

The 2026 LLM System Decision Manual

A blank operating system for signed workload decisions-not a ranking, forecast, or approval.

One authority chain, no implied winner

2026-08-29	2026-08-29	2026-12-31
IMPAKT Exact system/configuration Signed workload shortlist Shared confirmation pilot + common gates D/P/R/A result Comparative frontier + selected-posture fit Exact route/register + signed decision		
UNINSTANTIATED · UNKNOWN · N/S · NO AUTHORITY		
2026-08-29 -> 2026-12-31 · 30 pages · 38 screening models + 7 seat options · public outcomes N/S		
TAKEAWAY Start with one signed workload shortlist; SCALE only after the full comparison and authority chain recomputes.	DECISION RULE No system advances from public facts. The signed shortlist, shared confirmation evidence, executive stack, and decision attestation must all close.	
Operating note Start with one signed workload shortlist; SCALE only after the full comparison and authority chain recomputes.	BOUNDARY Evidence begins 2026-08-29 and expires by 2026-12-31. Public company outcomes remain N/S; all released forms are UNINSTANTIATED/UNKNOWN.	

Four artifacts before one approval

Eligibility is a signed shortlist-not a score. Frontier, economics, and fit then use the same frozen cases and exact systems.

The four-artifact executive chain

Signed workload shortlist	COMPLETE / candidate included	UNKNOWN
Comparative performance-cost frontier	complete comparable rows / lead or tie	UNKNOWN
Four-view economics ledger	signed incumbent + eligible comparators	UNKNOWN
Selected-posture fit result	normalized + uncertainty + sensitivity	UNKNOWN
shortlist	shared sample/protocol	common+additive gates
D/P/R/A result	frontier	fit
exact route/register	rollback	decision attestation
Decide one workload x data class x exact system; mixed API/LOCAL/SEAT/HYBRID winners are permitted.		
TAKEAWAY A candidate can lead one workload and lose another; the manual never forces a portfolio-wide winner.	DECISION RULE Apply the order non-compensatorily: shortlist eligibility -> comparative frontier -> economics -> one selected-posture fit. Missing evidence yields UNKNOWN.	
Operating note A candidate can lead one workload and lose another; the manual never forces a portfolio-wide winner.	BOUNDARY SCALE additionally requires the applicable D/P/R/A result, exact route/register, rollback, gates, and signed decision packet.	

BLANK WORKLOAD SHORTLIST - SIGN BEFORE PILOT

All 45 rows close as SHORTLISTED, EXCLUDED_HARD_CONSTRAINT, or ELIGIBLE_NOT_SELECTED. Capacity limits selection without fabricating exclusions.

Signed shortlist passport

UNINSTANTIATED													
SHORTLIST ID				DIGEST				WORKLOAD					
DATA				POSTURE				MAX					
S	MAX2	RUN1	DIR	P100	C4510	x1	N500	>=3	C450	D1500	R1050	<=3000	
M	MAX3	RUN2	DIR	P100	C4510	x1	N500	>=3	C450	D3000	R1050	<=4500	
B	MAX4	RUN3	DIR	P100	C4510	x1	N500	>=3	C450	D4500	R1050	<=6000	
H	MAX6	RUN5	DIR	P100	C4510	x1	N500	>=3	C450	D7500	R1050	<=9000	
ALG ID		CAPACITY-RANK-v12		ALG SHA		e1134809339e		CAL ID		CAL SHA		BUD ID	BUDGET-[WORK]
BUD SHA		ELIG		HARD		NSEL		SHORT					
45: SHORTLIST / HARD EXCLUDE / ELIGIBLE NOT SELECTED · U UNTIL SIGNED													
RANK -> SIGNED TIE -> TOPOLOGY/CAPACITY -> SEEDED SHA													
10-CASE SCREEN · DISJOINT 500-CASE CONFIRM · >=3 ATTEMPTS													
CAND	CFG	TOP	STATE	RANK/TIE	FALS	BUDGET	CASESxATT						
CAND	CFG	TOP	STATE	RANK/TIE	FALS	BUDGET	CASESxATT						
CAND	CFG	TOP	STATE	RANK/TIE	FALS	BUDGET	CASESxATT						
CAND	CFG	TOP	STATE	RANK/TIE	FALS	BUDGET	CASESxATT						
CAND	CFG	TOP	STATE	RANK/TIE	FALS	BUDGET	CASESxATT						
CAND	CFG	TOP	STATE	RANK/TIE	FALS	BUDGET	CASESxATT						
CAND	CFG	TOP	STATE	RANK/TIE	FALS	BUDGET	CASESxATT						
CAND	CFG	TOP	STATE	RANK/TIE	FALS	BUDGET	CASESxATT						
OWNER				SIGNER									
SIGNATURE				SIGN DATE									
DECISION DATE				REOPEN									
A2 shortlist_frontier_fit_v12.json / shortlist_schema + shortlists[*].shortlist_payload.candidate_rows.													
Passport binds COMPLETE 45-row A2 disposition-annex digest; not only six visible slots.													
Manifest, full embedded name, SHA-256, byte count, and extraction instruction: p30.													
TAKEAWAY				DECISION RULE									
Signed calibration ranks eligible systems reproducibly; disjoint confirmation decides outcomes only after the bounded shortlist is signed.				Apply signed hard constraints first, then the exact calibration metric, tolerance, topology/capacity policy, and seeded tie-break. Select only the bounded top rows and reconcile the all-in ledger.									
Operating note				BOUNDARY									
Signed calibration ranks eligible systems reproducibly; disjoint confirmation decides outcomes only after the bounded shortlist is signed.				Hard exclusion alone needs constraint evidence. Eligible nonselection needs calibration evidence, rank/tie disposition, reason, owner/date, and reopen rule; all released values remain blank.									

Sign one all-in workload budget

Small/Medium/Big/Huge defaults bound the whole selection program-not confirmation alone-and keep every resource dimension explicit.

Signed all-in workload budget

S	MAX2	RUN1	DIR	P100	C4510	x1	N500	>=3	C450	D1500	R1050	<=3000
M	MAX3	RUN2	DIR	P100	C4510	x1	N500	>=3	C450	D3000	R1050	<=4500
B	MAX4	RUN3	DIR	P100	C4510	x1	N500	>=3	C450	D4500	R1050	<=6000
H	MAX6	RUN5	DIR	P100	C4510	x1	N500	>=3	C450	D7500	R1050	<=9000

CODES: S/M/B/H posture · MAX max candidates/workload · RUN executable candidates · DIR DIRECT_ONLY

P shared calibration-pool cases · C screening cases/eligible system · x screening attempts/case

N unique confirmation cases · >= minimum independent attempts/confirmation · C45 calibration executions at 45 eligible

D direct confirmation executions · R nonconfirmation reserve · <= maximum all-in executions

A2 shortlist_frontier_fit_v12.json / capacity_scaled_defaults + all_in_budget_*

A3 IMPAKT_LLM_Decision_Workbook_2026_v1.json / payload.posture_selection_rubric_v1 · manifest p30.

	CASES/ATT	HUMAN H	GPU H/KWH	CASH/FX	DATA EXP
CAL					
DIRECT					
FACTORIAL					
SEAT					
LOCAL					
ROLLBACK					
OBS					

ATTEMPTS = UNITS x REPEATS · DIRECT = SHORTLIST x 500 x >=3 + REPEATS

SELECT SMALLEST PRESET MEETING SIGNED MINIMUM SCOPE; ALL 7 TOTALS <= SIGNED CEILINGS

INSUFFICIENT RESOURCES = STOP-REDESIGN · N/A = SIGNED 0 · ONE FX POLICY

RISK U · HARD RISK->HUGE/RISK CONTROLS · CUSTOM BOUNDED->NAMED POSTURE · NO CONTROL WEAKENING

SELECT SHA	LEDGER ID	LEDGER SHA	POSTURE	MODE
ATT TOTAL	ATT CAP	ALL CAPS	OWNER	SIGN/DATE

TAKEAWAY

Choose one posture and factorial mode, then sign all seven stage rows, exact sums, ceilings, units, currency/FX, and evidence before execution.

Operating note

Choose one posture and factorial mode, then sign all seven stage rows, exact sums, ceilings, units, currency/FX, and evidence before execution.

DECISION RULE

Reconcile calibration, direct confirmation, incremental factorial, seat adoption, LOCAL load, rollback drills, and observation. Every resource total must be within its signed cap.

BOUNDARY

Defaults are non-ranking planning limits. Missing, mixed-unit/currency, or over-cap values remain UNKNOWN/FAIL; HUGE_RISK_OVERRIDE adds controls but never budget or candidates automatically.

One denominator: accepted useful output

Count every eligible task and generation attempt. Reward only completed outputs that pass signed quality and policy criteria through observation.

Zero-crossing AUO accounting

AUO_rate = Accepted_useful_outputs / All_eligible_tasks		Cost_per_AUO = Fully_loaded_cost(all attempts, failures, retries, fallbacks, evaluations, reviews, corrections) / Accepted_useful_outputs		Intention_to_treat_denominator = All_eligible_tasks, including timeout, abandonment, and failure	
Every generation call, including retry, rejected draft, fallback, evaluation, and shadow call.					
If accepted useful outputs = 0, Cost per AUO is not finite; report the loss and never suppress the zero.					
events_retained_in_workload_attempt_and_cost_ledgers					
A critical error discovered inside the signed observation window revokes acceptance.					
left_to_right					
Task / volume	eligible tasks	task class	volume		
observation window	Interaction / load	attempts	retries		
rejected drafts	evaluation/shadow	Service / recovery	fallback		
timeout	abandonment	exception	rollback		
Human / error cost	review	rework	correction		
downstream error					
AUO		Accepted useful output only			
signed workload		acceptance rubric		data class	
owner		budget		expiry	
Non-crossing: every attempt enters the cost ledger; only accepted tasks cross the AUO gate.					
Method: M005, R005. Full trace: page 25. Method: R001.					

TAKEAWAY AUO is an outcome; attempts are workload and cost ledger events.	DECISION RULE Keep every eligible task and attempt in its required denominator or ledger.
Unit rule AUO is an outcome; attempts are workload and cost-ledger events.	BOUNDARY Company numerator, denominator, quality, and cost values remain N/S.

Keep seats, APIs, local, and Harness separate

Seats, API routes, checkpoints, generic scaffolds, and capital-H modules are separate procurement and control objects. Keep the incumbent visible.

Nine architectures with explicit product boundaries

ID + topology	Seat / API / local	Generic scaffold	Capital-H state
T0 - Incumbent/human	NO / NO / NO	WORKFLOW AS-IS	OPTIONAL
T1 - Business seats	YES / NO / NO	PRODUCT WORKFLOW	NO
T2 - Direct frontier API	NO / YES / NO	APPLICATION-OWNED	NO
T3 - Managed open-weight endpoint	NO / YES / NO	APPLICATION-OWNED	OPTIONAL
T4 - Company-controlled local/cloud	NO / NO / YES	APPLICATION-OWNED	OPTIONAL
T5 - Local default, frontier escalation	NO / YES / YES	ROUTING SCAFFOLD	OPTIONAL
T6 - Frontier default, local sensitive path	NO / YES / YES	ROUTING SCAFFOLD	OPTIONAL
T7 - Multi-provider through application-owned gateway/generic routing scaffold	NO / MULTIPLE / OPTIONAL	APPLICATION-OWNED ROUTING	OPTIONAL DOCUMENTED MODULES
T8 - Same exact route/scaffold; capital-H documented modules off/on	ROUTE + MODEL FIXED	UNCHANGED	OFF / ON FACTORIAL

YES	NO	OPTIONAL
UNCHANGED	MULTIPLE	Hybrid economics and ownership continue on page 11.
AI-cost telemetry and FinOps	natural-language AI Rules guidance	OPA/Rego checks on supported feature resources
documented MLOps pipeline pattern	Coverage, enforcement, LLM routing, quality, savings, effectiveness, and outcomes are N/S.	Evidence: R001, R002, P001-P003, H001-H007. Full trace: pages 25-30.

TAKEAWAY Coding, marketing, and ingestion may rationally select different systems.	DECISION RULE Eliminate a topology only through a documented hard constraint.
Ownership Application code owns workload logic and routing; enabled capital-H modules remain separately evidenced and costed.	BOUNDARY Common API shape is not portability; documented modules do not prove effectiveness.

The calendar is a signed evidence constraint

Every governing window now has a nonblank workflow plus exact sample, protocol, and gate bindings.
Five review days means calendar days in America/Mexico_City.
Executable concurrent calendar

CODING · E 2026-10-12->2026-11-08 · O 2026-10-12->2026-11-08

MARKETING · E 2026-10-12->2026-11-08 · O 2026-10-12->2026-11-08

DATA_INGESTION · E 2026-10-12->2026-11-08 · O 2026-10-12->2026-11-08

GENERAL_KNOWLEDGE · E 2026-10-12->2026-11-08 · O 2026-10-12->2026-11-22

SIGNED-REVIEW-DAY-ARITHMETIC-v12 CALENDAR_DAYS America/Mexico_City NOT_APPLICABLE

Latest signed observation end + 5 Mexico City calendar days.

2026-11-22T00:00:00-06:00

2026-11-27T00:00:00-06:00

OWNER

SIGNER

SIGNATURE

DATE

2026-11-27T00:00:00-06:00

2026-11-30T00:00:00-06:00

2026-12-28T00:00:00-06:00

2026-12-31

SCALE

CAPPED PILOT-EXTEND-TIE

STOP-REDESIGN

TAKEAWAY

Decision timing derives from the longest signed observation-not from an unbound calendar note.

DECISION RULE

Run evaluation and observation concurrently. Decide no earlier than 2026-11-27 only after the governing general-knowledge sample/protocol/gates are COMPLETE and five local calendar days elapse.

Operating note

Decision timing derives from the longest signed observation-not from an unbound calendar note.

BOUNDARY

An incomplete governing or workload protocol yields CAPPED PILOT-EXTEND-TIE or STOP-REDESIGN; no SCALE date. Rollout begins after decision and reaches 100% no earlier than 2026-12-28.

Candidate systems: exact routes, not brand names

The 38 screening models remain the bounded research universe; seven seat options are separate decision-capable systems on page 15.

Exact model screening catalog

SYSTEM · TOPOLOGY · ROUTE/ARTIFACT · PRECISION · CONFIG · STATE					
OAI-SOL-A	API	gpt-5.6-sol	N/S	CFG-OAI-SOL-A-v1	N/S
OAI-TERRA-A	API	gpt-5.6-terra	N/S	CFG-OAI-TERRA-A-v1	N/S
OAI-LUNA-A	API	gpt-5.6-luna	N/S	CFG-OAI-LUNA-A-v1	N/S
ANT-FABLE-A	API	claude-fable-5	N/S	CFG-ANT-FABLE-A-v1	N/S
ANT-OPUS-A	API	claude-opus-5	N/S	CFG-ANT-OPUS-A-v1	N/S
ANT-SONNET-A	API	claude-sonnet-5	N/S	CFG-ANT-SONNET-A-v1	N/S
HY4-A	API	hy4-preview	N/S	CFG-HY4-A-v1	N/S
MM-M3-A	API	MiniMax-M3	N/S	CFG-MM-M3-A-v1	N/S
QW-27-A	API	qwen3.8-27b	N/S	CFG-QW-27-A-v1	N/S
QW-MAX-A	API	qwen3.8-max	N/S	CFG-QW-MAX-A-v1	N/S
QW-FLASH-A	API	qwen3.8-flash	N/S	CFG-QW-FLASH-A-v1	N/S
GLM-FULL-A	API	glm-5.3	N/S	CFG-GLM-FULL-A-v1	N/S
GLM-FLASH-A	API	glm-5.3-flash	N/S	CFG-GLM-FLASH-A-v1	N/S
DS-FLASH-A	API	deepseek-v4-flash	N/S	CFG-DS-FLASH-A-v1	N/S
DS-PRO-A	API	deepseek-v4-pro	N/S	CFG-DS-PRO-A-v1	N/S
KIMI-K3-A	API	kimi-k3	N/S	CFG-KIMI-K3-A-v1	N/S
GEM-FLASH-A	API	gemini-3.7-flash	N/S	CFG-GEM-FLASH-A-v1	N/S
HY4-BF16-L	LOCAL	tencent/Hy4-preview	BF16	CFG-HY4-BF16-L-v1	N/S
HY4-FP8-L	LOCAL	tencent/Hy4-preview-FP8	FP8	CFG-HY4-FP8-L-v1	N/S
MM-M3-BF16-L	LOCAL	MiniMaxAI/MiniMax-M3	BF16	CFG-MM-M3-BF16-L-v1	N/S
MM-M3-MXFP8-L	LOCAL	MiniMaxAI/MiniMax-M3-MXFP8	MXFP8	CFG-MM-M3-MXFP8-L-v1	N/S
MM-M3-NVFP4-L	LOCAL	nvidia/MiniMax-M3-NVFP4	NVFP4	CFG-MM-M3-NVFP4-L-v1	N/S
QW-27-BF16-L	LOCAL	Qwen/Qwen3.8-27B	BF16	CFG-QW-27-BF16-L-v1	N/S
QW-27-FP8-L	LOCAL	Qwen/Qwen3.8-27B-FP8	FP8	CFG-QW-27-FP8-L-v1	N/S
QW-MAX-BASE-L	LOCAL	Qwen/Qwen3.8-2.4T-A95B	BASE	CFG-QW-MAX-BASE-L-v1	N/S
QW-MAX-FP8-L	LOCAL	Qwen/Qwen3.8-2.4T-A95B-FP8	FP8	CFG-QW-MAX-FP8-L-v1	N/S
QW-FLASH-BASE-L	LOCAL	Qwen/Qwen3.8-Flash-Next	M-BASE	CFG-QW-FLASH-BASE-L-v1	N/S
QW-FLASH-FP8-L	LOCAL	Qwen/Qwen3.8-Flash-Next-FP8	FP8	CFG-QW-FLASH-FP8-L-v1	N/S
GLM-FULL-MIXED-L	LOCAL	zai-org/GLM-5.3	MIXED	CFG-GLM-FULL-MIXED-L-v1	N/S
GLM-FULL-BF16-L	LOCAL	zai-org/GLM-5.3-BF16	BF16	CFG-GLM-FULL-BF16-L-v1	N/S
GLM-FLASH-MIXED-L	LOCAL	zai-org/GLM-5.3-Flash	mixed FP8	CFG-GLM-FLASH-MIXED-L-v1	N/S
GLM-FLASH-BF16-L	LOCAL	zai-org/GLM-5.3-Flash-BF16	BF16	CFG-GLM-FLASH-BF16-L-v1	N/S
DS-FLASH-MIXED-L	LOCAL	deepseek-ai/DeepSeek-V4-Flash-0731	M-CKPT	CFG-DS-FLASH-MIXED-L-v1	N/S
DS-PRO-L	LOCAL	deepseek-ai/DeepSeek-V4-Pro-0813	M-CKPT	CFG-DS-PRO-L-v1	N/S
MUSE-BF16-L	LOCAL	meta-models/Muse-Glimmer-30B	BF16	CFG-MUSE-BF16-L-v1	N/S
MUSE-Q4KM-L	LOCAL	R36	Q4_K_M	CFG-MUSE-Q4KM-L-v1	N/S
MUSE-Q4KXL-L	LOCAL	R37	Q4_K_XL	CFG-MUSE-Q4KXL-L-v1	N/S
KIMI-K3-MXFP4-L	LOCAL	R38	MXFP4	CFG-KIMI-K3-MXFP4-L-v12	N/S
SEAT-CHATGPT-BUSINESS		SEAT-CLAUDE-TEAM-STANDARD		SEAT-CLAUDE-TEAM-PREMIUM	
SEAT-CLAUDE-ENTERPRISE		SEAT-MINIMAX-PLUS		SEAT-MINIMAX-MAX	
SEAT-MINIMAX-ULTRA					
R36 = meta-models/Muse-Glimmer-30B-GGUF Q4_K_M + companions					
R37 = meta-models/Muse-Glimmer-30B-GGUF dynamic Q4_K_XL + companions					
R38 = moonshotai/Kimi-K3 MXFP4 screening variant; exact revision/files N/S until instantiated					
PRECISION: BASE = base precision · M-BASE = mixed base · Q4_K_XL = dynamic Q4_K_XL					
PRECISION: MIXED = mixed precision · M-CKPT = mixed checkpoint					
KIMI=MXFP4 ONLY · 38 MODELS (17 API / 21 LOCAL) · 7 SEATS					
TAKEAWAY			DECISION RULE		
Every formal test uses an exact configuration digest; family names and public bytes never substitute.			Screen by exact candidate/configuration ID, then bind one workload/data class. The repaired Kimi row is MXFP4 only; exact revision/files remain required before COMPLETE.		
Operating note			BOUNDARY		
Every formal test uses an exact configuration digest; family names and public bytes never substitute.			Public performance, fit, savings, and readiness remain N/S. HYBRID is assembled only from separately shortlisted exact components.		

Unknown is a stop unless strictly measurable

Every workload candidate receives the same four core gates; additive falsifiers and topology/risk gates can only add stops.

Fail-closed admission and exact gate hashes

Whole gate-support file SHA-256		b0ad4300cd3c8aeda4964c04672827cf6e71529483b4d820aff6bd01940a2c8e		
Expected-set catalog hash		0c06eb94fd61713a04a121249bfd6f827585e3c9d3ca8aad60fab0b47df2aed7		
Actual-template catalog hash		2c313e3a3ac2532a68a73734ddaf99d3e37e0577bfb15b95f8274d0b293dfdbe		
Whole-artifact canonical hash		1c4caee81f1853a056d0f21ee78f8819f3563631984b4ba97cc58263e41ff1c4		
coding	CORE-CODING-QUALITY-v12	CORE-CODING-SAFETY-v12	CORE-CODING-SERVICE-v12	CORE-CODING-AUO-ECONOMICS-v12
marketing	CORE-MARKETING-QUALITY-v12	CORE-MARKETING-SAFETY-v12	CORE-MARKETING-SERVICE-v12	CORE-MARKETING-AUO-ECONOMICS-v12
data_ingestion	CORE-DATA-INGESTION-QUALITY-v12	CORE-DATA-INGESTION-SAFETY-v12	CORE-DATA-INGESTION-SERVICE-v12	CORE-DATA-INGESTION-AUO-ECONOMICS-v12
PASS		typed signed threshold met; evidence valid		
FAIL		stop/redesign; never waived to SCALE		
UNKNOWN		missing/incomplete/incomparable; blocks SCALE		
N/S		not publicly substantiated; not evidence		
shortlist	shared sample/protocol		common+additive gates	
factorial	frontier		selected-posture fit	
route/register	rollback		decision attestation	
A1 expected_gate_catalog_v12.json / common_core_gate_catalog + expected_gate_sets. A1 / actual_gate_set_templates: metric/unit/denominator/operator/threshold/evidence-grade/expiry/owner/result templates. Extract in PDF Attachments panel; manifest, full SHA-256, byte count, and instructions: p30.				
TAKEAWAY		DECISION RULE		
No stale or ambiguous hash can identify the controlling gate support.		Admit only a signed shortlisted exact system on the shared frame. Any core, crosscut, additive, sample, protocol, or executive-stack UNKNOWN blocks SCALE.		
Operating note		BOUNDARY		
No stale or ambiguous hash can identify the controlling gate support.		A strictly bounded, non-sensitive pilot may carry only signed runtime/quality/capacity/resilience UNKNOWNs; FAIL and non-waivable control UNKNOWNs stop.		

Public performance is N/S; cheapness is tariff-only

Tariffs price one declared reference attempt; they do not establish accepted useful outputs, equivalent-task cost, fit, or savings.

Illustrative reference attempt: 10K input + 2K output

MODEL / API TARIFFS	11	PLANS / LOCAL CLOCKS	10
OAI-SOL-A · gpt-5.6-sol · S001/S004 .008*\$4+.002*\$0.40+.002*\$20 0.0728 USD · T2/C · S>=11-21		GLM-FULL-A · glm-5.3 · S023 .008*\$1.40+.002*\$0.26+.002*\$4.40 0.02052 USD · T3/C · C0	
OAI-TERRA-A · gpt-5.6-terra · S002 .008*\$2+.002*\$0.20+.002*\$12 0.0404 USD · T3/C · C0		GLM-FLASH-A · promo G0 · glm-5.3-flash · S023 .008*\$0.075+.002*\$0.015+.002*\$0.25 0.00113 USD · T5/C · G0	
OAI-LUNA-A · gpt-5.6-luna · S003 .008*\$0.20+.002*\$0.02+.002*\$1.20 0.00404 USD · T5/C · C0		GLM-FLASH-A · list G1 · glm-5.3-flash · S023 .008*\$0.15+.002*\$0.03+.002*\$0.50 0.00226 USD · T5/C · G1	
ANT-FABLE-A · claude-fable-5 · S005 .010*\$10+.002*\$50 0.2000 USD · T1/C · C0		DS-FLASH-A · off-peak DO · deepseek-v4-flash · S030 .008*\$0.22+.002*\$0.007+.002*\$0.66 0.003094 USD · T5/C · DO	
ANT-OPUS-A · claude-opus-5 · S005 .010*\$5+.002*\$25 0.1000 USD · T2/C · C0		DS-FLASH-A · peak DP · deepseek-v4-flash · S030 .008*\$0.44+.002*\$0.014+.002*\$1.32 0.006188 USD · T4/C · DP	
ANT-SONNET-A · claude-sonnet-5 · S005 .010*\$2+.002*\$10 0.0400 USD · T3/C · C0		DS-PRO-A · off-peak DO · deepseek-v4-pro · S030 .008*\$0.66+.002*\$0.022+.002*\$1.98 0.009284 USD · T4/C · DO	
HY4-A · hy4-preview · S058 .008*RMB6+.002*RMB0.3+.002*RMB18 0.0846 RMB · N/S/C · HY		DS-PRO-A · peak DP · deepseek-v4-pro · S030 .008*\$1.32+.002*\$0.044+.002*\$3.96 0.018568 USD · T4/C · DP	
MM-M3-A · MiniMax-M3 · S046 .008*\$0.30+.002*\$0.06+.002*\$1.20 0.00492 USD · T5/C · C0		KIMI-K3-A · kimi-k3 · S052 .008*RMB20+.002*RMB2+.002*RMB100 0.3640 RMB · N/S/C · C0	
QW-27-A · qwen3.8-27b · S059 .008*\$0.50+.002*\$0.10+.002*\$3 0.0102 USD · T4/C · C0		GEM-FLASH-A · gemini-3.7-flash · S054 .010*\$0.75+.002*\$3.75 0.0150 USD · T4/C · GF	
QW-MAX-A · qwen3.8-max · S018 .008*\$2+.002*\$0.25+.002*\$6 0.0285 USD · T3/C · C0		MUSE-BF16-L · local NA · meta-models/Muse-Glimmer-30B · S048 Registry: hosted SKU/tariff N/S · NON-TARIFF	
QW-FLASH-A · qwen3.8-flash · S019 .008*\$0.15+.002*\$0.016+.002*\$0.47 0.002172 USD · T5/C · C0			

Reference attempt: 8K uncached + 2K cache-hit when published + 2K output; otherwise 10K standard input. Not an AUO.

T5 · 0 <= c <= \$0.005	T4 · \$0.005 < c <= \$0.020	T3 · \$0.020 < c <= \$0.050	T2 · \$0.050 < c <= \$0.100	T1 · c > \$0.100
C0 · C0 · current at cutoff	HY · HY · Guangzhou 2026-08-28	G0 · G0 · through 2026-09-09 24:00 UTC+8	G1 · G1 · after 2026-09-09 24:00 UTC+8; future until effective	
DO · DO · outside weekday 01:00-04:00 / 06:00-10:00 UTC	DP · DP · weekday 01:00-04:00 / 06:00-10:00 UTC	GF · GF · introductory through 2026-12-31	NA · NA · local artifact only	
Qwen: implicit cache; never stack explicit cache.	RMB rows unbanded until a dated company FX scenario.	PUBLIC PERFORMANCE N/S	S048 is local-only; hosted SKU/tariff N/S at registry level.	
Promotional pricing available at least through 2026-11-21.		Exact promotion end N/S; exact reversion price N/S.		
Post-promotion cost N/S until fresh verification and a signed post-promotion bound/contract.		Signed post-promotion bound/contract: _____ owner/date: _		

TAKEAWAY Use the reference attempt for tariff comparison; calculate cost/AUO only from observed accepted outputs with a nonzero AUO denominator.	DECISION RULE Compare dated tariff scenarios per reference attempt, then decide from private accepted outcomes, all-attempt cost, and cost/AUO.
Status Tariff facts are C-grade; public performance and company outcomes remain N/S.	BOUNDARY The 8K/2K/2K bundle is not an AUO. Public performance and company outcomes remain N/S; post-promotion Sol pricing requires a contract scenario.

Show all four cost views before ROI

Separate cash from economic cost and 2026 activity from committed obligations. Reconcile every view to one workload and counterfactual.

Full-cost and hybrid routing ledger

Component	2026 cash	2026 economic	Committed cash	Committed economic
Model/API/seat	Paid in-window invoices; public list tariff only as assumption	In-window use plus economic entitlement treatment	Non-cancellable and post-2026 payments	Committed cash plus opportunity/terminal treatment
Local infrastructure	In-window purchase/lease cash	Depreciation/capital charge; defensible residual only here	Committed purchases, leases, reserved capacity	Committed cash plus defensible terminal/residual treatment
People/operations	Paid labor and contractors in-window	Fully loaded labor and opportunity cost	Committed support, staffing, on-call	Committed labor plus opportunity cost
Attempts/review	All eligible attempts, retries, evaluation, fallback, review	Same plus downstream correction burden	Committed quality/review obligations	Committed burden plus economic correction risk
Control plane	Application gateway/scaffold and enabled module cash	Ownership and operating burden	Committed licenses and integration	Committed controls plus opportunity cost
Risk/resilience	Incident, rollback, warm fallback cash	Expected downside and insurance	Committed standby and continuity	Committed downside plus terminal/exit treatment

All four outputs use the same signed workload, period, AUO definition, allocation driver, and named incumbent/counterfactual.

C_HYBRID = fixed local + local variable + frontier API/seat + application routing/control + enabled capital-H modules + duplicate/shadow/evaluation traffic + warm fallback + cross-environment network/data controls + multi-system operations/incident cost.

Use raw-direction tails: cost and burden adverse case = P90; value, NPV, and quality adverse case = P10; median = P50. A common P10/P90 label is permitted only after every input is transformed to a signed higher-is-better utility.

SCALE

All gates PASS; signed cost/AUO and adverse-case economics clear.

CAPPED PILOT-EXTEND-TIE

CAPPED PILOT-EXTEND-TIE only when zero mandatory controls Fail, every non-waivable field is Pass/known, remaining Unknowns are explicitly risk-owned bounded runtime/quality/capacity/resilience questions, and the pilot is segregated and non-sensitive.

STOP-REDESIGN

Any gate, floor, capacity, or downside limit fails.

A3 IMPAKT_LLM_Decision_Workbook_2026_v1.json / payload.economics_ledger_templates: comparator x component x four-view workbook; required before ROI · method M005/R001/R005 · manifest p30

policy/version ____ · dated cash-flow timing ____ · discount ____ · FX/currency ____ · tax ____

benefit basis/realization ____ · residual/terminal ____ · downside method ____

ROI numerator ____ · ROI denominator ____ · anti-double-count checks ____ · finance owner/sign/date ____

FINANCE-POLICY-v1

TAKEAWAY

One ledger must output four views and reconcile to cost per AUO, NPV, and ROI.

DECISION RULE

Approve no ROI from a blended cost number or unnamed counterfactual.

Four outputs

Report 2026 cash, 2026 economic, committed cash, and committed economic separately.

BOUNDARY

Blank finance-policy fields keep NPV, ROI, savings, and adverse-case economics N/S.

Innovation becomes a signed candidate result

Native anchors remain predeclared; each shortlisted system produces one comparable result row on the same confirmation cases.

Candidate innovation-result contract

Approved backlog addressed			approved item count + signed approved value		
Private-holdout uplift			AUO-rate percentage points + correction minutes		
Approval-cycle reduction			median calendar days per approved item		
Migration/regression avoided			engineer-hours + escaped regression count		
Replacement resilience			recovery days + AUO-quality loss		
CAND	CONFIG	NATIVE	INTERVAL	EVID SHA	U
CAND	CONFIG	NATIVE	INTERVAL	EVID SHA	U
CAND	CONFIG	NATIVE	INTERVAL	EVID SHA	U
CAND	CONFIG	NATIVE	INTERVAL	EVID SHA	U
CAND	CONFIG	NATIVE	INTERVAL	EVID SHA	U
CAND	CONFIG	NATIVE	INTERVAL	EVID SHA	U
CAND	CONFIG	NATIVE	INTERVAL	EVID SHA	U
CAND	CONFIG	NATIVE	INTERVAL	EVID SHA	U
Completed candidate IDs equal the signed shortlist exactly.		Signed monotone native-unit mapping; no cross-unit addition before normalization.		BCa case-cluster bootstrap, 95%; signed alternative requires a new comparison-design digest.	
Recompute every declared weight/anchor bound; unstable leader becomes TIE.		No signed material lead after uncertainty and sensitivity means TIE; incomplete means UNKNOWN.		FRONTIER-<WORKLOAD>-v12 and FIT-<WORKLOAD>-v12	
OWNER		FINANCE		SIGNATURE/DATE	
TAKEAWAY		DECISION RULE			
Bind every completed row to the signed frontier and the selected-posture fit; no roadmap claim can fill a private result.		Freeze weights, anchors, normalization, uncertainty, and tie rule before confirmation. Complete exactly the signed shortlist rows and no others.			
Operating note		BOUNDARY			
Bind every completed row to the signed frontier and the selected-posture fit; no roadmap claim can fill a private result.		Roadmaps remain D-grade scenarios. Blank, incomparable, or sensitivity-unstable results are UNKNOWN and block a material-lead claim.			

Local bytes are evidence, never a disposition

All 21 LOCAL identities retain source bytes in original units and exact normalized integer bytes; signed company constraints decide feasibility later.

Normalized LOCAL capacity and signed constraints

GB = 1,000,000,000 B		GiB = 1,073,741,824 B		TB = 1,000,000,000,000 B		TiB = 1,099,511,627,776 B	
PRECISION: BASE = base precision · M-BASE = mixed base · Q4_K_XL = dynamic Q4_K_XL							
PRECISION: MIXED = mixed precision · M-CKPT = mixed checkpoint							
MUSE-Q4KM-L	Q4_K_M	20 GB -> 20000000000 B	S049 / C	CFG-MUSE-Q4KM-L-v1	U		
MUSE-Q4KXL-L	Q4_K_XL	23 GB -> 23000000000 B	S049 / C	CFG-MUSE-Q4KXL-L-v1	U		
MUSE-BF16-L	BF16	59.6 GB -> 59600000000 B	S048 / E	CFG-MUSE-BF16-L-v1	U		
QW-27-FP8-L	FP8	31 GB -> 31000000000 B	S013 / C	CFG-QW-27-FP8-L-v1	U		
QW-27-BF16-L	BF16	N/S	S012 / N/S	CFG-QW-27-BF16-L-v1	U		
QW-FLASH-FP8-L	FP8	180 GB -> 180000000000 B	S017 / C	CFG-QW-FLASH-FP8-L-v1	U		
QW-FLASH-BASE-L	M-BASE	N/S	S016 / N/S	CFG-QW-FLASH-BASE-L-v1	U		
MM-M3-NVFP4-L	NVFP4	250 GB -> 250000000000 B	S045 / C	CFG-MM-M3-NVFP4-L-v1	U		
GLM-FLASH-MIXED-L	mixed FP8	N/S	S028 / N/S	CFG-GLM-FLASH-MIXED-L-v1	U		
MM-M3-MXFP8-L	MXFP8	440 GB -> 440000000000 B	S043 / C	CFG-MM-M3-MXFP8-L-v1	U		
GLM-FLASH-BF16-L	BF16	598.5 GiB -> 642634481664 B	S029 / C	CFG-GLM-FLASH-BF16-L-v1	U		
GLM-FULL-MIXED-L	MIXED	N/S	S026 / N/S	CFG-GLM-FULL-MIXED-L-v1	U		
HY4-FP8-L	FP8	0.8 TB -> 800000000000 B	S039 / C	CFG-HY4-FP8-L-v1	U		
MM-M3-BF16-L	BF16	854 GB -> 854000000000 B	S042 / E	CFG-MM-M3-BF16-L-v1	U		
DS-PRO-L	M-CKPT	893 GB -> 893000000000 B	S035 / C	CFG-DS-PRO-L-v1	U		
KIMI-K3-MXFP4-L	MXFP4	N/S	S051 / N/S	CFG-KIMI-K3-MXFP4-L-v12	U		
GLM-FULL-BF16-L	BF16	1.5 TB -> 1500000000000 B	S027 / C	CFG-GLM-FULL-BF16-L-v1	U		
HY4-BF16-L	BF16	1.56 TB -> 1560000000000 B	S038 / C	CFG-HY4-BF16-L-v1	U		
QW-MAX-FP8-L	FP8	2.4 TB -> 2400000000000 B	S015 / C	CFG-QW-MAX-FP8-L-v1	U		
QW-MAX-BASE-L	BASE	N/S	S014 / N/S	CFG-QW-MAX-BASE-L-v1	U		
DS-FLASH-MIXED-L	M-CKPT	N/S	S034 / N/S	CFG-DS-FLASH-MIXED-L-v1	U		
CONSTR ID				CONSTR SHA			
WORKLOAD				DATA			
CONFIG ID				CONFIG SHA			
artifact_bytes	LIMIT		OBS		SHA		U
peak_resident_memory	LIMIT		OBS		SHA		U
kv_cache	LIMIT		OBS		SHA		U
runtime_overhead	LIMIT		OBS		SHA		U
replicas	LIMIT		OBS		SHA		U
n_plus_one	LIMIT		OBS		SHA		U
interconnect	LIMIT		OBS		SHA		U
throughput	LIMIT		OBS		SHA		U
p95_latency	LIMIT		OBS		SHA		U
queue_depth	LIMIT		OBS		SHA		U
power	LIMIT		OBS		SHA		U
storage	LIMIT		OBS		SHA		U
PACKET ID				PACKET SHA			
OWNER				SIGNER			
SIGNATURE				DATE			
OVERALL	UNKNOWN	21 LOCAL · NORM · 0 DISP		HW/RUNTIME/EVID/TOPOLOGY U/N/S			
TAKEAWAY		DECISION RULE					
A capacity packet supplements the shared workload pilot. It cannot prove quality, fit, frontier position, or SCALE readiness.		Normalize through the versioned registry, bind one workload/data class/ configuration, then compute 12 typed capacity dimensions against signed constraints.					
Operating note		BOUNDARY					
A capacity packet supplements the shared workload pilot. It cannot prove quality, fit, frontier position, or SCALE readiness.		Every released disposition is UNKNOWN. KIMI-K3-MXFP4-L is MXFP4 only; exact revision/files/runtime and target hardware remain blank until signed.					

The CEO ledger starts empty

Break-even and realized benefits compare the incumbent with every signed eligible comparator; no topology is mandatory after a signed hard constraint.

Benefits, charges, and exact break-even boundary

Managed fully loaded cost avoided	N/S	Local fixed infrastructure	N/S
		Local variable energy/compute	N/S
Redeployable capacity or labor	N/S	People, review, operations	N/S
		Controls, evaluation, enabled modules	N/S
Risk-adjusted option value	N/S	Fallback, incident, migration, financing	N/S

$\text{Break_even_AUO_volume} = (\text{Fixed_local_cost} - \text{Fixed_managed_cost}) / (\text{Managed_variable_cost_per_AUO} - \text{Local_variable_cost_per_AUO})$

Denominator must be positive; both exact systems must meet equivalent gates and capacity must remain feasible.

$\text{Energy_cost} = \text{IT_energy_kWh} \times \text{PUE} \times \text{tariff}$

OR $\text{Energy_cost} = \text{Metered_facility_wall_energy_kWh} \times \text{tariff}$

Never use both boundaries or apply PUE twice; facility-wall energy already includes facility load.

ALL VALUES N/S METHOD M005/R001/R005 · TRACE P25

BENEFIT-REALIZATION-v1

Released time is cash/economic value only when actually redeployed to funded work, removes an avoidable cost, or creates measured contribution margin.

Unredeployed released time is capacity, not cash savings.

Field	Baseline	Candidate	Realized evidence/owner
task time + review/rework	_____	_____	_____
paid / eligible / active / AUO-producing	_____	_____	_____
released hours + realization fraction	_____	_____	_____
funded redeployment / avoidable cost / contribution margin	_____	_____	_____
Count each seat, subscription, labor hour, retry, review minute, benefit, and shared control allocation once.			

BENEFIT-REALIZATION-v1

FINANCE-POLICY-v1

Comparator set = incumbent UNION every signed eligible exact system; no managed/local quota and no excluded topology.

TAKEAWAY Use an incumbent plus all signed eligible comparators; count realized benefit only after funded redeployment, avoidable cost, or measured contribution margin.	DECISION RULE Compute equivalent-gate economics for the incumbent and every signed eligible comparator, then apply the one-count and benefit-realization policies.
Operating note Use an incumbent plus all signed eligible comparators; count realized benefit only after funded redeployment, avoidable cost, or measured contribution margin.	BOUNDARY Do not force managed or LOCAL after a signed hard-constraint exclusion. Ineligible systems stay out; missing comparable evidence remains UNKNOWN.

A seat option is a tested system, not a price row

Seven seat plans now have exact configuration, sample, protocol, gate, route, and decision templates for each workload.

Decision-capable seat option path

SEAT-CHATGPT-BUSINESS		ChatGPT Business	USD 20/user/month annual or USD 25 monthly; executed cadence/terms N/S	C/S/P=S01	U
SEAT-CLAUDE-TEAM-STANDARD		Claude Team standard	USD 20/seat/month annual or USD 25 monthly; executed cadence/terms N/S	C/S/P=S02	U
SEAT-CLAUDE-TEAM-PREMIUM		Claude Team premium	USD 100/seat/month annual or USD 125 monthly; executed cadence/terms N/S	C/S/P=S03	U
SEAT-CLAUDE-ENTERPRISE		Claude Enterprise	USD 20/seat/month plus API-rate usage, annual; executed terms N/S	C/S/P=S04	U
SEAT-MINIMAX-PLUS		MiniMax Token Plan Plus	USD 22/subscription/month; exact shared quota/terms N/S	C/S/P=S05	U
SEAT-MINIMAX-MAX		MiniMax Token Plan Max	USD 55/subscription/month; exact shared quota/terms N/S	C/S/P=S06	U
SEAT-MINIMAX-ULTRA		MiniMax Token Plan Ultra	USD 132/subscription/month; exact shared quota/terms N/S	C/S/P=S07	U
eligible users		assigned seats	active seats	completed cases	
entitlement-block events		paid interactions	accepted useful outputs		
ROW Sxx: SEAT_ID[Sxx] is the full seat candidate ID printed at left. CFG[Sxx] = 'CFG-I-' + SEAT_ID[Sxx] + '-v12'. P-C[Sxx] = 'PROTOCOL-' + SEAT_ID[Sxx] + '-CODING-v12'. P-M[Sxx] = 'PROTOCOL-' + SEAT_ID[Sxx] + '-MARKETING-v12'. P-I[Sxx] = 'PROTOCOL-' + SEAT_ID[Sxx] + '-DATA-INGESTION-v12'. C/S/P=Sxx binds the row's CFG plus all three protocol identities above.					
Count each seat/subscription once; keep seat fees, API usage, labor, retries, review, and controls in separate ledger fields before reconciliation.					
Seat entitlement never substitutes for API contract, API tariffs, or workload evidence; hybrid allocations reconcile each exact target separately.					
TAKEAWAY		DECISION RULE			
Measure eligibility, assigned/active seats, completed cases, entitlement blocks, paid interactions, and AUOs without mixing API spend.		A seat advances only through the same signed shortlist, shared confirmation frame, common gates, frontier, fit, route, rollback, and decision chain.			
Operating note		BOUNDARY			
Measure eligibility, assigned/active seats, completed cases, entitlement blocks, paid interactions, and AUOs without mixing API spend.		Published plan prices are tariff facts only. Executed plan/version/client/contract, adoption, capacity, and company outcomes remain N/S or UNKNOWN.			

Implementation burden becomes a candidate result

Eight native measures now produce one signed, uncertainty-aware candidate row on the same workload frame; time and labor remain separate until normalization.

Candidate implementation-result contract

integration labor	person-hours	BASE	CAND	INTERVAL	U
evaluation labor	person-hours	BASE	CAND	INTERVAL	U
migration burden	person-hours + calendar days	BASE	CAND	INTERVAL	U
incident recovery	minutes/event	BASE	CAND	INTERVAL	U
review/rework	minutes/AUO	BASE	CAND	INTERVAL	U
operational toil	person-hours/ month	BASE	CAND	INTERVAL	U
service tail	p95/p99 milliseconds	BASE	CAND	INTERVAL	U
capacity headroom	percentage points	BASE	CAND	INTERVAL	U
RESULT ID		CAND/CFG		SAMPLE/PROT	
RESULT+CI		SENS		DIGEST	
signed monotone per- metric mapping to 0- 100; favorable direction explicit	selected-posture weights total 100%; gates remain non- compensatory	case-clustered 95% interval propagated through normalization	recompute declared bounds; unstable leader becomes TIE	bind candidate row into FRONTIER- <WORKLOAD>-v12 and FIT- <WORKLOAD>-v12	
TAKEAWAY		DECISION RULE			
The result feeds the comparative frontier and selected-posture fit, not an unsupported demo-speed score.		Freeze baselines, mappings, materiality, uncertainty, and sensitivity before confirmation; compute one result for every signed shortlisted candidate.			
Operating note		BOUNDARY			
The result feeds the comparative frontier and selected-posture fit, not an unsupported demo-speed score.		Never combine heterogeneous units before signed normalization. Missing measures, unstable sensitivity, or incomparable baselines yield UNKNOWN or TIE.			

Coding compares every exact system on one core

Build/test quality, unsafe-change, service, and AUO economics are mandatory for all 45 systems; candidate falsifiers are additive.

Coding · 45 exact operating paths

CODING	PATH · EXACT SYSTEM/CONFIG · SAMPLE · PROTOCOL · Q SAFE SVC AUO · ADD/AGG
A01/A	OAI-SOL-A C/S=A01 P=A01 U U U U 1 U
A02/A	OAI-TERRA-A C/S=A02 P=A02 U U U U 1 U
A03/A	OAI-LUNA-A C/S=A03 P=A03 U U U U 1 U
A04/A	ANT-FABLE-A C/S=A04 P=A04 U U U U 3 U
A05/A	ANT-OPUS-A C/S=A05 P=A05 U U U U 1 U
A06/A	ANT-SONNET-A C/S=A06 P=A06 U U U U 1 U
A07/A	HY4-A C/S=A07 P=A07 U U U U 2 U
A08/A	MM-M3-A C/S=A08 P=A08 U U U U 3 U
A09/A	QW-27-A C/S=A09 P=A09 U U U U 1 U
A10/A	QW-MAX-A C/S=A10 P=A10 U U U U 1 U
A11/A	QW-FLASH-A C/S=A11 P=A11 U U U U 2 U
A12/A	GLM-FULL-A C/S=A12 P=A12 U U U U 3 U
A13/A	GLM-FLASH-A C/S=A13 P=A13 U U U U 2 U
A14/A	DS-FLASH-A C/S=A14 P=A14 U U U U 3 U
A15/A	DS-PRO-A C/S=A15 P=A15 U U U U 3 U
A16/A	KIMI-K3-A C/S=A16 P=A16 U U U U 1 U
A17/A	GEM-FLASH-A C/S=A17 P=A17 U U U U 2 U
L18/L	HY4-BF16-L C/S=L18 P=L18 U U U U 1 U
L19/L	HY4-FP8-L C/S=L19 P=L19 U U U U 1 U
L20/L	MM-M3-BF16-L C/S=L20 P=L20 U U U U 1 U
L21/L	MM-M3-MXFP8-L C/S=L21 P=L21 U U U U 1 U
L22/L	MM-M3-NVFP4-L C/S=L22 P=L22 U U U U 1 U
L23/L	QW-27-BF16-L C/S=L23 P=L23 U U U U 1 U
L24/L	QW-27-FP8-L C/S=L24 P=L24 U U U U 1 U
L25/L	QW-MAX-BASE-L C/S=L25 P=L25 U U U U 1 U
L26/L	QW-MAX-FP8-L C/S=L26 P=L26 U U U U 1 U
L27/L	QW-FLASH-BASE-L C/S=L27 P=L27 U U U U 1 U
L28/L	QW-FLASH-FP8-L C/S=L28 P=L28 U U U U 1 U
L29/L	GLM-FULL-MIXED-L C/S=L29 P=L29 U U U U 1 U
L30/L	GLM-FULL-BF16-L C/S=L30 P=L30 U U U U 1 U
L31/L	GLM-FLASH-MIXED-L C/S=L31 P=L31 U U U U 1 U
L32/L	GLM-FLASH-BF16-L C/S=L32 P=L32 U U U U 1 U
L33/L	DS-FLASH-MIXED-L C/S=L33 P=L33 U U U U 1 U
L34/L	DS-PRO-L C/S=L34 P=L34 U U U U 1 U
L35/L	MUSE-BF16-L C/S=L35 P=L35 U U U U 1 U
L36/L	MUSE-Q4KM-L C/S=L36 P=L36 U U U U 1 U
L37/L	MUSE-Q4KXL-L C/S=L37 P=L37 U U U U 1 U
L38/L	KIMI-K3-MXFP4-L C/S=L38 P=L38 U U U U 1 U
S39/S	SEAT-CHATGPT-BUSINESS C/S=S39 P=S39 U U U U 1 U
S40/S	SEAT-CLAUDE-TEAM-STANDARD C/S=S40 P=S40 U U U U 1 U
S41/S	SEAT-CLAUDE-TEAM-PREMIUM C/S=S41 P=S41 U U U U 1 U
S42/S	SEAT-CLAUDE-ENTERPRISE C/S=S42 P=S42 U U U U 1 U
S43/S	SEAT-MINIMAX-PLUS C/S=S43 P=S43 U U U U 1 U
S44/S	SEAT-MINIMAX-MAX C/S=S44 P=S44 U U U U 1 U
S45/S	SEAT-MINIMAX-ULTRA C/S=S45 P=S45 U U U U 1 U

FILE b0ad4300 · EXPECT 0c06eb94 · ACTUAL 2c313e3a · CANON 1c4caee8

GATES: Q QUALITY · SAFE SAFETY · SVC SERVICE · AUO ECON · ADD ADDITIVE · U UNKNOWN

CFG[Axx] = 'CFG-I' + CID[Axx] + ('-v12' if TOPO[Axx]=SEAT or CID[Axx]=KIMI-K3-MXFP4-L else '-v1').

S[Axx] = 'SAMPLE-' + CID[Axx] + '-CODING-v12'.

P[Axx] = 'PROTOCOL-' + CID[Axx] + '-CODING-v12'.

A1 expected_gate_catalog_v12.json / actual_gate_set_templates · CODE · U ONLY->exact CAND+CFG+S+P · p30

TAKEAWAY Use the same frozen cases, protocol, four core definitions, typed thresholds, units, denominators, and evidence rules for every shortlisted exact system.	DECISION RULE Expected gates equal the common core plus additive falsifiers and applicable crosscuts/topology/risk gates; actual IDs must match exactly.
Operating note Use the same frozen cases, protocol, four core definitions, typed thresholds, units, denominators, and evidence rules for every shortlisted exact system.	BOUNDARY Released rows are UNKNOWN. LOCAL capacity, seat entitlement, public benchmarks, and additive falsifiers cannot replace any common core result.

Marketing compares every exact system on one core

Brand/compliance quality, unsafe publication, service, and AUO economics are mandatory for all 45 systems; candidate falsifiers are additive.

Marketing · 45 exact operating paths

MARKETING	PATH · EXACT SYSTEM/CONFIG	SAMPLE · PROTOCOL	Q	SAFE	SVC	AUO	ADD/AGG		
A01/A	OAI-SOL-A	C/S=A01	P=A01	U	U	U	U	1	U
A02/A	OAI-TERRA-A	C/S=A02	P=A02	U	U	U	U	1	U
A03/A	OAI-LUNA-A	C/S=A03	P=A03	U	U	U	U	1	U
A04/A	ANT-FABLE-A	C/S=A04	P=A04	U	U	U	U	3	U
A05/A	ANT-OPUS-A	C/S=A05	P=A05	U	U	U	U	1	U
A06/A	ANT-SONNET-A	C/S=A06	P=A06	U	U	U	U	2	U
A07/A	HY4-A	C/S=A07	P=A07	U	U	U	U	1	U
A08/A	MM-M3-A	C/S=A08	P=A08	U	U	U	U	2	U
A09/A	QW-27-A	C/S=A09	P=A09	U	U	U	U	1	U
A10/A	QW-MAX-A	C/S=A10	P=A10	U	U	U	U	1	U
A11/A	QW-FLASH-A	C/S=A11	P=A11	U	U	U	U	1	U
A12/A	GLM-FULL-A	C/S=A12	P=A12	U	U	U	U	1	U
A13/A	GLM-FLASH-A	C/S=A13	P=A13	U	U	U	U	3	U
A14/A	DS-FLASH-A	C/S=A14	P=A14	U	U	U	U	2	U
A15/A	DS-PRO-A	C/S=A15	P=A15	U	U	U	U	1	U
A16/A	KIMI-K3-A	C/S=A16	P=A16	U	U	U	U	3	U
A17/A	GEM-FLASH-A	C/S=A17	P=A17	U	U	U	U	2	U
L18/L	HY4-BF16-L	C/S=L18	P=L18	U	U	U	U	1	U
L19/L	HY4-FP8-L	C/S=L19	P=L19	U	U	U	U	1	U
L20/L	MM-M3-BF16-L	C/S=L20	P=L20	U	U	U	U	1	U
L21/L	MM-M3-MXFP8-L	C/S=L21	P=L21	U	U	U	U	1	U
L22/L	MM-M3-NVFP4-L	C/S=L22	P=L22	U	U	U	U	1	U
L23/L	QW-27-BF16-L	C/S=L23	P=L23	U	U	U	U	1	U
L24/L	QW-27-FP8-L	C/S=L24	P=L24	U	U	U	U	1	U
L25/L	QW-MAX-BASE-L	C/S=L25	P=L25	U	U	U	U	1	U
L26/L	QW-MAX-FP8-L	C/S=L26	P=L26	U	U	U	U	1	U
L27/L	QW-FLASH-BASE-L	C/S=L27	P=L27	U	U	U	U	1	U
L28/L	QW-FLASH-FP8-L	C/S=L28	P=L28	U	U	U	U	1	U
L29/L	GLM-FULL-MIXED-L	C/S=L29	P=L29	U	U	U	U	1	U
L30/L	GLM-FULL-BF16-L	C/S=L30	P=L30	U	U	U	U	1	U
L31/L	GLM-FLASH-MIXED-L	C/S=L31	P=L31	U	U	U	U	1	U
L32/L	GLM-FLASH-BF16-L	C/S=L32	P=L32	U	U	U	U	1	U
L33/L	DS-FLASH-MIXED-L	C/S=L33	P=L33	U	U	U	U	1	U
L34/L	DS-PRO-L	C/S=L34	P=L34	U	U	U	U	1	U
L35/L	MUSE-BF16-L	C/S=L35	P=L35	U	U	U	U	1	U
L36/L	MUSE-Q4KM-L	C/S=L36	P=L36	U	U	U	U	1	U
L37/L	MUSE-Q4KXL-L	C/S=L37	P=L37	U	U	U	U	1	U
L38/L	KIMI-K3-MXFP4-L	C/S=L38	P=L38	U	U	U	U	1	U
S39/S	SEAT-CHATGPT-BUSINESS	C/S=S39	P=S39	U	U	U	U	1	U
S40/S	SEAT-CLAUDE-TEAM-STANDARD	C/S=S40	P=S40	U	U	U	U	1	U
S41/S	SEAT-CLAUDE-TEAM-PREMIUM	C/S=S41	P=S41	U	U	U	U	1	U
S42/S	SEAT-CLAUDE-ENTERPRISE	C/S=S42	P=S42	U	U	U	U	1	U
S43/S	SEAT-MINIMAX-PLUS	C/S=S43	P=S43	U	U	U	U	1	U
S44/S	SEAT-MINIMAX-MAX	C/S=S44	P=S44	U	U	U	U	1	U
S45/S	SEAT-MINIMAX-ULTRA	C/S=S45	P=S45	U	U	U	U	1	U

FILE b0ad4300 · EXPECT 0c06eb94 · ACTUAL 2c313e3a · CANON 1c4caee8

GATES: Q QUALITY · SAFE SAFETY · SVC SERVICE · AUO ECON · ADD ADDITIVE · U UNKNOWN

CFG[Axx] = 'CFG-I' + CID[Axx] + ('-v12' if TOPO[Axx]=SEAT or CID[Axx]=KIMI-K3-MXFP4-L else '-v1').

S[Axx] = 'SAMPLE-' + CID[Axx] + '-MARKETING-v12'.

P[Axx] = 'PROTOCOL-' + CID[Axx] + '-MARKETING-v12'.

A1 expected_gate_catalog_v12.json / actual_gate_set_templates · MKT · U ONLY->exact CAND+CFG+S+P · p30

TAKEAWAY

Use the same frozen cases, protocol, four core definitions, typed thresholds, units, denominators, and evidence rules for every shortlisted exact system.

DECISION RULE

Expected gates equal the common core plus additive falsifiers and applicable crosscuts/topology/risk gates; actual IDs must match exactly.

Operating note

Use the same frozen cases, protocol, four core definitions, typed thresholds, units, denominators, and evidence rules for every shortlisted exact system.

BOUNDARY

Released rows are UNKNOWN. LOCAL capacity, seat entitlement, public benchmarks, and additive falsifiers cannot replace any common core result.

Ingestion compares every exact system on one core

Loss/repair quality, safety/privacy, service, and AUO economics are mandatory for all 45 systems; candidate falsifiers are additive.

Data Ingestion · 45 exact operating paths

INGEST	PATH · EXACT SYSTEM/CONFIG · SAMPLE · PROTOCOL · Q	SAFE	SVC	AUO	ADD/AGG
A01/A	OAI-SOL-A	C/S=A01	P=A01	U	U
A02/A	OAI-TERRA-A	C/S=A02	P=A02	U	U
A03/A	OAI-LUNA-A	C/S=A03	P=A03	U	U
A04/A	ANT-FABLE-A	C/S=A04	P=A04	U	U
A05/A	ANT-OPUS-A	C/S=A05	P=A05	U	U
A06/A	ANT-SONNET-A	C/S=A06	P=A06	U	U
A07/A	HY4-A	C/S=A07	P=A07	U	U
A08/A	MM-M3-A	C/S=A08	P=A08	U	U
A09/A	QW-27-A	C/S=A09	P=A09	U	U
A10/A	QW-MAX-A	C/S=A10	P=A10	U	U
A11/A	QW-FLASH-A	C/S=A11	P=A11	U	U
A12/A	GLM-FULL-A	C/S=A12	P=A12	U	U
A13/A	GLM-FLASH-A	C/S=A13	P=A13	U	U
A14/A	DS-FLASH-A	C/S=A14	P=A14	U	U
A15/A	DS-PRO-A	C/S=A15	P=A15	U	U
A16/A	KIMI-K3-A	C/S=A16	P=A16	U	U
A17/A	GEM-FLASH-A	C/S=A17	P=A17	U	U
L18/L	HY4-BF16-L	C/S=L18	P=L18	U	U
L19/L	HY4-FP8-L	C/S=L19	P=L19	U	U
L20/L	MM-M3-BF16-L	C/S=L20	P=L20	U	U
L21/L	MM-M3-MXFP8-L	C/S=L21	P=L21	U	U
L22/L	MM-M3-NVFP4-L	C/S=L22	P=L22	U	U
L23/L	QW-27-BF16-L	C/S=L23	P=L23	U	U
L24/L	QW-27-FP8-L	C/S=L24	P=L24	U	U
L25/L	QW-MAX-BASE-L	C/S=L25	P=L25	U	U
L26/L	QW-MAX-FP8-L	C/S=L26	P=L26	U	U
L27/L	QW-FLASH-BASE-L	C/S=L27	P=L27	U	U
L28/L	QW-FLASH-FP8-L	C/S=L28	P=L28	U	U
L29/L	GLM-FULL-MIXED-L	C/S=L29	P=L29	U	U
L30/L	GLM-FULL-BF16-L	C/S=L30	P=L30	U	U
L31/L	GLM-FLASH-MIXED-L	C/S=L31	P=L31	U	U
L32/L	GLM-FLASH-BF16-L	C/S=L32	P=L32	U	U
L33/L	DS-FLASH-MIXED-L	C/S=L33	P=L33	U	U
L34/L	DS-PRO-L	C/S=L34	P=L34	U	U
L35/L	MUSE-BF16-L	C/S=L35	P=L35	U	U
L36/L	MUSE-Q4KM-L	C/S=L36	P=L36	U	U
L37/L	MUSE-Q4KXL-L	C/S=L37	P=L37	U	U
L38/L	KIMI-K3-MXFP4-L	C/S=L38	P=L38	U	U
S39/S	SEAT-CHATGPT-BUSINESS	C/S=S39	P=S39	U	U
S40/S	SEAT-CLAUDE-TEAM-STANDARD	C/S=S40	P=S40	U	U
S41/S	SEAT-CLAUDE-TEAM-PREMIUM	C/S=S41	P=S41	U	U
S42/S	SEAT-CLAUDE-ENTERPRISE	C/S=S42	P=S42	U	U
S43/S	SEAT-MINIMAX-PLUS	C/S=S43	P=S43	U	U
S44/S	SEAT-MINIMAX-MAX	C/S=S44	P=S44	U	U
S45/S	SEAT-MINIMAX-ULTRA	C/S=S45	P=S45	U	U
FILE b0ad4300 · EXPECT 0c06eb94 · ACTUAL 2c313e3a · CANON 1c4caee8					
GATES: Q QUALITY · SAFE SAFETY · SVC SERVICE · AUO ECON · ADD ADDITIVE · U UNKNOWN					
CFG[Axx] = 'CFG-I' + CID[Axx] + ('-v12' if TOPO[Axx]=SEAT or CID[Axx]=KIMI-K3-MXFP4-L else '-v1').					
S[Axx] = 'SAMPLE-' + CID[Axx] + '-DATA-INGESTION-v12'.					
P[Axx] = 'PROTOCOL-' + CID[Axx] + '-DATA-INGESTION-v12'.					
A1 expected_gate_catalog_v12.json / actual_gate_set_templates · INGEST · U ONLY->exact CAND+CFG+S+P · p30					
TAKEAWAY		DECISION RULE			
Use the same frozen cases, protocol, four core definitions, typed thresholds, units, denominators, and evidence rules for every shortlisted exact system.		Expected gates equal the common core plus additive falsifiers and applicable crosscuts/topology/risk gates; actual IDs must match exactly.			
Operating note		BOUNDARY			
Use the same frozen cases, protocol, four core definitions, typed thresholds, units, denominators, and evidence rules for every shortlisted exact system.		Released rows are UNKNOWN. LOCAL capacity, seat entitlement, public benchmarks, and additive falsifiers cannot replace any common core result.			

Choose a budgeted D/P/R/A design before claiming effects

DIRECT_ONLY, FRACTIONAL_SCREEN, TARGETED_CONFIRMATORY, and FULL_FACTORIAL make burden and permissible claims explicit across the same sixteen-cell registry.

Executable budgeted D/P/R/A result

D · direct exact system plus generic scaffold			P · capital-H pass-through modules and policies			R · application-owned routing policy			A · alternate exact target		
RESULT	CAND/CFG	WORK/DATA	SHORT SHA	SAMPLE SHA	BUDGET ID	BUDGET SHA					
DESIGN MODE ☐ DIRECT ☐ FRACTIONAL ☐ TARGETED ☐ FULL											
D0P0R0A0	A:U	U	D0P0R0A1	A:U	U	D0P0R1A0	A:U	U	D0P0R1A1	A:U	U
CFG			CFG			CFG			CFG		
CxA			CxA			CxA			CxA		
THR			THR			THR			THR		
W/SHA			W/SHA			W/SHA			W/SHA		
D0P1R0A0	A:U	U	D0P1R0A1	A:U	U	D0P1R1A0	A:U	U	D0P1R1A1	A:U	U
CFG			CFG			CFG			CFG		
CxA			CxA			CxA			CxA		
THR			THR			THR			THR		
W/SHA			W/SHA			W/SHA			W/SHA		
D1P0R0A0	A:U	U	D1P0R0A1	A:U	U	D1P0R1A0	A:U	U	D1P0R1A1	A:U	U
CFG			CFG			CFG			CFG		
CxA			CxA			CxA			CxA		
THR			THR			THR			THR		
W/SHA			W/SHA			W/SHA			W/SHA		
D1P1R0A0	A:U	U	D1P1R0A1	A:U	U	D1P1R1A0	A:U	U	D1P1R1A1	A:U	U
CFG			CFG			CFG			CFG		
CxA			CxA			CxA			CxA		
THR			THR			THR			THR		
W/SHA			W/SHA			W/SHA			W/SHA		
MODE SHA d2d3d1b10d6ee99dba6e12791a94a2dbb8c01bdb32f21c67ebc7c609169e1715						SELECTED FACTORIAL RESULT ID + SHA					
MODE DEFINES APPLICABLE CELLS · OTHER CELLS N/A = 0						APPLICABLE: CASES x ATTEMPTS = EXECUTIONS · REUSE ONCE					
7 STAGES / 7 DIMENSIONS = LEDGER TOTAL <= EACH CAP						DIRECT:NO EFFECT · FRACTIONAL:ALIASED · TARGETED:SIGNED · FULL:DECLARED					
OWNER			SIGNER			SIGNATURE/DATE			DIGEST		
TAKEAWAY The signed mode makes cells APPLICABLE or N/A, reconciles exact cases and attempts with the all-in ledger, and limits every estimand before execution.						DECISION RULE DIRECT_ONLY uses only D1P0R0A0 and claims no D/P/R/A effect. Other modes must match their exact signed cell set, arithmetic, evidence, and claim ceiling.					
Operating note The signed mode makes cells APPLICABLE or N/A, reconciles exact cases and attempts with the all-in ledger, and limits every estimand before execution.						BOUNDARY Generic scaffold is application-owned; capital-H Harness scope is module-specific; application routing owns target selection. All released effects are N/S/ UNKNOWN.					

Residency is the whole route

Residency covers every real ingress, work, record, control, and fallback path. A regional endpoint or local server is insufficient.

Data-class control matrix

Class	Examples	Minimum route controls	Default deployment rule
D0 Public	Already public, non-sensitive	Identity, integrity, abuse, logging	Normal security baseline
D1 Internal	Routine internal drafts	Org identity, encryption, approved terms, deletion	Approved managed or local route
D2 Confidential	Customer/company confidential, source code	DPA, use/retention, region, RBAC, audit, incident	Exact endpoint/build approval
D3 Restricted/regulated	PHI, payment, sensitive personal, regulated decisions	Counsel/security approval, eligible features, keys/network, SIEM, isolation, human oversight	Strict dedicated route
D4 Sovereign/critical secrets	Export-controlled, classified, crown-jewel	No uncontrolled external path; physical, personnel, supply-chain, no-egress controls	Explicit threat-model approval
ALL ROUTES N/S		Risk/route method: M001, M002, R001, R002, R004. Full trace: page 25.	ROUTE STAGE COUNT = 5
INGRESS: client > application gateway/generic scaffold > queue			
CAPITAL-H SERVICES: each enabled service receiving metadata, traces, artifacts, or control traffic			
WORK: model > tools > retrieval			
RECORDS: cache > logging > support > backup			
FALLBACK: every fallback destination and control path			
MANAGED CHECK: Plan, endpoint, region, use/retention, subprocessors, keys/network, logs, incident, continuity.			
LOCAL CHECK: License, integrity/SBOM, vulnerabilities, telemetry/egress, secrets, storage/backups, admin/physical isolation, recovery.			
PASS only when every mandatory and failover path has approved geography, use, retention, deletion, authority, and fallback.		REQUIRED: geography / use / retention / deletion / authority / fallback	
NAMED AUTHORITY		POLICY + VERSION	OWNER / SIGNATURE / DATE
MIXED / UNCERTAIN: Use the most restrictive applicable class or STOP-REDESIGN.			
NO SILENT LOWERING: Never lower class across tools, outputs, telemetry, records, backups, support, or fallback.			
CATALOG ID: RISK-CONTROL-CATALOG-v6			
CATALOG SHA: a4f25e84fc5049695ec75e1716aa5519853f3e1f40b66b1d35c0002ae0eb6c1d			
ACTIVATION: Any trigger applies the full Huge/risk gate set.			
TAKEAWAY		DECISION RULE	
Any unknown mandatory or failover path fails production.		One named authority signs every workload/data-class route.	
Endpoint limit		BOUNDARY	
Managed and local checks stay attached to the same end-to-end route and its actual data/control traffic.		Mixed or uncertain data takes the most restrictive class or stops.	

Sign the frontier; score only the selected posture

At most six shortlist rows compare native evidence and uncertainty. One selected company posture supplies weights; the other three remain irrelevant.

Comparative frontier + one selected-posture fit

SHORTLIST ID		SHORTLIST SHA		DESIGN ID	
DESIGN SHA		FRONTIER ID		FRONTIER SHA	
CAND	CONFIG	QUALITY+CI	COST/AUO+CI	TAIL+CI	U
CAND	CONFIG	QUALITY+CI	COST/AUO+CI	TAIL+CI	U
CAND	CONFIG	QUALITY+CI	COST/AUO+CI	TAIL+CI	U
CAND	CONFIG	QUALITY+CI	COST/AUO+CI	TAIL+CI	U
CAND	CONFIG	QUALITY+CI	COST/AUO+CI	TAIL+CI	U
CAND	CONFIG	QUALITY+CI	COST/AUO+CI	TAIL+CI	U
CAND	CONFIG	QUALITY+CI	COST/AUO+CI	TAIL+CI	U
CAND	CONFIG	QUALITY+CI	COST/AUO+CI	TAIL+CI	U

POSTURE
☐ Small
☐ Medium
☐ Big
☐ Huge

WEIGHTS=100	NORM SHA	95% METHOD	SENSITIVITY	FIT ID

RESULT	☐ MATERIAL LEADER ☐ TIE ☐ UNKNOWN	FIT SHA	
--------	---	---------	--

shortlist COMPLETE	frontier COMPLETE	selected-posture fit COMPLETE	material lead or signed tie disposition	all non-compensatory gates PASS
--------------------	-------------------	-------------------------------	---	---------------------------------

OWNER	SIGNER	SIGNATURE/DATE

A2 shortlist_frontier_fit_v12.json / fit_results.
A3 IMPAKT_LLM_Decision_Workbook_2026_v1.json / payload.fit_execution_schema_v1.
Eight-dimension mapping + direction + weight + sensitivity + posture-rationale + result-tie trace.
Required before MATERIAL LEADER or TIE; manifest, SHA-256, bytes, and extraction instruction: p30.

TAKEAWAY A material leader must survive uncertainty and signed sensitivity. Otherwise record TIE; incomplete comparison remains UNKNOWN.	DECISION RULE Complete frontier rows exactly equal the signed shortlisted candidates. Normalize native dimensions only after mappings/weights are frozen and signed.
Operating note A material leader must survive uncertainty and signed sensitivity. Otherwise record TIE; incomplete comparison remains UNKNOWN.	BOUNDARY Gates never compensate. Public benchmarks, byte bands, or plan prices cannot fill company result cells; SCALE requires COMPLETE frontier and fit digests.

Sign one exact route for one workload decision

One workload x data class x exact system binds shortlist, sample, protocol, context, route, active register, rollback, and every used target.

Exact-system workload route form

WORK	DATA	CFG ID	CFG SHA	SHORT ID	SHORT SHA	SAMPLE ID
SAMPLE SHA	PROT ID	PROT SHA	GATE ID	GATE SHA	BUDGET ID	BUDGET SHA
CTX ID	CTX SHA	ROUTE ID	ROUTE SHA	SET ID		
SET SHA	REGISTER ID	REGISTER ATTEST	ROLLBACK ID	ROLLBACK SHA		
PRIMA RY	☐ USED ☐ UNUS ED	TARGET ID	TARGET SHA	EXECUTION	SHARE	ACCOUNTING
FALLB ACK	☐ USED ☐ UNUS ED	TARGET ID	TARGET SHA	EXECUTION	SHARE	ACCOUNTING
RETRY	☐ USED ☐ UNUS ED	TARGET ID	TARGET SHA	EXECUTION	SHARE	ACCOUNTING
EVALU ATION	☐ USED ☐ UNUS ED	TARGET ID	TARGET SHA	EXECUTION	SHARE	ACCOUNTING
SHADO W	☐ USED ☐ UNUS ED	TARGET ID	TARGET SHA	EXECUTION	SHARE	ACCOUNTING
5 EXCLUSIVE ROLES · PLANNED=100% · REALIZED=100% · RECONCILE ONCE						
SIGN<=ACTIVATE<=ATTEST<=DECIDE · COVER THROUGH 2026-12-28						
ONE WORKLOAD x DATA CLASS x EXACT SYSTEM · 5 ROLES						
TARGETS / SHARES / EVIDENCE / COST / REGISTER / DECISION = U/N/S						
OWNER		SIGNER		SIGNATURE/DATE		
TAKEAWAY		DECISION RULE				
Each used flow must equal a COMPLETE configuration digest; five mutually exclusive shares reconcile to 100% planned and realized.		Bind the exact context candidate/configuration to the primary target; alternatives may differ only through separately COMPLETE exact configurations and signed route roles.				
Operating note		BOUNDARY				
Each used flow must equal a COMPLETE configuration digest; five mutually exclusive shares reconcile to 100% planned and realized.		UNUSED roles carry a reason and no target. Capacity is supplemental for LOCAL; seat, API, LOCAL, and hybrid routes retain separate contracts and evidence identities.				

BLANK WORKLOAD DECISION CARD - NOT AN APPROVAL

Sixteen separately auditable identity pairs and three locked decision digests bind one workload, data class, exact system, and complete executive stack.

Auditable workload decision stack

WORKLOAD	DATA	POSTURE	DECISION AT
OBJECT · INDEPENDENT ID · INDEPENDENT DIGEST			
Configuration		Active register	
ID	SHA	ID	SHA
Signed shortlist		D/P/R/A result	
ID	SHA	ID	SHA
Sample manifest		Comparative frontier	
ID	SHA	ID	SHA
Pilot protocol		Selected-posture fit	
ID	SHA	ID	SHA
Capacity packet / N/A		Rollback plan	
ID	SHA	ID	SHA
Decision context		Actual gate set	
ID	SHA	ID	SHA
Route set		All-in budget ledger	
ID	SHA	ID	SHA
Member route(s)		Decision packet ID/state	
ID	SHA	ID	SHA
PAYLOAD SHA		ATTESTATION SHA	
☐ Executed contract evidence attached		☐ Executed control evidence attached	
Computed aggregate UNKNOWN		LOCKED UNKNOWN UNTIL EVERY REQUIRED ID/DIGEST/STATE RECOMPUTES	
DISPOSITION		☐ SCALE ☐ CAPPED PILOT-EXTEND-TIE ☐ STOP-REDESIGN	
BOUNDED CONDITION			
UNKNOWN / EXPIRY / EXTENSION			
CEO		FINANCE OWNER	
SIG	DATE	SIG	DATE
WORKLOAD OWNER		CONTROL OWNER	
SIG	DATE	SIG	DATE
TAKEAWAY		DECISION RULE	
The computed aggregate is UNKNOWN until every selected artifact is COMPLETE, current, exact, signed, and mechanically reconciled.		Recompute the stack: shortlist, all-in budget, sample/protocol, capacity when LOCAL, context/configuration, route/register, factorial, frontier, fit, rollback, actual gates, contract/control evidence, and decision attestation.	
Operating note		BOUNDARY	
The computed aggregate is UNKNOWN until every selected artifact is COMPLETE, current, exact, signed, and mechanically reconciled.		No released control is selected or signed. Any blank, stale digest, FAIL, prohibited waiver, non-A/B production-critical evidence, route mismatch, or expired window blocks SCALE.	

How to read the evidence register

Decode the register before using it: methodology, repository doctrine, chronology, and license exceptions are source-scoped.

Method, repository, and artifact anchors

A8	B8
M001 · NIST GenAI Profile AI 600-1 AP 4,9,18,21-22,24 · METHOD LOC nist.gov · D ret 2026-08-28 · G C V METHOD L not certification/law/counsel/system test · X=X-M A ledger:M001 · 6c49d89a	R004 · AI production-readiness gate AP 2,4,16,19,21-23 · METHOD LOC repo/readiness-gate · D WT 2026-08-28 · G C V METHOD L not assessment/red team/attestation/assurance · X=X-M A ledger:R004 · 6c49d89a
M002 · NIST adversarial ML taxonomy AP 9,18-19,21 · METHOD LOC nist.gov · D ret 2026-08-28 · G C V METHOD L framework is not certification or threat resistance · X X-M A ledger:M002 · 6c49d89a	R005 · Cost per completed workflow AP 2-3,5,7,9,11-12,14,17-18,22-24 · METHOD/ID LOC repo/cost-workflow · D WT 2026-08-28 · G C V IO+METHOD L no filled data, price table, saving, ROI, or payback · X X-M A ledger:R005 · 6c49d89a
M003 · MLPerf Inference Datacenter AP 3,13,16-19 · METHOD LOC mlcommons.org · D ret 2026-08-28 · G C V METHOD L public scenario ≠ company workload/config · X=X-M A ledger:M003 · 6c49d89a	R006 · Evaluation oracle release control AP 16,18-20,23 · METHOD LOC repo/oracle-release · D WT 2026-08-28 · G C V METHOD+TRACE L method is not production or independent evidence · X X-M A ledger:R006 · 6c49d89a
M004 · LM Evaluation Harness AP 3,9,17-18,20 · METHOD/BENCH-V LOC github.com · D ret 2026-08-28 · G C V BENCH-V+METHOD+TRACE L tool/benchmark ≠ company outcome; pin config · X=X-M A ledger:M004 · 6c49d89a	S024 · GLM-5.3 launch AP 8 · METHOD/ART/LIC/STAT/DATE LOC z.ai · D 2026-08-14 · G C V DATE+ARCH+BENCH-V+METHOD L vendor methods vary by task · X X-A/X-S A ledger:S024 · 6c49d89a
M005 · HELM evaluation method AP 3,5,11,14 · METHOD/BENCH-V LOC crfm.stanford.edu · D ret 2026-08-28 · G C V BENCH-V+METHOD L rank is screening, not workload/business outcome · X X-M A ledger:M005 · 6c49d89a	S025 · GLM-5.3-Flash launch AP 8,17 · METHOD/ID/STAT/DATE LOC z.ai · D 2026-08-26 · G C V DATE+METHOD+MOD+BENCH-V L vendor-run evaluation · X X-I/X-S A ledger:S025 · 6c49d89a
R001 · API, self-hosted, or hybrid AP 2-7,9,11-12,14,21-24 · METHOD LOC repo/api-self-hosted-or-hybrid · D WT 2026-08-28 · G C V METHOD L doctrine is not legal/security/procurement approval · X X-M A ledger:R001 · 6c49d89a	S037 · Tencent Hy4 repository AP 8 · METHOD/ID/ART/LIC LOC github.com · D 2026-08-28 · G C V ART+CTX+TOOLS+REAS+GUIDE+LIC L official repository only · X X-I/X-A A ledger:S037 · 6c49d89a
R002 · Five enterprise AI boundaries AP 6,20-21,23 · METHOD/BENCH-V LOC repo/five-boundaries · D WT 2026-08-28 · G C V BENCH-V+METHOD L framework is not deployed/effective control · X X-M A ledger:R002 · 6c49d89a	S044 · MiniMax community license AP 8 · METHOD/ART/LIC LOC huggingface.co · D 2026-08-28 · G C V LIC+TRACE+POLICY L not OSI-approved; legal review · X X-A A ledger:S044 · 6c49d89a
R003 · Private LLM benchmark method AP 2-3,15,19,23 · METHOD/BENCH-V LOC repo/private-benchmark · D WT 2026-08-28 · G C V BENCH-V+METHOD L template; no runs or results · X X-M A ledger:R003 · 6c49d89a	S056 · MiniMax M3 commits AP 8 · ART/LIC/STAT/DATE LOC huggingface.co · D init 2026-06-12 · G C V DATE+ART L weights chronology, not API launch · X X-A/X-S A ledger:S056 · 6c49d89a

AP=use/pages · LOC=link · D=date · G=grade · V=verifies · L=limit · X=transfer · A=archive
PASS=met FAIL=miss UNKNOWN=stop N/S=not scorable A=exact B=repro C=first-party D=adjacent E=assumption/nondecisive
ID=id STAT=status DATE=date ART=artifact CTX=context IO=I/O MOD=mode TOOLS=tool REAS=reason PRICE=tariff CUT=cutoff
PLAN=plan QUOTA=quota LIC=license BYTE=bytes ARCH=arch PARAM=params BENCH-V=bench RET=retain FB=fallback
REFUSAL=refuse COMPAT=compat METHOD=method TRACE=trace POLICY=control GUIDE=guide ENC=encoder LANG=lang
X-I=id#fit/ready X-A=artifact#runtime/result X-T=tariff#realized cost X-S=status#continuity/result X-M=method#control/result
Fresh verify + new archive before approval. Ledger ≠ source capture. Repository boundary: artifact ≠ runtime/outcome.

TAKEAWAY 16 catalog records; no public company outcome.	DECISION RULE Use codes only with the visible grade, transfer, applicability, freshness, and repository rules.
Decode Read AP/LOC/D/G/V/L/X/A, then claim and transfer codes before following any source.	BOUNDARY Public method and repository evidence cannot substitute for exact private approval evidence.

Plans, Harness, and operating sources

Plans, product quotas, capital-H modules, and status notices define commercial/control objects-not workload outcomes or API equivalence.

Plans, modules, and operating evidence

A	8	B	7
P001 · ChatGPT Work + Codex pricing AP 6,15 · METHOD/PRICE/PLAN LOC learn.chatgpt.com · D 2026-08-28 · G C V PRICE+PLAN+POLICY L shared task use; no API; Enterprise quoted · X X-T A ledger:P001 · 6c49d89a		H006 · Feature-policy checks AP 6,15-16,20 · METHOD LOC developer.harness.io · D 2026-08-28 · G C V METHOD+POLICY L not retroactive; route/threshold/rollback separate · X X-M A ledger:H006 · 6c49d89a	
P002 · Claude plans and pricing AP 6,15 · PRICE/PLAN LOC claude.com · D 2026-08-28 · G C V PRICE+PLAN+QUOTA L usage limits; API separate; contract-specific · X X-T A ledger:P002 · 6c49d89a		H007 · AWS MLOps tutorial AP 6,15-16,20 · METHOD LOC developer.harness.io · D 2026-08-28 · G C V POLICY+GUIDE L tutorial is not LLM/agent adequacy · X X-M A ledger:H007 · 6c49d89a	
P003 · MiniMax Token Plan AP 6,15 · PRICE/PLAN LOC platform.minimax.io · D 2026-08-28 · G C V PRICE+PLAN+QUOTA L not per-user or enterprise governance · X X-T A ledger:P003 · 6c49d89a		S006 · Claude Fable + Mythos launch AP 8 · METHOD/PRICE/PLAN/STAT/DATE LOC anthropic.com · D upd 2026-06-09/2026-07-01 · G C V DATE+STAT+PRICE+METHOD+POLICY+BENCH-V L launch and later status co-located · X X-T/X-S A ledger:S006 · 6c49d89a	
H001 · AI Cost Management overview AP 6,15,20,23 · METHOD LOC developer.harness.io · D 2026-08-28 · G C V TRACE+POLICY L visibility is not denial; validate coverage/lag · X X-I A ledger:H001 · 6c49d89a		S007 · Fable redeployment AP 8 · STAT/DATE LOC anthropic.com · D 2026-07-01 · G C V STAT L first-party status record · X X-S A ledger:S007 · 6c49d89a	
H002 · AI Cost quickstart AP 6,15,20 · METHOD LOC developer.harness.io · D 2026-08-28 · G C V GUIDE L setup is not coverage or accuracy proof · X X-M A ledger:H002 · 6c49d89a		S008 · Claude Sonnet 5 launch AP 8,17-19 · METHOD/ID/PRICE/PLAN LOC anthropic.com · D 2026-06-30 · G C V DATE+ID+TOOLS+PRICE+METHOD L price in S010; charts not rankings · X X-I/X-T/X-S A ledger:S008 · 6c49d89a	
H003 · Harness SDK integrations AP 6,15,20 · METHOD LOC developer.harness.io · D 2026-08-28 · G C V TRACE L trace estimate is not invoice; test bypass/privacy · X X-I A ledger:H003 · 6c49d89a		S010 · Claude release notes AP 8 · PRICE/PLAN/STAT/DATE LOC platform.claude.com · D upd 2026-08-28 · G C V PRICE+DATE L live changelog; Sept hike superseded · X X-T/X-S A ledger:S010 · 6c49d89a	
H004 · Cost capability matrix AP 6,15,20 · METHOD LOC developer.harness.io · D 2026-08-28 · G C V METHOD L confirm module and deployment coverage · X X-I A ledger:H004 · 6c49d89a		S033 · DeepSeek V4 Flash alias AP 8 · ART/LIC LOC huggingface.co · D 2026-08-28 · G C V ART L official DeepSeek repository · X X-A A ledger:S033 · 6c49d89a	
H005 · Harness AI Rules AP 6,15-16,20 · ID LOC developer.harness.io · D 2026-08-28 · G C V CTX L guidance is not enforcement · X X-M A ledger:H005 · 6c49d89a			

Card fields, claim codes, grades, and transfer codes: page 25.

TAKEAWAY 15 catalog records; no public company outcome.	DECISION RULE Keep seat, subscription, API, application control, and enabled module ledgers separate.
Separate Trace plan and module facts to the exact product, entitlement, connector, policy, and date.	BOUNDARY Published entitlement and module facts do not prove availability, enforcement, savings, or company fit.

Core API, Hunyuan, and MiniMax sources

Core OpenAI, Anthropic, Hunyuan, and MiniMax records support dated route, artifact, modality, tariff, and license facts.

Core API and exact artifact evidence

A	8	B	8
S001 · GPT-5.6 Sol AP 8,10,17-19 · ID/PRICE/PLAN LOC developers.openai.com · D live 2026-08-29 · G C V ID/CTX/IO/MOD/TOOLS/price+LC\$/promo>=2026-11-21 L exact promo end/reversion N/S; live page rules · X=X-I/X-T A ledger:S001 · 6c49d89a		S040 · Tencent Cloud model list AP 8,17-19 · METHOD/ID LOC intl.cloud.tencent.com · D upd 2026-08-28 · G C V ID+CTX+METHOD+IO+REAS+TOOLS L dated AI-translated Korean snapshot · X X-I A ledger:S040 · 6c49d89a	
S002 · GPT-5.6 Terra AP 8,10,17-19 · ID/PRICE/PLAN LOC developers.openai.com · D live 2026-08-28 · G C V ID+CTX+IO+MOD+TOOLS+PRICE L official live page · X X-I/X-T A ledger:S002 · 6c49d89a		S041 · MiniMax M3 launch AP 8,17-19 · METHOD/ID/ART/LIC LOC minimax.io · D 2026-06-01 · G C V DATE+ARCH+MOD+TOOLS+CTX+BENCH-V+METHOD L no June-1 weight availability · X X-I/X-A/X-S A ledger:S041 · 6c49d89a	
S003 · GPT-5.6 Luna AP 8,10,17-19 · ID/PRICE/PLAN LOC developers.openai.com · D live 2026-08-28 · G C V ID+CTX+IO+MOD+TOOLS+PRICE L official live page · X X-I/X-T A ledger:S003 · 6c49d89a		S042 · MiniMax M3 artifact AP 8,13 · ID/ART/LIC/BENCH-V LOC huggingface.co · D init 2026-06-12 · G C V ART+PARAM+CTX+MOD+REAS+BENCH-V L commit date comes from S056 · X X-I/X-A A ledger:S042 · 6c49d89a	
S004 · GPT-5.6 launch AP 10 · ID/PRICE/PLAN/STAT/DATE LOC openai.com · D refreshed 2026-08-29 · G C V pre/GA+alias+bench+temp\$+promo>=2026-11-21+REAS L vendor bench; exact promo end/reversion N/S · X=X-S/X-T A ledger:S004 · 6c49d89a		S043 · MiniMax M3 MXFP8 AP 8,13 · ART/LIC LOC huggingface.co · D init 2026-06-12 · G C V ART+BYTE L commit date S057; size is not runtime · X X-A A ledger:S043 · 6c49d89a	
S005 · Claude model overview AP 8,10,17-19 · ID/ART/LIC/PRICE/PLAN LOC platform.claude.com · D live 2026-08-28 · G C V ID+PRICE+CTX+IO+MOD+REAS+CUT L mutable live documentation · X X-I/X-T A ledger:S005 · 6c49d89a		S045 · MiniMax M3 NVFP4 AP 8,13 · ART/LIC/BENCH-V LOC huggingface.co · D 2026-08-28 · G C V ART+BYTE+BENCH-V L partner quant, not first-party · X X-A A ledger:S045 · 6c49d89a	
S036 · Tencent Hy4 preview launch AP 8,17-19 · METHOD/ID/ART/LIC LOC tencent.com · D 2026-08-28 · G C V ID+STAT+DATE+PARAM+BENCH-V+METHOD L launch-day/private evaluation · X X-I/X-A/X-S A ledger:S036 · 6c49d89a		S046 · MiniMax pay-go pricing AP 8,10 · METHOD/PRICE/PLAN LOC platform.minimax.io · D live 2026-08-28 · G C V PRICE+METHOD L current live list price · X X-T A ledger:S046 · 6c49d89a	
S038 · Hy4 preview BF16 AP 8,13 · METHOD/ID/ART/LIC LOC huggingface.co · D 2026-08-28 · G C V ART+BYTE+PARAM+METHOD+MOD L official Tencent artifact · X X-I/X-A A ledger:S038 · 6c49d89a		S055 · Claude Fable/Mythos ops AP 8,17-19 · ID LOC platform.claude.com · D live 2026-08-28 · G C V ID+IO+REFUSAL+FB+RET+TOOLS L procurement-relevant operational constraints · X X-I A ledger:S055 · 6c49d89a	
S039 · Hy4 preview FP8 AP 8,13 · ART/LIC LOC huggingface.co · D 2026-08-28 · G C V ART+BYTE L official Tencent quant · X X-A A ledger:S039 · 6c49d89a		S057 · MiniMax MXFP8 commits AP 8 · ART/LIC/STAT/DATE LOC huggingface.co · D init 2026-06-12 · G C V DATE+ART L weights chronology, not API launch · X X-A/X-S A ledger:S057 · 6c49d89a	

Card fields, claim codes, grades, and transfer codes: page 25.

TAKEAWAY 16 catalog records; no public company outcome.	DECISION RULE Use current route pages for cutoff facts and launch pages only for dated change/chronology facts.
Clock Keep cutoff tariff, temporary-change, launch chronology, API route, and artifact records distinct.	BOUNDARY Temporary prices, vendor benchmarks, and artifacts never establish horizon cost or company outcomes.

Qwen 3.8 API and checkpoint sources

Qwen records separate hosted 27B/Max/Flash routes from base and FP8 artifacts, limits, prices, and release chronology.

Qwen hosted and local evidence

A	6	B	5
S011 · Qwen Flash Next launch AP 8 · METHOD/ART/LIC/PRICE/PLAN LOC qwen.ai · D 2026-08-26 · G C V DATE+ARCH+BENCH-V+PRICE+METHOD L current Flash-price page · X X-A/X-T/X-S A ledger:S011 · 6c49d89a		S017 · Qwen Flash Next FP8 AP 8,13 · ART/LIC LOC huggingface.co · D 2026-08-26 · G C V ART+BYTE L official Qwen quant · X X-A A ledger:S017 · 6c49d89a	
S012 · Qwen3.8-27B AP 8,13 · METHOD/ID/ART/LIC LOC huggingface.co · D 2026-08-28 · G C V ART+PARAM+MOD+CTX+METHOD+LIC+BENCH-V L official Qwen artifact · X X-I/X-A A ledger:S012 · 6c49d89a		S018 · Qwen3.8-Max hosted AP 8,10,17-19 · ID/PRICE/PLAN LOC qwencloud.com · D live 2026-08-28 · G C V ID+MOD+TOOLS+PRICE+CTX+IO+REAS L hosted price; checkpoint terms separate · X X-I/X-T A ledger:S018 · 6c49d89a	
S013 · Qwen3.8-27B-FP8 AP 8,13 · ART/LIC LOC huggingface.co · D 2026-08-28 · G C V ART+BYTE+ARCH L official Qwen quant · X X-A A ledger:S013 · 6c49d89a		S019 · Qwen3.8-Flash hosted AP 8,10,17-19 · ID/PRICE/PLAN LOC qwencloud.com · D live 2026-08-28 · G C V ID+MOD+TOOLS+PRICE+CTX+IO+REAS L price differs from launch; checkpoint separate · X X-I/X-T A ledger:S019 · 6c49d89a	
S014 · Qwen3.8-2.4T-A95B AP 8,13 · ID/ART/LIC/BENCH-V LOC huggingface.co · D 2026-08-28 · G C V ART+PARAM+CTX+LIC+MOD+BENCH-V L custom license; hosted front end separate · X X-I/X-A A ledger:S014 · 6c49d89a		S020 · QwenCloud releases AP 8 · STAT/DATE LOC docs.qwencloud.com · D upd 2026-08-03...08-28 · G C V DATE L hosted dates only; no checkpoint date/bytes · X X-S A ledger:S020 · 6c49d89a	
S015 · Qwen3.8-2.4T-A95B-FP8 AP 8,13 · ART/LIC LOC huggingface.co · D 2026-08-28 · G C V ART+BYTE L datacenter first-party quant · X X-A A ledger:S015 · 6c49d89a		S059 · Qwen3.8-27B hosted AP 8,10,17-19 · ID/PRICE/PLAN LOC qwencloud.com · D live 2026-08-28/2026-08-19 · G C V ID+MOD+TOOLS+IO+PRICE+CTX L hosted/checkpoint equivalence unverified · X X-I/X-T A ledger:S059 · 6c49d89a	
S016 · Qwen Flash Next AP 8,13 · ID/ART/LIC LOC huggingface.co · D 2026-08-26 · G C V ART+ARCH+PARAM+MOD+CTX+LIC L preview; custom community license · X X-I/X-A A ledger:S016 · 6c49d89a			

Qwen hosted limits remain separate from local 262,144/YaRN context and checkpoint operation.
Card fields, claim codes, grades, and transfer codes: page 25.

TAKEAWAY 11 catalog records; no public company outcome.	DECISION RULE Match every claim to the exact hosted ID or checkpoint/precision and its own context boundary.
Pin Use exact Qwen route/checkpoint, precision, limit, tariff, and release-date records.	BOUNDARY Hosted context, cache, and tariff facts never transfer to a local checkpoint or company workload.

GLM 5.3 and DeepSeek V4 sources

GLM and DeepSeek records separate hosted routes, scheduled price clocks, dated checkpoints, aliases, precisions, and vision experiments.

GLM and DeepSeek exact-system evidence

A	6	B	6
S021 · GLM-5.3 docs AP 8,17-19 · ID/PRICE/PLAN LOC docs.z.ai · D live 2026-08-28 · G C V ID+CTX+IO+MOD+TOOLS+REAS+PRICE L current operational docs · X X-I/X-T A ledger:S021 · 6c49d89a		S029 · GLM-5.3-Flash-BF16 AP 8,13 · ART/LIC LOC huggingface.co · D 2026-08-26 · G C V ART+BYTE L official Z.ai artifact · X X-A A ledger:S029 · 6c49d89a	
S022 · GLM-5.3-Flash docs AP 8,17-19 · ID/PRICE/PLAN LOC docs.z.ai · D live 2026-08-28 · G C V ID+MOD+CTX+IO+TOOLS+REAS+PRICE L current operational docs · X X-I/X-T A ledger:S022 · 6c49d89a		S030 · DeepSeek pricing AP 8,10 · METHOD/ID/PRICE/PLAN LOC api-docs.deepseek.com · D sched 2026-08-16 · G C V IO+PRICE+METHOD L time-variable rates explicit · X X-I/X-T A ledger:S030 · 6c49d89a	
S023 · Z.ai pricing AP 8,10 · METHOD/PRICE/PLAN LOC docs.z.ai · D live 2026-08-28 · G C V PRICE+METHOD L higher Sep rate stays future · X X-T A ledger:S023 · 6c49d89a		S031 · DeepSeek V4 release AP 8,17-19 · ID/STAT/DATE/BENCH-V LOC api-docs.deepseek.com · D 2026-08-13 · G C V ID+STAT+CTX+IO+TOOLS+BENCH-V L official API release · X X-I/X-S A ledger:S031 · 6c49d89a	
S026 · GLM-5.3 artifact AP 8,13 · METHOD/ART/LIC/STAT/DATE LOC huggingface.co · D live 2026-08-28 · G C V ART+METHOD+ARCH+PARAM+LIC+STAT L counts separated; not recomputed · X=X-A/X-S A ledger:S026 · 6c49d89a		S032 · DeepSeek API updates AP 8,17-19 · METHOD/ID LOC api-docs.deepseek.com · D upd 2026-08-21 · G C V METHOD+MOD+COMPAT L vision experiment remains separate · X X-I A ledger:S032 · 6c49d89a	
S027 · GLM-5.3-BF16 AP 8,13 · ART/LIC LOC huggingface.co · D 2026-08-28 · G C V ART+BYTE L official Z.ai artifact · X X-A A ledger:S027 · 6c49d89a		S034 · DeepSeek V4 Flash 0731 AP 8,13 · ART/LIC/BENCH-V LOC huggingface.co · D 2026-07-31 · G C V ART+LIC+PARAM+BENCH-V L count not recomputed; DSpark included · X X-A A ledger:S034 · 6c49d89a	
S028 · GLM-5.3-Flash artifact AP 8,13 · METHOD/ID/ART/LIC LOC huggingface.co · D 2026-08-26 · G C V ART+LIC+METHOD+PARAM+MOD L official Z.ai artifact · X X-I/X-A A ledger:S028 · 6c49d89a		S035 · DeepSeek V4 Pro 0813 AP 8,13 · METHOD/ART/LIC/BENCH-V LOC huggingface.co · D 2026-08-13 · G C V ART+LIC+PARAM+BYTE+BENCH-V+METHOD L count not recomputed; stored total separate · X X-A A ledger:S035 · 6c49d89a	

GLM Flash list changes after 2026-09-09 24:00 UTC+8.

DeepSeek peak windows: weekday 01:00-04:00 and 06:00-10:00 UTC; experimental vision stays separate.

Card fields, claim codes, grades, and transfer codes: page 25.

TAKEAWAY 12 catalog records; no public company outcome.	DECISION RULE Treat current/scheduled GLM tariffs and DeepSeek off-peak/peak plus vision/text records as separate facts.
Separate Pin GLM/DeepSeek route, checkpoint, precision, price window, and modality before transfer.	BOUNDARY S026/S028 do not prove byte counts; aliases and deployment examples do not prove hostability or outcomes.

Emerging, launch, and closeout sources

Opus, Muse, Kimi, Gemini, and regional tariff records close the ledger with modality, artifact, pricing, and chronology boundaries.

Emerging-model evidence and closeout

A	5	B	5
S009 · Claude Opus 5 launch AP 8 · METHOD/ID/PRICE/PLAN LOC anthropic.com · D 2026-07-24 · G C V DATE+ID+PRICE+METHOD+BENCH-V+FB+POLICY L five-attempt/fallback method · X X-I/X-T/X-S A ledger:S009 · 6c49d89a		S051 · Kimi K3 artifact AP 8,13,17-19 · METHOD/ID/ART/LIC LOC huggingface.co · D 2026-07-16 · G C V ART+ID+PARAM+METHOD+MOD+CTX+LIC+BENCH-V L video unresolved pending API schema · X X-I/X-A A ledger:S051 · 6c49d89a	
S047 · Muse Glimmer launch AP 8 · METHOD/ID/ART/LIC LOC research.meta.ai · D 2026-08-10 · G C V DATE+ARCH+MOD+CTX+TOOLS+BENCH-V+METHOD L vendor-run evaluation · X X-I/X-A/X-S A ledger:S047 · 6c49d89a		S052 · Kimi API platform AP 8,10 · ID/PRICE/PLAN LOC platform.kimi.com · D live 2026-08-28 · G C V IO+PRICE L region/contract may vary · X X-I/X-T A ledger:S052 · 6c49d89a	
S048 · Muse Glimmer 30B artifact AP 8,10,13 · ART; p10 registry N/S-not tariff LOC huggingface.co · D 2026-08-10 · G C V ART+METHOD+PARAM+ENC+MOD+LANG+CUT+CTX L official Meta artifact · X X-I/X-A A ledger:S048 · 6c49d89a		S053 · Gemini 3.7 Flash reference AP 8,17-19 · ID/STAT/DATE LOC ai.google.dev · D live 2026-08-28 · G C V ID+STAT+CTX+IO+MOD+TOOLS+REAS L not price evidence · X X-I/X-S A ledger:S053 · 6c49d89a	
S049 · Muse Glimmer GGUF AP 8,13 · METHOD/ART/LIC LOC huggingface.co · D 2026-08-10 · G C V ART+BYTE+METHOD L hardware/runtime-specific · X X-A A ledger:S049 · 6c49d89a		S054 · Gemini 3.7 Flash launch AP 8,10 · METHOD/ID/PRICE/PLAN LOC blog.google · D 2026-08-13 · G C V DATE+METHOD+BENCH-V+IO+PRICE L vendor-run launch evidence · X X-I/X-T/X-S A ledger:S054 · 6c49d89a	
S050 · Moonshot model page AP 8 · ID/STAT/DATE LOC moonshot.ai · D 2026-07-16 · G C V DATE+ID L first-party product site · X X-I/X-S A ledger:S050 · 6c49d89a		S058 · Tencent model prices AP 10 · METHOD/ID LOC cloud.tencent.com · D upd 2026-08-28 · G C V METHOD+IO L confirm region and contract · X X-I A ledger:S058 · 6c49d89a	

Identity/evidence - MUST BE NAMED BEFORE CLOSEOUT

Gates/controls - MUST BE NAMED BEFORE CLOSEOUT

Economics/capacity - MUST BE NAMED BEFORE CLOSEOUT

Rollout/rollback - MUST BE NAMED BEFORE CLOSEOUT

Expiry/reopen - MUST BE NAMED BEFORE CLOSEOUT

Monthly: 2026-09-30, 2026-10-30, 2026-11-30 · final: 2026-12-18 · expiry/handoff: 2026-12-31

Working-tree methods/artifacts only; no deployment, repetition, or outcome inferred.

Bounded set: eight reproduced requested families plus Kimi K3 and Gemini 3.7 Flash; not a complete-market survey.

Card fields, claim codes, grades, and transfer codes: page 25.

EMBEDDED DECISION COMPANIONS

Open PDF Attachments panel; extract; verify SHA-256.

A1 expected_gate_catalog_v12.json | 10,567,988 bytes

A1 SHA-256 b0ad4300cd3c8aeda4964c04672827cf6e71529483b4d820aff6bd01940a2c8e

A2 shortlist_frontier_fit_v12.json | 895,605 bytes

A2 SHA-256 4549b39659fbc9a46faa5747586ebd8ff06d713187f52bc959cca7477e46c00c

A3 IMPAKT_LLM_Decision_Workbook_2026_v1.json | 840,767 bytes

A3 SHA-256 e7566cc74820d6df9c8d23689d8f0d9fca4735a56490e66e1a7675cd51a2731b

ROUTING A1 / common_core_gate_catalog + expected_gate_sets + actual_gate_set_templates

ROUTING A2 / shortlist_schema + shortlists + capacity_scaled_defaults + all_in_budget_* + fit_results

ROUTING A3 / payload.posture_selection_rubric_v1 + economics_ledger_templates + fit_execution_schema_v1

Catalog /Names /EmbeddedFiles only; no page file-attachment annotations.

TAKEAWAY	DECISION RULE
10 catalog records; no public company outcome.	Use each record only for its exact route/artifact, date, region, modality, license, and introductory-price window.
Close	BOUNDARY
Resolve remaining route, artifact, modality, price-window, and chronology questions before approval.	No unsupported September launch, video support, horizon price, or public workload outcome is inferred.