelligence per dollar, N = 1

THEIRS NeMo RL + Gym as repeatable post-training infrastructure; Nemotron 3 Ultra open weights, locally post-trainable. OURS The owner's corrections, preferences, and household environment — the most private RL signal on earth. JOINT A reference recipe for on-box personal post-training: NeMo RL on Spark/Station, rollouts receipted, weights never leaving hardware the family owns.

SEAM 6 · Home physical AI ↔ Jetson Thor / Isaac / PhysicsNeMo / Omniverse — actuators under the consent fabric

THEIRS Thor-class robotics compute; Omniverse simulation-ready worlds; agent-callable physics. OURS The household consent fabric and transparency log. JOINT Home robots as scoped-grant actuators of Agent One; a PhysicsNeMo/Omniverse twin of the garage for thermal, power, and solar siting of the grid's nodes. Sequenced last, honestly (§09).

05

The Enterprise Substrate · RHEL, the Alliance, and confidential computing

Households are the mission; enterprises fund the road. For the practice, the firm, and the agency, the substrate question is answered before it is

asked: enterprise AI in America runs overwhelmingly on Red Hat Enterprise Linux — and Red Hat now sits inside NVIDIA's Open Secure AI Alliance, contributing signed-patch supply-chain security (Lightwell, with IBM).

The RHEL commitment

— **Runtime target.** The Linux port of the 🤫 One agent runtime (Seam 1) is built and validated against RHEL for enterprise deployments — one codepath for DGX OS at home and RHEL in the rack, so an RIA practice's Ultra and a bank's on-prem cluster run the same consent fabric. **ROADMAP · DEMAND-GATED**

— **Certification path.** NVIDIA AI Enterprise's supported OS matrix is the target envelope; we pursue validation there rather than inventing our own. Stated as a path being pursued — **no certification is claimed as held today.** **ROADMAP**

— **Supply chain.** hussh ships its agent components with signed, attestable provenance, aligning with the Alliance's Lightwell/Safetensors direction — the same receipts discipline we apply to data, applied to our own binaries. **IN DEVELOPMENT**

Confidential computing, made legible

NVIDIA's Hopper- and Blackwell-generation confidential computing keeps models and data encrypted in use. What the enterprise buyer additionally needs is *legibility*: proof, per job, that the protection held. Seam 2's receipt-to-attestation binding gives the compliance officer a single verifiable record — grant, fields, TEE evidence, hash-chain position — replacing a stack of attestation PDFs with a query the owner can actually run. That is the enterprise version of the household promise: a transparency log their compliance officer can actually read.

WHO THIS UNLOCKS

The published One Puppy Practice motion (RIA practices, insurance FMO/IMOs, Apple-ecosystem MSPs) extends unchanged onto NVIDIA silicon for regulated workloads: client-data agents on Ultra hardware, RHEL-validated runtime for the firms whose IT

departments require it, receipts for every access. Same consent fabric, heavier iron, stricter auditors — satisfied by construction.

06

The Edge Grid · garages on Starlink, engineered honestly

The open letter states the north star plainly: Puppy One computers on Starlink, Agent One managing the fleet, sited where power is cheap and the sun is abundant — compute for the world, owned by the many. This section is that sentence's engineering rules.

Grid rules (recorded before scale, so scale cannot erase them)

— **Matched-book iron rule.** Fleet and rack capacity is bought back-to-back against committed demand only — no naked compute inventory. B200 rental pricing moved +24.4% in March 2026 alone; a broker holding unhedged capacity is a prop desk wearing a utility's clothes.

— **Church/state.** hussh is simultaneously ratings agency (Watt Score), reseller, and broker. Certification is firewalled from the resale P&L: flat published testing fees, open methodology, no pay-for-rank ever. The Moody's problem, solved by constitution rather than promise.

— **Delivered, not peak.** Every brokered job is a production benchmark someone else paid for, emitting measured tokens/s, tokens/J, tokens/$, and loop duration. The catalog's 0/100-scored honest starting position fills from live routing, not lab claims — the same philosophy NVIDIA's own record posts measure by.

— **Consent at every hop.** A federated job is a PCHP handshake between spaceIDs; the household that sells idle cycles gets the same receipts discipline as the household that buys them.

— **The power bill closes the loop.** R4–R6 owners pair hardware depreciation with demand-response and consented grid participation — the garage earns

on the electricity rail it already pays, which is what makes "sited where power is cheap and the sun is abundant" an economic sentence, not a poetic one. A PhysicsNeMo-class thermal/solar twin of the site (Seam 6) is the siting tool.

The sales motion at the door

Door to door, as written: **Kirkland first, Beverly Hills next, a check-in wedge in Las Vegas** — then research, finance, and public sector co-sell. When a warehouse owner puts accelerated compute on the grid and earns from it, the local edge and telecom community wins, the customer gets supercomputing at the lowest cost per watt, and the neighborhood gets better. Cold-start liquidity runs in closed loops we control both sides of — the Garage Grid's first federated node, then one campus loop (idle cluster nights as supply, researcher burst as demand), then 🤫 One power users, then the open market. Single-tenant matched book → campus loop → exchange.

07

America First 🇺🇸 · then the world, deliberately

The commitment is unconditional and already live on every hussh page: 🤫 **Private Agent One is free for every American citizen.** We do not sell your data, your attention, or your contacts. Sovereign supercomputing for the many begins at home.

This is also where the partnership's gravity is. NVIDIA's own July drumbeat — building in America, for America — and the personal-supercomputer line's framing as the personal computers of the AI era describe the same country this plan serves: one where the family, the practice, the campus, and the agency own their intelligence instead of renting custody of it. Personal sovereignty is national sovereignty at N = 1; a nation of citizens who own their agents is the sovereign-AI thesis, finished.

Sequenced honestly

FRONT	POSTURE TODAY

EVERY CITIZEN	Agent One free; subscription rail (§02) carries the burst economics; no citizen ever pays for the agent itself.
PRACTICES & FIRMS	Published Puppy family and Practice bundles, extended to Ultra/NVIDIA configurations per §05.
CAMPUSES	The campus loop is the grid's second liquidity stage — idle cluster nights as supply, researcher burst as demand; academic partnerships already in motion feed it.
FEDERAL	Public solutions pages live (federal, defense & national security). FedRAMP High authorization is **in pursuit and never claimed as held** — the honest status is itself the posture security buyers trust. DGX-class systems already inside federal institutions make the substrate conversation short.
THEN THE WORLD	Built in America, for the world — with the UAE and India tracks (where early traction is strongest) sequenced after the American foundation, not instead of it.

> **A supercomputer in every garage is not a metaphor for scale. It is a civic architecture: capability that is local, loyal, legible, and owned — by Americans first, and then by everyone.**

08

Buy · Build · Sell Together · the concrete program

#	WORKSTREAM	FIRST DELIVERABLE	WINDOW
W1	Ultra runtime port (Seam 1) — Agent One on DGX OS/NemoClaw	Resident Agent One on Spark; consent ceremony end-to-end on DGX OS	2 quarters from pod start
W2	PCHP × OpenShell/NOOA reference (Seam 2)	Open-source consent layer PR + receipt↔attestation binding demo	1–2 quarters
W3	Alliance membership + SPIFFE interop (Seam 3)	hussh application filed; spaceID→SPIFFE mapping spec	This quarter

W4	Burst boundary on Dynamo (Seam 4)	First governed-surge job: provision → run → receipt → teardown, keys never persisted	2–3 quarters
W5	Personal post-training recipe (Seam 5)	NeMo RL on Station, receipted rollouts, published recipe	3–4 quarters
W6	Subscription rail pilot (§02)	One platform partner, one metered burst SKU, live statements	Partner-gated

Sell together

Ultra lineup listed alongside NVIDIA's personal-AI-supercomputer channel; Puppy 100 certified-host program through the NVIDIA Partner Network (our cold-start scoping: 4–12 months unassisted — **sponsorship collapses it to a signature**); Inception for the startup track; a GTC demo as the public proof: a family's Agent One, fully local on Spark, every hop receipted, bursting to Rubin. Channel conflict dissolves by design — partners keep hardware margin; the runtime is ours.

Buy together — symmetric capital

— hussh buys NVIDIA silicon as durable capital assets — owned, not rented — in committed matched-book volumes across Spark → Station → RTX PRO → NVL-class, growing with the grid.

— With his blessing: NVentures in the **$25M seed — first round of a $250M vision** — capital to talent and infrastructure; seven of us today, hiring.

— And the personal ask, exactly as the letter says it: one conversation, one pilot, one garage full of Puppy One on NVIDIA accelerators, owned by the person who hosts it. Then the world.

09

Honest Status & Gates

COMPONENT	STATUS
PCHP consent core — ed25519 receipts, hash-chained log, grants/tombstones, field-scope enforcement	**IMPLEMENTED · TESTED**

Fetch pipeline + tail-tolerant executor (deadline, breakers, graceful partials)	**IMPLEMENTED · TESTED**
60-second live first-fetch against real sources — the defining demo	**IN DEVELOPMENT · CRITICAL PATH**
Subscription attestation (§02 VERIFY/ACTIVATE)	**IN DEVELOPMENT**
Governed-cloud burst boundary; Dynamo wiring (W4)	**DESIGNED · NOT WIRED**
Linux/DGX OS agent runtime port (W1); RHEL validation (§05)	**ROADMAP · DEMAND-GATED**
Watt Score catalog — systems scored at any evidence grade	**0 / 100 · FILLS FROM LIVE ROUTING**
Channel agreements executed; Alliance membership; any NVIDIA agreement	**ZERO TODAY — THIS DOCUMENT IS THE PROPOSAL**
FedRAMP High	**IN PURSUIT · NEVER CLAIMED AS HELD**

Gates, unchanged from our operating canon: G0 first-dollar unblock · G1 matched-book until brokerage GMV proven · G2 ≥40% runtime attach before channel scale · G3 external PCHP deployment behind the engagement + counsel gate · G4 market-data products only after liquidity. The critical path is the live first-fetch; everything else exists to make that minute trustworthy.

10

Sources

Read in full for this plan (all NVIDIA materials published July 2026; retrieved Aug 3, 2026):

— Open Secure AI Alliance — blogs.nvidia.com/blog/open-secure-ai-alliance (Jul 27)

— Vera Rubin: Intelligence per Dollar for Post-Training — blogs.nvidia.com (Jul 17)

— Agent Toolkit + PhysicsNeMo + CUDA-X — nvidianews.nvidia.com (Jul 26); Omniverse libraries (Jul 20) and Jetson Thor (Jul 15) via same channels

— World Record: MoE Pre-Training on GB300 NVL72 — developer.nvidia.com/blog (Jul 21)

— DGX Spark product page (NemoClaw, always-on local agents) — nvidia.com, retrieved Aug 3

hussh canon: the open letter (hushh.ai/one/nvidia); One Puppy Partner Master Pack + Personal Supercomputing Infrastructure technical specification (rev 2026-06-12); compute flywheel & Puppy 100 pages (wiki, Jul 2026).

Provided for study and queued, not yet incorporated: NVIDIA marketplace catalog pages; Naval Postgraduate School DGX coverage. They will be folded into the next revision rather than cited unread.
