BEYOND FRONTIER EVALUATION: AUDITABLE ENERGY, AFFORDABILITY, AND DATA-GOVERNANCE METRICS FOR AI DEPLOYED IN THE GLOBAL SOUTH

Anonymous authors
Paper under double-blind review

ABSTRACT

AI evaluation and governance are indexed to frontier models, yet most people meet AI through distilled, fine-tuned, or platform-embedded systems, often in low- and middle-income countries where rules written for frontier labs propagate weakly. We argue that evaluation should target diffusion capacity in the technology-policy sense: whether a deployed system is affordable, runs on local infrastructure, and can be governed where it lands. We operationalize it with three instruments an auditor can compute or verify within one jurisdiction from public evidence. Joules per Task (JpT) labels inference energy with its measurement boundary and task harness, since rankings can reverse when either changes; in a provenance-preserving collation, published per-prompt figures for text generation (measured and estimated) span nearly three orders of magnitude, and boundary choice alone moves one provider's median-prompt estimate by 2.4×. A Price Accessibility Index (PAI) relates price to income: on a 20-economy panel, a USD 20 monthly subscription exceeds the UN 2%-of-monthly-income affordability line for broadband in 19 economies, burdens range from 6.5× to about 82× the US level, and a purchasing-power discount alone leaves half the panel above the line. A data-value-retention (DVR) vector grades disclosure of onshore storage, consent, local revenue sharing, and portability. Verdict rules grade each axis separately, withhold positive verdicts below a host-country governance threshold, and let oversight and rights failures cap a grade rather than be averaged away; by construction they are monotone in adverse evidence and admit no aggregate that can offset a veto, and each identified gaming route has a closing rule. On three focal cases (an open-weight model release, a data-centre build-out, and a mobile-money platform), the rules return capped and ungraded verdicts alongside positive ones; a re-verified dataset of 92 documented engagements by 39 Chinese AI firm groups in 24 countries grounds the cases descriptively. We keep in-sample consistency checks separate from a forward test specified in advance, and release the dataset, code, and JpT task suite.

1 INTRODUCTION

Technical AI evaluation and governance take the frontier laboratory as their privileged object: how to evaluate, align, and constrain the most capable models trained in a handful of compute-rich jurisdictions (Reuel et al., 2024; Anderljung et al., 2023; Shevlane et al., 2023). For most of humanity that object is the wrong one. A consumer in Lagos, a small enterprise in Jakarta, or a clinic in Nairobi meets AI as a distilled, fine-tuned, localized, or platform-embedded system—an open-weight model on leased racks, a national language model rebuilt on an imported base, a credit-scoring pipeline inside a mobile-money app—and the question that matters there is whether the AI that arrives is affordable, runs on the infrastructure that exists, and can be held to account by the institutions that exist.

The mismatch is structural. Frontier capability is converging while deployment capacity diverges: on the AI Index's arena convention (Chatbot Arena scores) the gap between the best open- and closed-weight models was about 3.4% in March 2026, a fluctuation rather than a chasm (Stanford

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

HAI, 2026), while roughly 32 countries host dedicated AI data centres, US and Chinese operators run more than 90% of dedicated AI compute (Hawkins et al., 2025; Lehdonvirta et al., 2024), and emerging economies outside China hold half the world's internet users but under 10% of data-centre capacity (International Energy Agency, 2025). We call the gap between what can be built and what can be delivered, afforded, and governed the *diffusion deficit*. It widens under geo-economic fragmentation, in which rules written in one jurisdiction propagate weakly across the "islands" of the resulting order (Farrell & Newman, 2019; Witt, 2019; Tang, 2023): artifacts move on product cycles, governance on decade timescales, so the useful evaluation is one that can be performed *within* the jurisdiction where a system lands.

This paper proposes an evaluation standard for that setting. We take the object of evaluation to be *diffusion capacity* in the technology-policy sense (Ding, 2024; Foster & Heeks, 2013): whether a deployed system is affordable (accessibility), runs and scales on local infrastructure (adaptability), and can be governed by the receiving society (institutional compatibility). We operationalize each axis with an instrument that an auditor can compute or verify from public evidence within one jurisdiction, and we specify how verdicts are issued without ever producing a composite score, under demanding constraints: every number carries its provenance, every grade traces to a stated cut-point, no aggregate can launder an adverse finding, and the standard must be able to return adverse verdicts on the very deployments its proponents might favour.

Contributions. (i) We reframe the target of technical AI evaluation from frontier capability to diffusion capacity and derive three evaluation axes from a definition rather than a list, keeping the standard separate from any hypothesis about supplier behaviour (Section 3). (ii) We specify three instruments: JpT, a boundary- and harness-labelled energy-per-task tuple, with a provenance-preserving collation of published values and a proposition stating exactly when rankings are invariant; PAI, a price-to-income index with a metered-consumption extension and a 20-economy panel recomputed from World Bank data; and DVR, a four-component disclosure vector with an explicit order (Section 4). (iii) We state the verdict rules—per-axis grades with published cut-points, a jurisdictional governance gate, and per-axis veto caps—prove monotonicity and non-compensation, and enumerate the gaming routes with the rule that closes each (Section 5). (iv) We run the rules on three focal deployments, obtaining positive, neutral, ungraded, capped, and withheld components (Section 6), and separate in-sample consistency checks from a forward test specified in advance (Section 7). Dataset, code, panel, and the JpT task suite are released as supplementary material.

2 Related work

Technical AI governance. Reuel et al. (2024) organize technical AI governance into six capacities—assessment, access, verification, security, operationalization, and ecosystem monitoring. We adopt the capacities and change the target: our instruments are assessment and verification tools, the governance gate an operationalization condition, and periodic re-audits ecosystem monitoring, all pointed at the deployment tail rather than at frontier training runs. Work on audit access (Casper et al., 2024) and extreme-risk evaluation (Shevlane et al., 2023) shares the frontier indexing we depart from; the end-to-end audit framework of Raji et al. (2020), with its insistence on an evidence trail, is the procedural model our disclosure-based grading follows.

Measuring inference energy. Per-query energy figures are reported at incompatible boundaries and on incompatible tasks: single-accelerator measurements (Luccioni et al., 2024), per-token leaderboards (Chung et al., 2025), latency-based estimates for closed models (Jegham et al., 2025), a corporate self-report that computes the same prompt at two boundaries (Google, 2025a), and deployment-scale accounting that argues several widely cited figures are overstated (Oviedo et al., 2026). MLPerf Power measures at the wall and excludes facility overhead (Tschand et al., 2025); the Software Carbon Intensity (SCI) specification counts per functional unit and includes embodied carbon (Green Software Foundation, 2024); neither fixes a task distribution for deployment auditing. JpT is a reporting discipline on top of these, not a competitor.

Affordability and transparency indices. The Broadband Commission's "1 for 2" target—entry-level broadband at under 2% of monthly GNI per capita—is the accepted affordability line for connectivity (Broadband Commission for Sustainable Development, 2018), preceded by a 5% line

(Alliance for Affordable Internet, 2016); no comparable line exists for AI services, so PAI supplies the ratio and, pending an AI-specific line, adopts the broadband cut-points as stated, contestable defaults. Composite indices of AI transparency and responsibility (Bommasani et al., 2023; Global Center on AI Governance, 2024) confront the same choice between per-indicator reporting and aggregation that DVR faces; we side with per-indicator reporting for the reason the ESG-rating literature gives: aggregation manufactures disagreement and invites reactivity (Berg et al., 2022; Espeland & Sauder, 2007; Strathern, 1997).

Diffusion, infrastructure, and data in the Global South. Ding (2024) makes diffusion, not invention, the mechanism of technological power; Lehdonvirta et al. (2024) and Hawkins et al. (2025) map the material geography of compute (see also Crawford, 2021); development research makes diffusion to low-income users the dependent variable (Foster & Heeks, 2013) and warns that inclusion on platform-dictated terms can deepen inequality (Heeks, 2022). The data-colonialism critique (Couldry & Mejias, 2019) and the institutional-void and bypass literatures (Khanna & Palepu, 2010; Prado & Trebilcock, 2019; Verkuil, 2007) supply the G axis; digital public infrastructure (Eaves et al., 2024), a civic layer of identity, payments, and consent-based exchange, is the nearest governance comparator, to which our instruments add an energy and affordability layer.

3 Diffusion capacity as the evaluation target

Definition and axes. We define the diffusion capacity of a deployed AI system in a jurisdiction as the conjunction of three conditions a receiving society must satisfy for the system to reach and remain with its users: **accessibility** (users can afford and operate it: price, commodity hardware, language coverage), **adaptability** (it runs and scales on local infrastructure: energy, grid, compute), and **institutional compatibility** (the receiving society can govern it: data rights and institutions). The three axes are labelled S (inclusive algorithms), E (sustainable computing), and G (equitable data) after the environmental, social, and governance (ESG) letters they borrow; the labels name the constraint each axis makes visible, not the condition itself. The decomposition is a definition with a defense, not a theorem: each condition is necessary (a system that is unaffordable, cannot run, or is banned does not reach or remain), the conditions are pairwise distinct (each can fail while the other two hold), and the set is exhaustive for reach-and-remain in a falsifiable sense—a counter-example would be a factor whose absence prevents sustained use despite affordability, infrastructural fit, and governability. The obvious candidate, reliability or safety, fails that test: systems demonstrably diffuse and persist at scale while unsafe, so safety conditions whether diffusion *benefits* users, not whether it occurs. A second candidate, demand, is a property of the market rather than of the deployment, and the standard takes it as given; we know of no other and invite counter-examples.

Where safety belongs, and what the verdicts are not. Safety, accountability, and misuse therefore enter the standard in two places rather than as a fourth axis: as a jurisdictional *governance gate*—no positive verdict issues where the receiving jurisdiction lacks a comprehensive data-protection statute in force and a supervisory authority with sanction powers in operation, because the same technology's effect depends on the institutions that receive it (Carter & Carter, 2025; Heeks, 2022)—and as *veto caps*: documented surveillance-capable deployment without independent oversight, biometric abuse, systematic exclusion, or exit-foreclosing lock-in caps the affected axis at its floor. We state the scope explicitly. The verdicts are *diffusion-trustworthiness* verdicts—whether a deployment can be trusted to reach, scale, and remain governable where it lands—not model-behaviour safety certifications, and they contain no task-level reliability floor: a clinical triage model that clears every rule here may still be unfit for its domain, and domain evaluation is a separate instrument a regulator should require in addition, not a component we have folded in.

The island boundary and three auditor modes. An "island" is the jurisdiction whose users the deployment serves and whose regulator would act on the verdict. Because many such jurisdictions consume models hosted elsewhere, we distinguish three auditor modes, each labelled in the record: *measure* (open weights run on hardware the auditor controls), *verify* (a provider's disclosed, boundary-labelled value re-run on a subset), and *reference* (a third-party measurement cited as such). For remote closed services JpT is thus a verification-and-attestation protocol rather than a measurement the auditor performs, and the grade reflects that (Section 5).

Standard versus hypothesis. The standard evaluates deployments and does not presuppose why suppliers behave as they do; the conjecture that motivated it (that under fragmented conditions some suppliers internalize accessibility, energy, and data-governance criteria as design parameters rather than compliance overlays) is stated, checked for consistency, and paired with a forward test in Section 7 and Appendix F, and nothing in the instruments or rules depends on it.

4 THREE INSTRUMENTS

4.1 JOULES PER TASK (JPT), BOUND TO SUSTAINABLE COMPUTING (E)

For a system s serving a task distribution $\mathcal{T}$ under a fixed harness H (prompts, decoding parameters, batch condition), let $E_s^b(\tau)$ be the energy in joules attributed to task τ at measurement boundary $b \in \{\text{chip},\ \text{node},\ \text{node+PUE}\}$: accelerator energy only; whole-server energy including idle accelerators, host CPU and memory; and server energy multiplied by the facility's measured power-usage effectiveness (PUE). Then

$$\mathrm{JpT}_b(s; \mathcal{T}, H) \ = \ \mathbb{E}_{\tau \sim \mathcal{T}}\big[E_s^b(\tau)\big], \tag{1}$$

reported as the boundary-labelled tuple $\langle \mathrm{JpT}_{\text{chip}}, \mathrm{JpT}_{\text{node}}, \mathrm{JpT}_{\text{node+PUE}} \rangle$ with the harness identifier. Wide-area network and end-device energy are deliberately outside the tuple; an end-to-end figure may be reported alongside, labelled as such, but is not comparable across deployments. Write $\mathrm{JpT}_b(s) = \kappa_b(s)\,\mathrm{JpT}_{\text{chip}}(s)$ with overhead factor $\kappa_b(s) \geq 1$.

Proposition 1 (Boundary–harness invariance)**.** *Fix $\mathcal{T}$ and H. For systems s_1, s_2 with $\mathrm{JpT}_{chip}(s_1) < \mathrm{JpT}_{chip}(s_2)$, the ranking at boundary b reverses if and only if $\kappa_b(s_1)/\kappa_b(s_2) > \mathrm{JpT}_{chip}(s_2)/\mathrm{JpT}_{chip}(s_1)$. Hence unequal overhead factors are necessary but not sufficient for a boundary-induced reversal; under a fixed $(b, \mathcal{T}, H)$ the ranking is invariant to the reporting convention.*

The proof is a one-line rearrangement (Appendix C). The condition is not academic: one provider's own disclosure puts the same median prompt at 0.10 Wh (360 J) at the accelerator and 0.24 Wh (864 J) at node+PUE, a $2.4\times$ gap from boundary choice alone (Google, 2025a). The proposition turns on the ratio of two systems' overhead factors, not on one system's boundary spread; but where deployments differ in overhead by as much as one provider's own boundary gap, chip-level energies within $2.4\times$ of each other can be reordered.

What the published record looks like. Figure 1 collates published per-task values for text generation without normalizing them, since normalizing would erase the provenance a regulator needs. They span nearly three orders of magnitude: about 151 J for a frugal single-accelerator generation (Luccioni et al., 2024); 360–864 J for a median consumer prompt depending on boundary (Google, 2025a); about 1.1 kJ for a deployment-scale frontier median (Oviedo et al., 2026); 1.5–6.4 kJ for short and long GPT-4o-class prompts; and 110–141 kJ for the heaviest long prompts, of which the two reasoning-mode flagships (121 and 141 kJ) are estimated at $74\times$ and $86\times$ the lightest long-prompt value in the same evaluation (GPT-4.1 nano, 0.454 Wh, about 1.6 kJ) (Jegham et al., 2025). Per-token rates of 0.15–0.4 J on H100-class hardware (Chung et al., 2025) anchor the low end. The peer-reviewed correction of Oviedo et al. (2026)—that several widely cited per-query figures are overstated 4–$20\times$ at deployment scale—sharpens the argument rather than weakening it: applied to the estimate-class rows it would compress the span to roughly one-and-a-half to two-and-a-half orders of magnitude, so the three-order figure describes the spread of the *published record*, which is what a regulator has to read, not true energy.

Harness and grading. JpT values are comparable only under a fixed public harness. We release a reference suite (40 prompts in five families—short QA, summarization, translation into eight lower-resource languages, code, and multi-step reasoning—with greedy decoding and two batch conditions) and a harness that writes the boundary tuple per prompt; the collation above reports published values on their own harnesses and claims no results on the suite. Standards anchors are MLPerf Power (Tschand et al., 2025) and SCI (Green Software Foundation, 2024). For infrastructure deployments the E instrument is the per-site *measured* PUE, water-usage effectiveness (WUE), and grid carbon intensity, for which nominal product specifications do not substitute.

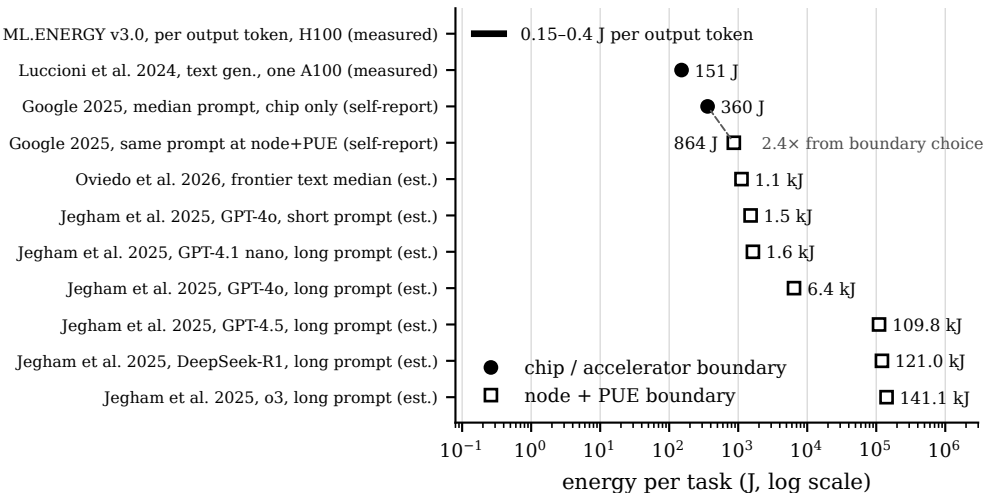

Figure 1: Per-task inference energy for text generation as published, on one log scale, each value carrying its source's own boundary and characterization (measured, self-report, or estimate); the underlying table with sources is in Appendix A. Filled circles are chip/accelerator-boundary values; open squares include node overhead and PUE. Nominally comparable tasks span nearly three orders of magnitude (151 J to 141 kJ), the deployment-scale median of Oviedo et al. (2026) sits between the short-query and reasoning-mode clusters, and the same prompt moves $2.4\times$ under boundary choice alone—the case for boundary-labelled, task-relative reporting rather than any single number.

4.2 Price Accessibility Index (PAI), bound to inclusive algorithms (S)

For a service s priced at p_s US dollars per month in jurisdiction j with annual per-capita income y_j (GNI, Atlas method, current US dollars),

$$\mathrm{PAI}(s, j) \;=\; \frac{12\, p_s}{y_j}, \tag{2}$$

the share of monthly per-capita income the service consumes, compared against the Broadband Commission's 2% line (Broadband Commission for Sustainable Development, 2018). The relative burden $\mathrm{PAI}(s, j)/\mathrm{PAI}(s, \mathrm{US})$ expresses the same quantity against a reference economy. PAI is a transparent ratio because, for an audit instrument, trivial recomputability from public data beats sophistication. Two extensions handle the cases a subscription price does not cover. For *metered* services priced per token, p_s is the monthly cost of a disclosed reference consumption basket—we use a light basket (0.2M input and 0.05M output tokens) and a heavy one (2M and 0.5M)—so that a PAI for an API is always a PAI *for a basket*, and the basket is part of the record. For *zero-price* artifacts (open weights, free tiers), PAI is undefined rather than maximal, and S is graded on the remaining accessibility criteria: a documented commodity-hardware floor and documented coverage of the jurisdiction's official languages. A purchasing-power variant replaces p_s by $p_s\, k_j$, with k_j the ratio of the purchasing-power-parity (PPP) conversion factor to the market exchange rate; we report it as a sensitivity rather than as the index, because Atlas-method income is exchange-rate based and calling the index "PPP-adjusted" would misdescribe it, and the pricing exercise below uses it as a *pricing* proposal, which is a different use from indexing affordability.

The panel. Figure 2 applies a US$20 monthly reference price—the common flagship consumer tier—to twenty low- and middle-income economies. Nineteen exceed the 2% line; the median burden is 9.0% of monthly income; and the burden ranges from $6.5\times$ to $82\times$ the US level. A three-stage exercise, of which we can populate two (a purchasing-power discount, then a 2%-of-monthly-income welfare cap, then a cost floor we cannot populate because no vendor discloses marginal inference cost), yields prices in the US$1.68–10.83 band and shows that the discount alone leaves ten of the twenty economies above the line: purchasing-power pricing is necessary but not sufficient. Vendors have since introduced regional consumer tiers priced near US$5 in dozens of low- and middle-income markets (OpenAI, 2025; Google, 2025b), inside this band, but no vendor discloses the pricing

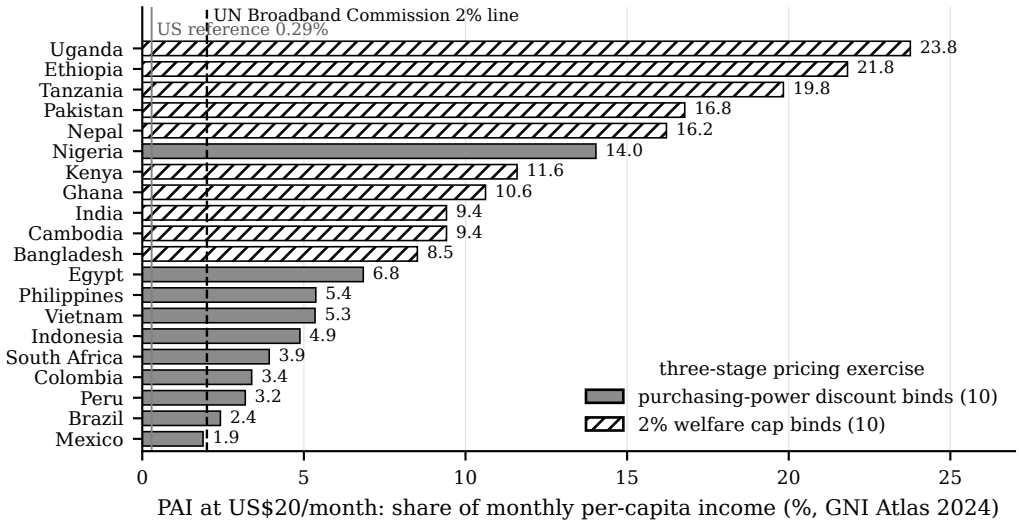

Figure 2: The 20-economy panel at a US$20 monthly reference price, recomputed from 2024 World Bank values (World Bank, 2024). Nineteen of twenty economies exceed the 2% line (Mexico, 1.9%, is the exception); the median is 9.0% and the burden ranges from $6.5\times$ (Mexico) to $82\times$ (Uganda) the US reference of 0.29%. Shading records which stage of the three-stage pricing exercise binds: the purchasing-power discount for ten economies, the 2% welfare cap for the other ten—including, because currency shocks push market rates far from PPP, the discount rather than the cap for Nigeria and Egypt.

algorithm, which is the disclosure gap PAI targets. The metered extension (Appendix B) shows why the basket must be stated: at list prices accessed in September 2026, every provider's smallest tier clears the 2% line in all twenty economies under the light basket, whereas flagship tiers under the heavy basket clear it in none to ten of them.

4.3 Data-value retention (DVR), bound to equitable data (G)

G is reported as a four-component disclosure vector $\mathrm{DVR}(s, j) = \langle$onshore storage, informed consent, local revenue sharing, portability$\rangle$. Each component takes a value in the ordered set $\mathsf{U} \prec \mathsf{D}^- \prec \mathsf{D}^+ \prec \mathsf{V}$: undisclosed; disclosed with the commitment absent; disclosed with the commitment present; disclosed, present, and verified by a regulator filing, audit, or independent test. Vectors are compared componentwise (a partial order); the *floor* of the axis is grade F, which a vector of all-U-or-D^- components also yields, and a veto cap sets the grade to F with an adverse flag, so "capped" is never confused with "not yet disclosed". Disclosure is the auditable proxy for retention, not retention itself; a provider that discloses offshore storage is more transparent than one that discloses nothing and is graded accordingly, but the commitment it discloses is absent, and the grade records both facts.

5 How verdicts are issued

The standard issues verdicts, not scores, by the rules in Table 1: each axis is graded ordinally on its own instrument; the governance gate withholds positive grades below the jurisdictional threshold; and veto caps pin an axis to its floor on documented adverse evidence. Undisclosed quantities leave an axis ungraded rather than presumed, because presumption in either direction is what a disclosure standard exists to remove. Three properties follow from the construction.

Proposition 2 (Monotonicity and no cross-axis compensation)**.** *Under the rules of Table 1: (i) additional adverse evidence never raises any verdict component, and (ii) no verdict component can be raised by evidence on another axis.*

Table 1: Verdict rules v0.1: per-axis grades from stated cut-points, the governance gate, and the veto caps. Grades are P (positive), N (neutral), F (failing), U (ungraded); W marks a positive grade withheld by the gate. No aggregate is computed. Cut-points are parameters of the standard, published so that a verdict can be contested by contesting a cut-point.

Axis / rule	Grade
S, priced tier	PAI $\leq 2\%$: P; $2\% < \text{PAI} \leq 5\%$: N; PAI $> 5\%$: F (the 2% and 5% lines of Broadband Commission for Sustainable Development, 2018; Alliance for Affordable Internet, 2016). Metered services are priced on the disclosed basket.
S, zero price	Hardware floor and official-language coverage both documented: P; one: N; neither: U.
E, model service	Tuple disclosed at node+PUE on a fixed public harness and verified (measure or verify mode): P; any boundary-labelled value, including a third-party reference: N; disclosed value contradicted by verification by more than $2\times$: F; no boundary-labelled value: U.
E, facility	Per-site measured PUE (and WUE, grid intensity where available) disclosed and verified: P; disclosed: N; contradicted: F; undisclosed: U.
G (DVR)	All four components $\succeq \mathsf{D}^+$ with at least two V: P; at least two components disclosed: N; a documented adverse finding (breach, sanction, or a disclosed commitment shown false): F; three or more U: U.
Governance gate	Positive grades issue only in jurisdictions with (a) a comprehensive data-protection statute in force and (b) a supervisory authority with sanction powers in operation; elsewhere P becomes W.
Veto caps	Documented surveillance-capable deployment without independent oversight, biometric abuse, systematic exclusion, or exit-foreclosing lock-in caps the affected axis at F, with the flip condition (the disclosure that lifts the cap) recorded. Caps are per axis.

(i) holds because every rule maps adverse evidence to a weakly lower grade or a withdrawn positive; (ii) because each grade is a function of its own axis's evidence and the gate only lowers (Appendix C). *Remark (no composite).* The rule set defines no function from the axis grades to a scalar, so there is no composite *within the standard* against which a veto-capped axis could be offset. This is a property of the design, not a theorem about all scoring functions: anyone can compute a composite outside the standard, and the standard's answer is that such a composite is not one of its verdicts. We refuse a single number deliberately: cross-firm ordinal ranking is where reactivity is strongest (Espeland & Sauder, 2007), and per-axis verdicts with flip conditions are offers of a compliance path rather than standing indictments.

Gaming the standard. An audit instrument earns trust by stating how it would be gamed and which rule closes each route. Appendix G lists seven: boundary, harness, and basket shopping (closed by the mandatory tuple, the fixed suite, and the disclosed basket), tier gaming and disclosure theatre (closed by the zero-price rule and the $\mathsf{D}^+ \prec \mathsf{V}$ order), composite lobbying (closed by the remark above), and gate gaming—routing a deployment's nominal jurisdiction through a gate-passing host while its users sit elsewhere—closed by defining the island as the jurisdiction of the users served.

6 THREE FOCAL CASES

The cases span the three axes and the three kinds of artifact the standard must handle: an open-weight release with a hosted metered service, an infrastructure build-out, and a platform-embedded system. Two of them are rows of CAIGS-92 (Chinese AI in the Global South), an event-level dataset of 92 documented, landed engagements by 39 Chinese AI firm groups in 24 countries, re-verified row by row and released with the paper (Appendix E); the open-weight release falls outside the dataset's landed-deployment criterion and is included because the standard must handle that artifact type. The dataset grounds the cases descriptively (only 23 of 92 engagements carry a company-disclosed scale figure, the disclosure gap the instruments target) and identifies no causal effect. Every input to a verdict is a matter of public record, reconstructed rather than gathered by fieldwork on the firms. Table 2 gives the result of running the rules; the rule application is in the released code.

Table 2: Verdicts on the focal cases under Table 1. P positive, N neutral, F failing, U ungraded, W withheld by the gate; $F^{\dagger}$ capped by a veto (raw grade U); "–" not applicable. Inputs and sources: Appendix D.

Deployment	Jurisdiction (gate)	S	E	G	Binding rule
Open-weight release (DeepSeek-R1) + provider's metered API (flash tier, light basket)	Nigeria (pass)	P	N	N	S: basket PAI 0.08% (pro tier 0.32%, also P; zero-price artifact N); E: third-party references only; G: two of four components disclosed, storage disclosed as offshore
Same deployment	Cambodia (fail)	W	N	N	gate: no comprehensive statute in force
Data-centre build-out (Huawei Cloud Africa; Konza national cloud)	Kenya (pass)	–	U	$F^{\dagger}$	E: per-site measured PUE undisclosed; G: surveillance/lock-in veto, no disclosed oversight or exit instrument
Mobile-money platform (OPay: biometric KYC, automated fraud and credit scoring)	Nigeria (pass)	P	U	N	S: zero price, USSD channel, official language; E: no disclosure; G: storage and consent disclosed, retention and portability not; oversight present, veto does not fire

An open-weight release. The release of DeepSeek-R1 under an open licence, with distilled variants that run on commodity hardware, converted frontier-class capability into a non-rival input (DeepSeek-AI, 2024; 2025); the provider's own hosted API is metered. Under the light basket at September 2026 list prices the provider's flash tier costs about US$0.12 per month, a PAI of 0.08% in Nigeria (pro tier: 0.32%), so S is positive for the priced service; the zero-price artifact has a documented hardware floor for its distilled variants but disclosed language coverage only for Chinese and English, an N under the zero-price rule, and Appendix D carries both grades. E is neutral: third-party values exist at labelled boundaries (a long reasoning-mode prompt estimated at 33.6 Wh at node+PUE; Jegham et al., 2025) but the provider discloses nothing and no value is on a fixed public harness. G for the hosted service is neutral: the privacy policy discloses storage in the provider's home jurisdiction (D^{-} for a Nigerian user) and a consent regime (D^{+}), and is silent on revenue retention and portability; for weights run locally, G is determined by the deployer, not the model. The same inputs in Cambodia, where a comprehensive data-protection statute exists only in draft (Cambodia MPTC, 2025), yield W on S: the gate withholds the positive verdict exactly where the framework's critics would expect a supply-side standard to wave it through. Single-vendor adoption telemetry places this model in the 11–15% band of generative-AI market share across most of Africa, Nigeria included, and far higher in politically aligned economies (China 89%, Belarus 56%, Cuba 49%, Russia 43%) (Microsoft AI Economy Institute, 2026); both readings, accessibility-driven and alignment-driven, are live, and neither enters a verdict.

A data-centre build-out. Huawei Cloud's African build-out—a local public cloud in Johannesburg (2019) (Reuters, 2024), a local cloud in Lagos (2024) (Huawei Cloud, 2024), and a national cloud at Konza with an on-site substation financed by a concessional loan (Konza Technopolis, 2025)—is built from the vendor's prefabricated modular product line in grid-constrained geographies, where the binding constraint is often absolute electricity availability rather than carbon intensity. The auditable quantity, per-site measured PUE under local thermal and grid conditions, is undisclosed and nominal product figures (≤ 1.15) do not substitute, so E is ungraded rather than presumed. G is capped: the public record of the same vendor's earlier safe-city deployments in the host country, together with a single-vendor national cloud with no disclosed independent-oversight or exit instrument, fires the surveillance-and-lock-in veto. The flip conditions—publish per-site measured PUE; disclose a binding oversight instrument and portability guarantees—require no trust in the vendor.

A platform-embedded system. OPay, a Nigerian mobile-money platform with tens of millions of users, operates biometric identity verification at onboarding and automated fraud and credit-

risk scoring; these deployed automated systems are the audited objects. The 2024 onboarding freeze ordered by the central bank over know-your-customer (KYC) gaps, lifted after remediation (TechCabal, 2024), enters the verdict only as evidence for gate condition (b)—a supervisor that can and does sanction—not as an AI-governance event, which it was not. S is positive: zero price, a USSD (feature-phone dial-code) channel, and service in the official language. E is ungraded: the platform is hosted on a Lagos cloud region (Huawei Cloud, 2025) with no energy disclosure. G is neutral: onshore storage under Nigerian financial-data rules and a consent regime are disclosed; revenue retention and portability are not. The concentration of a re-identifiable biometric and transactional corpus under one operator is recorded as a monitoring item; because an independent data-protection authority and a sectoral supervisor exist (Nigeria, 2023), the veto does not fire.

7 Consistency checks and a forward test

Two kinds of empirical claim attach to the standard, and we keep them apart. The first is an *in-sample consistency check*: the verdict components in Table 2 and the dataset's descriptive pattern are consistent with the motivating conjecture that, on this corridor, deeper integration coincides with more governance-active hosts rather than with weaker ones (Appendix F states the conjecture, its rivals, and the discriminating indicators). Because the indicators are scored on the evidence that motivated the conjecture, this is a consistency check and nothing more: no effect size, no adjudicated rival. The second is a *forward test specified in advance*. The non-Chinese AI-compute build-out now under way in Africa—GPU facilities in South Africa with announced extensions to Nigeria and Kenya (Cassava Technologies, 2025)—will be audited with the same instruments and rules at a fixed date (the first quarter of 2028) by the released code. The pre-specified prediction is that verdict components will track host-jurisdiction governance (gate status, disclosed oversight instruments) rather than supplier origin: the same host should produce the same G outcomes for a US-supplied facility as for a Chinese-supplied one. If adverse G components instead concentrate by supplier regardless of host, then either the instruments are not supplier-neutral or supplier origin carries information the conjecture denies; we pre-commit to separating the two by re-auditing each adverse component against its flip condition rather than by re-reading the outcome (protocol: Appendix F).

8 Discussion and limitations

When cross-border standards propagation is attenuated, the useful instrument is one an independent auditor can compute, verify, or reference within an island from public evidence; JpT, PAI, and DVR are that, and every policy use is phrased as disclose, reference, verify, never as a market-access gate. Six limitations bound the claims. (1) The cut-points in Table 1 are defaults, argued for and published, not calibrated on a pilot; the rules have not been run at scale, so we claim no stable cross-target comparability. (2) The JpT collation reports published values on their own harnesses; no results on the released suite are claimed. (3) GNI per capita is a mean, so in high-inequality economies PAI understates how prohibitive access is below the mean. (4) DVR grades disclosure, not retention; only the V level touches practice. (5) The dataset is a census of documented engagements, skewed toward announced projects and used descriptively only, and the focal cases rest on public-record reconstruction with some single-source inputs, labelled as such. (6) The standard contains no domain-reliability floor, by design.

9 Conclusion

The next decade of AI governance will be decided as much in Lagos and Jakarta as in the frontier laboratories, and evaluation should be indexed accordingly. We have proposed a standard that evaluates AI where it is deployed—boundary-labelled energy per task, income-relative price access, and data-governance disclosure, combined by rules that grade each axis on its own evidence, withhold positive verdicts below a host-country governance threshold, and cap rather than average away oversight and rights failures—and shown on three deployments that it returns positive, neutral, ungraded, capped, and withheld components, each with the disclosure that would change it: a standard recomputable by the regulator who has to act on it, and able to say no.

AI USE STATEMENT

We used generative AI tools (Anthropic's Claude) in this work, and disclose the uses by the categories of the ICLR AI Policy for Authors. *Tasks in the required-disclosure list:* helping to develop and refine the conceptual framework and the verdict rules; refining the forward-test conjecture, its rivals, and its discriminating indicators; assisting in formulating the precise statements of Propositions 1 and 2 and in drafting their proofs; implementing the methods (the panel, metered-basket, figure, verdict-rule, and dataset-release scripts); and drafting the interpretation of the focal-case verdicts and of the panel results. *Tasks in the recommended-disclosure list:* drafting and editing text for readability, suggesting structure, identifying and summarizing related literature, sourcing and fact-checking cited figures against their primary sources, creating the figures through the released scripts, drafting the JpT task-suite prompts, and formatting references. We did not use generative AI to generate synthetic data; translation, transcription, and dataset cleaning are not applicable to this work. All AI-assisted content was reviewed by the authors: every cited figure was checked against its primary source, the proofs were checked line by line, all code was run and the values reported in the paper were compared with its outputs, and the literature identified with AI assistance was read before citation. The authors take full responsibility for the content of this work, including text, claims, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

The paper proposes an evaluation standard and applies it to public-record evidence about named commercial deployments; it collects no personal data and involves no human subjects. The verdicts are diffusion-trustworthiness verdicts, not safety certifications, and are not rankings of countries or vendors: no composite or ordinal ranking exists, every policy use is phrased as disclose, reference, or verify, and each adverse component carries the disclosure that would lift it. The dual-use risk we cannot design away is selective quotation of adverse components in bloc politics; the mitigation is that verdicts are per axis, per jurisdiction, and published with their flip conditions. The released dataset compiles publicly reported corporate engagements with per-row source URLs and contains no personal information. The paper advocates evidence standards, not any vendor's or bloc's deployment.

REPRODUCIBILITY STATEMENT

The supplementary material contains: the 20-economy panel with the World Bank values used and the script that re-fetches them and recomputes every number in Section 4.2 and Figure 2; the metered-basket computation with the list prices used and their access dates; the JpT collation table with per-row sources (Appendix A) and the figure script; the JpT reference task suite (40 prompts, CC0) and the measurement harness; the verdict-rule implementation and the focal-case inputs that produce Table 2; and the CAIGS-92 dataset with codebook and the axis-coding rule behind Figure 3. Appendix C gives the proofs, Appendix D the per-case evidence and sources, and Appendix F the forward-test protocol.

REFERENCES

Alliance for Affordable Internet. The 2015–16 affordability report. Technical report, Alliance for Affordable Internet (A4AI), Washington, DC, 2016.

Markus Anderljung, Joslyn Barnhart, Anton Korinek, Jade Leung, Cullen O'Keefe, Jess Whittlestone, et al. Frontier AI regulation: Managing emerging risks to public safety. arXiv:2307.03718, 2023.

Florian Berg, Julian F. Kölbel, and Roberto Rigobon. Aggregate confusion: The divergence of ESG ratings. *Review of Finance*, 26(6):1315–1344, 2022.

Rishi Bommasani, Kevin Klyman, Shayne Longpre, Sayash Kapoor, Nestor Maslej, Betty Xiong, Daniel Zhang, and Percy Liang. The foundation model transparency index. arXiv:2310.12941, 2023.

Broadband Commission for Sustainable Development. 2025 targets: "Connecting the other half". ITU/UNESCO Broadband Commission, Geneva. Target 2: entry-level broadband affordable at less than 2% of monthly GNI per capita, 2018.

Cambodia MPTC. Draft law on personal data protection. Final draft under consultation; no comprehensive statute in force at the time of writing, 2025.

Erin Baggott Carter and Brett L. Carter. Exporting the tools of dictatorship: The politics of China's technology transfers. *Perspectives on Politics*, 23(3):1089–1108, 2025.

Stephen Casper, Carson Ezell, Charlotte Siegmann, Noam Kolt, Taylor Lynn Curtis, Benjamin Bucknall, et al. Black-box access is insufficient for rigorous AI audits. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*, pp. 2254–2272, 2024.

Cassava Technologies. Cassava and NVIDIA to build Africa's first AI factory. Company announcement, 2025 (South Africa first; Egypt, Kenya, Morocco, and Nigeria announced to follow), 2025.

Jae-Won Chung, Jeff J. Ma, Ruofan Wu, Jiachen Liu, Oh Jun Kweon, Yuxuan Xia, Zhiyu Wu, and Mosharaf Chowdhury. The ML.ENERGY benchmark: Toward automated inference energy measurement and optimization. arXiv:2505.06371; leaderboard v3.0, 2025.

Nick Couldry and Ulises A. Mejias. *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Stanford University Press, 2019.

Kate Crawford. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press, 2021.

DeepSeek. DeepSeek privacy policy. chat.deepseek.com, version in force 2025 (data stored on servers in the People's Republic of China), 2025.

DeepSeek. DeepSeek API pricing. api-docs.deepseek.com, accessed 21 September 2026, 2026.

DeepSeek-AI. DeepSeek-V3 technical report. arXiv:2412.19437, 2024.

DeepSeek-AI. DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning. arXiv:2501.12948, 2025.

Jeffrey Ding. *Technology and the Rise of Great Powers: How Diffusion Shapes Economic Competition*. Princeton University Press, 2024.

David Eaves, Mariana Mazzucato, and Beatriz Vasconcellos. Digital public infrastructure and public value: What is 'public' about DPI? Technical Report Working Paper 2024/05, UCL Institute for Innovation and Public Purpose, 2024.

Wendy Nelson Espeland and Michael Sauder. Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology*, 113(1):1–40, 2007.

Henry Farrell and Abraham L. Newman. Weaponized interdependence: How global economic networks shape state coercion. *International Security*, 44(1):42–79, 2019.

Christopher Foster and Richard Heeks. Conceptualising inclusive innovation: Modifying systems of innovation frameworks to understand diffusion of new technology to low-income consumers. *European Journal of Development Research*, 25(3):333–355, 2013.

Global Center on AI Governance. Global index on responsible AI 2024. Technical report, Global Center on AI Governance, Cape Town, 2024.

Google. Measuring the environmental impact of delivering AI at Google scale. C. Elsworth, K. Huang, D. Patterson, I. Schneider, R. Sedivy, S. Goodman, B. Townsend, P. Ranganathan, J. Dean, A. Vahdat, B. Gomes, and J. Manyika; arXiv:2508.15734 (corporate self-report), 2025a.

Google. Google AI Plus: a lower-priced consumer tier. Company announcements, September–October 2025 (launched in Indonesia and extended to more than 40 countries at about US$4.99 per month), 2025b.

Google. Gemini API pricing. ai.google.dev/gemini-api/docs/pricing, accessed 21 September 2026, 2026.

Green Software Foundation. Software carbon intensity (SCI) specification. ISO/IEC 21031:2024, 2024.

Zoe Jay Hawkins, Vili Lehdonvirta, and Boxi Wu. AI compute sovereignty: Infrastructure control across territories, cloud providers, and accelerators. SSRN working paper 5312977, 2025.

Richard Heeks. Digital inequality beyond the digital divide: Conceptualizing adverse digital incorporation in the Global South. *Information Technology for Development*, 28(4):688–704, 2022.

Huawei Cloud. Launch of the Nigeria local cloud region (Lagos). Company announcement, December 2024, 2024.

Huawei Cloud. OPay accelerates financial inclusion in Nigeria with Huawei Cloud Lagos. Customer case study, August 2025, 2025.

International Energy Agency. Energy and AI. Technical report, IEA, Paris, 2025.

Nidhal Jegham, Marwan Abdelatti, Chan Young Koh, Lassad Elmoubarki, and Abdeltawab Hendawi. How hungry is AI? Benchmarking energy, water, and carbon footprint of LLM inference. arXiv:2505.09598 (preprint), 2025.

Kenya. Data protection act, 2019. Act No. 24 of 2019, establishing the Office of the Data Protection Commissioner, 2019.

Tarun Khanna and Krishna G. Palepu. *Winning in Emerging Markets: A Road Map for Strategy and Execution*. Harvard Business Press, 2010.

Konza Technopolis. Konza national data centre and cloud: Phase-one infrastructure completion. Republic of Kenya, October 2025 (China Eximbank concessional financing; Huawei as contractor), 2025.

Vili Lehdonvirta, Boxi Wu, and Zoe Hawkins. Compute North vs. Compute South: The uneven possibilities of compute-based AI governance around the globe. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES 2024)*, 2024.

Alexandra Sasha Luccioni, Yacine Jernite, and Emma Strubell. Power hungry processing: Watts driving the cost of AI deployment? In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*, pp. 85–99, 2024.

Microsoft AI Economy Institute. Global AI adoption in 2025: A widening digital divide (AI diffusion report). Technical report, Microsoft, 2026. January 2026.

Nigeria. Nigeria data protection act, 2023. Act establishing the Nigeria Data Protection Commission (NDPC), 2023.

OpenAI. ChatGPT Go: a lower-priced consumer tier for price-sensitive markets. Company announcements, August–November 2025 (launched in India, then extended to dozens of low- and middle-income markets at roughly US$4.5–5 per month), 2025.

OpenAI. API pricing. openai.com/api/pricing, accessed 21 September 2026, 2026.

Felipe Oviedo, Fiodar Kazhamiaka, Esha Choukse, Allen Kim, Amy Luers, Melanie Nakagawa, Ricardo Bianchini, and Juan M. Lavista Ferres. Energy use of AI inference, efficiency pathways, and test-time scaling. *Joule*, 10:102430, 2026. arXiv:2509.20241.

Mariana Mota Prado and Michael J. Trebilcock. *Institutional Bypasses: A Strategy to Promote Reforms for Development*. Cambridge University Press, 2019.

Inioluwa Deborah Raji, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20)*, pp. 33–44, 2020.

Anka Reuel, Ben Bucknall, Stephen Casper, Tim Fist, Lisa Soder, Onni Aarne, et al. Open problems in technical AI governance. arXiv:2407.14981, 2024.

Reuters. Huawei Cloud sees fast business growth in South Africa. News report, 24 October 2024 (local cloud region launched in Johannesburg in 2019), 2024.

Toby Shevlane, Sebastian Farquhar, Ben Garfinkel, Mary Phuong, Jess Whittlestone, Jade Leung, et al. Model evaluation for extreme risks. arXiv:2305.15324, 2023.

Stanford HAI. Artificial intelligence index report 2026. Technical report, Stanford Institute for Human-Centered Artificial Intelligence, 2026.

Marilyn Strathern. 'Improving ratings': Audit in the British university system. *European Review*, 5 (3):305–321, 1997.

Yao Tang. *Where Does the Money Come From, Vol. 4: The Archipelagized Economy*. New Star Press, Beijing, 2023.

TechCabal. Nigerian fintechs resume customer signups as CBN lifts freeze. 3 June 2024. Onboarding pause ordered by the Central Bank of Nigeria in April 2024 and lifted after KYC remediation, 2024.

Arya Tschand, Arun Tejusve Raghunath Rajan, Sachin Idgunji, Anirban Ghosh, Jeremy Holleman, Csaba Király, et al. MLPerf Power: Benchmarking the energy efficiency of machine learning systems from microwatts to megawatts for sustainable AI. In *IEEE International Symposium on High-Performance Computer Architecture (HPCA 2025)*, 2025.

Paul R. Verkuil. *Outsourcing Sovereignty: Why Privatization of Government Functions Threatens Democracy and What We Can Do About It*. Cambridge University Press, 2007.

Michael A. Witt. De-globalization: Theories, predictions, and opportunities for international business research. *Journal of International Business Studies*, 50(7):1053–1077, 2019.

World Bank. World development indicators: GNI per capita (Atlas method), PPP conversion factors and official exchange rates, 2024 values. Washington, DC. Indicators NY.GNP.PCAP.CD, PA.NUS.PPP, PA.NUS.FCRF (API, accessed 21 September 2026), 2024.

A THE JPT COLLATION, WITH SOURCES

Table 3 lists the values plotted in Figure 1. Conversions use 1 Wh = 3600 J and are the sources' own units restated; "measured", "self-report", and "estimate" follow each source's characterization of its own numbers.

Table 3: Published per-task inference energy for text generation, as reported.

Source	System, task	Boundary	As reported	Joules	Character
ML.ENERGY v3.0 (Chung et al., 2025)	open-weight LLMs on H100, per output token	chip	0.15–0.4 J/token	(rate)	measured
Luccioni et al. (2024)	text generation, single A100, unbatched	chip	0.042 Wh	151	measured
Google (2025a)	median Gemini text prompt	chip (active accelerator)	0.10 Wh	360	self-report
Google (2025a)	same prompt	node+PUE (PUE 1.09)	0.24 Wh	864	self-report
Oviedo et al. (2026)	frontier text query, deployment-scale median	node+PUE	0.31 Wh	1,116	estimate
Jegham et al. (2025)	GPT-4o, short prompt	node+PUE	0.42 Wh	1,512	estimate
Jegham et al. (2025)	GPT-4.1 nano, long prompt	node+PUE	0.454 Wh	1,634	estimate
Jegham et al. (2025)	GPT-4o, long prompt	node+PUE	1.788 Wh	6,437	estimate
Jegham et al. (2025)	GPT-4.5, long prompt	node+PUE	30.5 Wh	109,800	estimate
Jegham et al. (2025)	DeepSeek-R1, long prompt (reasoning)	node+PUE	33.6 Wh	120,960	estimate
Jegham et al. (2025)	o3, long prompt (reasoning)	node+PUE	39.2 Wh	141,120	estimate

B PAI: PANEL VALUES, THE THREE-STAGE EXERCISE, AND THE METERED EXTENSION

Panel. Table 4 gives the inputs and outputs for Figure 2: 2024 GNI per capita (Atlas), the PPP-to-market-rate ratio k_j, PAI at US$20, the relative burden, the PPP-discounted price ($20k_j$), the welfare-cap price ($0.02\,y_j/12$), and which stage binds. Reference burdens: United States 0.29% ($y = 82{,}910$), China 1.76%, world 1.79%. The cost floor of the third stage is omitted because no vendor discloses marginal inference cost; where it exists it can only raise the final price.

Table 4: The 20-economy panel (World Bank 2024 values, accessed 21 September 2026).

Economy	GNI pc	k_j	PAI (%)	Burden (×US)	PAI after PPP (%)	$20k_j$ (US$)	Cap price (US$)	Binds
Mexico	12,760	0.542	1.9	6.5	1.0	10.83	21.27	PPP
Brazil	9,930	0.461	2.4	8.3	1.1	9.22	16.55	PPP
Peru	7,530	0.477	3.2	11.0	1.5	9.54	12.55	PPP
Colombia	7,090	0.354	3.4	11.7	1.2	7.09	11.82	PPP
South Africa	6,110	0.405	3.9	13.6	1.6	8.10	10.18	PPP
Indonesia	4,920	0.299	4.9	16.9	1.5	5.98	8.20	PPP
Vietnam	4,490	0.288	5.3	18.5	1.5	5.76	7.48	PPP
Philippines	4,470	0.337	5.4	18.5	1.8	6.75	7.45	PPP
Egypt	3,510	0.138	6.8	23.6	0.9	2.75	5.85	PPP
Bangladesh	2,820	0.258	8.5	29.4	2.2	5.16	4.70	cap
Cambodia	2,550	0.329	9.4	32.5	3.1	6.58	4.25	cap
India	2,550	0.244	9.4	32.5	2.3	4.89	4.25	cap
Ghana	2,260	0.301	10.6	36.7	3.2	6.02	3.77	cap
Kenya	2,070	0.321	11.6	40.1	3.7	6.42	3.45	cap
Nigeria	1,710	0.119	14.0	48.5	1.7	2.38	2.85	PPP
Nepal	1,480	0.253	16.2	56.0	4.1	5.06	2.47	cap
Pakistan	1,430	0.240	16.8	58.0	4.0	4.80	2.38	cap
Tanzania	1,210	0.282	19.8	68.5	5.6	5.64	2.02	cap
Ethiopia	1,100	0.328	21.8	75.4	7.2	6.56	1.83	cap
Uganda	1,010	0.331	23.8	82.1	7.9	6.61	1.68	cap

Metered services. Table 5 prices two reference baskets at list prices accessed on 21 September 2026 (DeepSeek, 2026; OpenAI, 2026; Google, 2026) and counts the panel economies in which the resulting PAI is at or under 2%. Two features of current list prices matter for auditing: one provider prices by time of day, with peak hours defined in UTC around its home business day, so the same basket has two PAIs; and every provider's smallest tier clears the line everywhere under light use, while flagship tiers under heavy use clear it in at most half the panel, and the most expensive tier nowhere. A PAI for a metered service is therefore meaningless without its basket, which is why the basket is part of the record.

Table 5: PAI for metered services under a light basket (0.2M input, 0.05M output tokens per month) and a heavy basket (2M, 0.5M). List prices in US$ per million tokens.

Provider	Model (tier)	In	Out	Light US$/mo	$\leq 2\%$ (of 20)	Heavy US$/mo	$\leq 2\%$ (of 20)
DeepSeek	deepseek-flash (peak)	0.30	1.20	0.12	20	1.20	20
DeepSeek	deepseek-flash (off-peak)	0.15	0.60	0.06	20	0.60	20
DeepSeek	deepseek-v4-pro (peak)	1.32	3.96	0.46	20	4.62	10
OpenAI	GPT-5.6 Luna	0.20	1.20	0.10	20	1.00	20
OpenAI	GPT-5.6 Terra	2.00	12.00	1.00	20	10.00	5
OpenAI	GPT-5.6 Sol	5.00	30.00	2.50	15	25.00	0
Google	Gemini 3.5 Flash-Lite	0.30	2.50	0.18	20	1.85	18
Google	Gemini 3.8 Flash	0.75	3.75	0.34	20	3.38	14
Google	Gemini 3.1 Pro Preview	2.00	12.00	1.00	20	10.00	5

C PROOFS

Proposition 1. With $\mathrm{JpT}_b(s) = \kappa_b(s)\,\mathrm{JpT}_{\text{chip}}(s)$, the ranking at b reverses iff $\kappa_b(s_1)\mathrm{JpT}_{\text{chip}}(s_1) > \kappa_b(s_2)\mathrm{JpT}_{\text{chip}}(s_2)$, i.e. iff $\kappa_b(s_1)/\kappa_b(s_2) > \mathrm{JpT}_{\text{chip}}(s_2)/\mathrm{JpT}_{\text{chip}}(s_1)$, which

exceeds 1 by hypothesis. If $\kappa_b(s_1) \leq \kappa_b(s_2)$ the inequality cannot hold, so unequal overhead is necessary; if $1 < \kappa_b(s_1)/\kappa_b(s_2) \leq \mathrm{JpT}_{\text{chip}}(s_2)/\mathrm{JpT}_{\text{chip}}(s_1)$ it does not hold, so unequal overhead is not sufficient. Under a fixed $(b, \mathcal{T}, H)$ both systems are reported under the same convention and the order is that of JpT_b itself. □

Proposition 2. (i) Enumerate the rules: a higher PAI, a lower disclosure level, a contradicted disclosure, a documented adverse finding, a failed gate condition, and a documented veto trigger each map to a weakly lower grade (P→N→F, P→W, any→F) or leave U unchanged; no rule maps adverse evidence upward. (ii) Each axis grade is a function of that axis's instrument and its own veto flags; the gate is a function of jurisdictional facts and only lowers. (iii) The rule set's outputs are the tuple of axis grades and flags; it defines no mapping from that tuple to a scalar, so "compensation" would have to be a function outside the rule set, whose value is not a verdict of the standard. □

D Focal-case inputs and sources

Open-weight release and hosted service. Weights and distilled variants: DeepSeek-AI (2025; 2024). Metered prices: DeepSeek (2026) (light basket at peak rates, US$0.12 per month; PAI 0.08% in Nigeria, 0.06% in Cambodia). Energy references: 33.6 Wh per long reasoning prompt at node+PUE (Jegham et al., 2025); per-token measurements (Chung et al., 2025). DVR: storage on servers in the provider's home jurisdiction and consent regime disclosed in the privacy policy (DeepSeek, 2025); revenue retention and portability undisclosed. Gate: Nigeria, Data Protection Act 2023 and the NDPC (Nigeria, 2023); Cambodia, draft law only (Cambodia MPTC, 2025). Telemetry: Microsoft AI Economy Institute (2026), whose economy-level map bins DeepSeek's market share among generative-AI users (most African economies, Nigeria included, in the 11–15% band; a few East and Southern African economies higher; no exact African percentages are printed), states that African usage is estimated at two to four times other regions, and attributes part of African uptake to promotion and telecom-partner integration by Chinese firms—a supply-push reading we note as a rival to the accessibility reading.

Data-centre build-out. Sites and financing: Reuters (2024); Huawei Cloud (2024); Konza Technopolis (2025). E: nominal product PUE $\leq$ 1.15; per-site measured PUE undisclosed for every African site we could find. G: onshore storage disclosed for the national cloud; consent, revenue retention, and exit undisclosed; veto on the public record of safe-city deployments by the same vendor in the host country and the absence of a disclosed oversight or exit instrument. Gate: Kenya, Data Protection Act 2019 and the ODPC (Kenya, 2019).

Platform-embedded system. Users and hosting: Huawei Cloud (2025). Supervisory capacity: the April–June 2024 onboarding freeze and its lifting (TechCabal, 2024). DVR: onshore storage under Nigerian financial-data rules and consent disclosed in the platform's privacy policy; revenue retention and portability undisclosed. Gate: Nigeria (2023). The biometric and transactional corpus is recorded as a monitoring item; the veto requires the absence of independent oversight, which is not the case.

E The CAIGS-92 dataset

CAIGS-92 (Chinese AI in the Global South, v3.0) is an event-level compilation of 92 documented, landed engagements by 39 Chinese AI firm groups across 24 countries, one row per documented deployment (a firm group may appear more than once in a country), assembled from public sources with one URL-carrying source record per row (91 of 92 rows rest on two or more sources; median 3). Rows were screened from a 144-candidate pool under three inclusion criteria (Chinese or verifiably Chinese-backed firm; directly AI-related; landed deployment) and then re-verified against their sources, a pass that corrected 287 fields; reliability is graded high for 80 rows, medium for 10, and low for 2, and the 52 excluded Global-South candidates are listed with reasons. Country-level context (data-protection law, AI strategy, income group) is joined from a 37-country reference table with in-cell vintages. The dataset, its codebook and reference table, the exclusion list, and the axis-coding rule are released with the paper.

Scope and limits: the dataset is a census of documented engagements, not of all engagements; compiled from public reporting, it skews toward large or announced projects and includes only landed

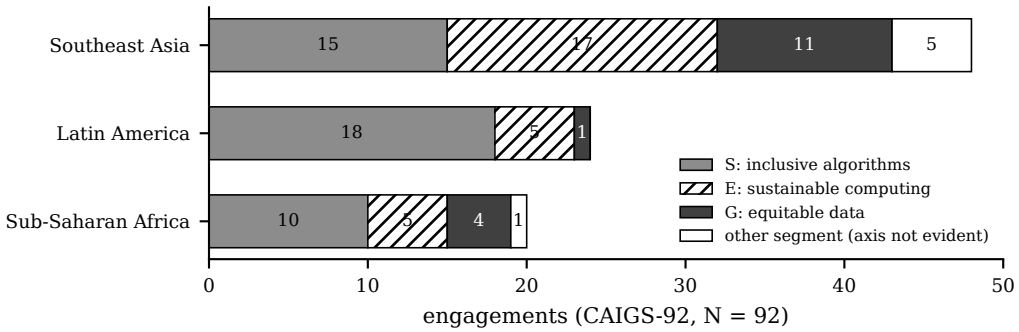

Figure 3: CAIGS-92 engagements by region and by the diffusion axis most directly engaged, coded from the deployment's principal AI segment (cloud and computing infrastructure and AI hardware → E; fintech and public-security systems → G; consumer and sectoral algorithms, models, and robotics → S; other → not evident). The rule is a coarse first-pass, released with the data; it returns "not evident" for six rows.

deployments by design. Only 23 of 92 rows carry a company-disclosed scale figure (3 rest on media estimates and 66 disclose none or leave the field uncoded); disclosed figures mix investment and transaction-volume semantics, so we aggregate no monetary totals. Entry timing is bimodal (27 entries in 2018–19; 27 in 2024–26). The dataset is used descriptively; it identifies no causal effect and cannot test the governance gate.

F THE MOTIVATING CONJECTURE, ITS RIVALS, AND THE FORWARD-TEST PROTOCOL

The conjecture. The standard was motivated by an empirical conjecture about supplier behaviour under fragmented conditions: that for a subset of suppliers, accessibility, energy, and data-governance criteria cease to be external compliance overlays and become internal design parameters of the systems they export (efficiency as a business model, inclusion as a market-entry precondition, data governance as trust infrastructure). This is a claim about firms, testable against rivals, and not a normative endorsement; the standard does not depend on it.

Rivals and discriminating indicators. Rivals are cost arbitrage (firms follow price and margin), geopolitical substitution (export-control-forced bloc alignment), state-capital export (surplus capacity), regulatory arbitrage (depth where governance is weakest), market segmentation (price localization without capability transfer), and institutional-quality attraction of foreign direct investment (profit-seekers also prefer strong institutions). Three indicators discriminate: local capability building (corpus co-building, talent programmes, localization), which arbitrage and capital export do not predict; diffusion depth tracking host governance rather than camp alignment, which geopolitical substitution and regulatory arbitrage do not predict; and local data-value retention (onshore storage, revenue sharing, portability), which every rival predicts to be one-way outflow. On the dataset the first has only a coarse proxy (89 of 92 engagements name a local partner, 12 a university; the free-text localization notes are released but not yet coded); the second is consistent with the corridor-level variation we describe qualitatively below; and the third cannot be scored on CAIGS-92, which codes no data-governance fields—it is observable only on the focal cases, where onshore storage is disclosed in two of three and revenue sharing in none. The investment-attraction rival is observationally equivalent on depth alone and separates only jointly on the first and third indicators, the third of which is partial; the conjecture is therefore *not* currently distinguished from an institutional-quality story, and we say so rather than restate a decision rule that cannot fire.

Forward-test protocol. *Object:* non-Chinese-supplied AI-compute facilities in Africa landed between 2025 and 2027 (the South African GPU facilities and their announced Nigerian and Kenyan extensions). *Date:* first quarter of 2028. *Instruments:* E (per-site measured PUE, WUE, grid intensity)

and G (DVR, oversight and exit instruments) under Table 1, by the released code. *Prediction:* within a host jurisdiction, G components and veto outcomes do not differ by supplier origin; across hosts they track gate status and disclosed oversight instruments. *Refutation:* adverse G components concentrating by supplier regardless of host. *Evidence classes:* regulator filings, procurement contracts, facility disclosures, and the same public-record reconstruction used for the focal cases, with every value labelled measure, verify, or reference.

G GAMING ROUTES AND CLOSING RULES

Table 6: Attack surface of the instruments and the rule that closes each route.

Attack	Mechanism	Closing rule
Boundary shopping	report JpT at the chip boundary only	boundary tuple mandatory (Proposition 1)
Harness shopping	choose a favourable prompt suite	fixed public suite; otherwise declared non-comparable
Basket shopping	price an API on an unrepresentative basket	basket disclosed and recorded; two reference baskets
Tier gaming	a nominal zero-price tier escapes PAI	zero-price artifacts graded on hardware floor and language coverage
Disclosure theatre	DVR components "disclosed" but unverifiable	$\mathsf{D}^{+} \prec \mathsf{V}$; P requires two verified components
Composite lobbying	pressure to publish a single ranking	no aggregate exists (Proposition 2)
Gate gaming	route the nominal jurisdiction through a gate-passing host	the island is the jurisdiction of the users served

H GLOSSARY

Diffusion deficit: the gap between converging frontier capability and diverging deployment capacity. **Diffusion capacity**: the conjunction of accessibility, adaptability, and institutional compatibility (Section 3). **S, E, G**: the three axes (inclusive algorithms, sustainable computing, equitable data). **JpT, PAI, DVR**: the three instruments. **Boundary tuple**: ⟨chip, node, node+PUE⟩. **Auditor modes**: measure, verify, reference. **Governance gate**: conditions (a)–(b) of Table 1. **Veto cap**: a documented adverse finding capping an axis at its floor, with a flip condition. **Flip condition**: the disclosure that changes a verdict component. **CAIGS-92**: the dataset of Appendix E.