

Measuring AI by Completed Outcomes

Status: public working report Author: Njål Gaute Solland Date: 2026-07-28

1. Purpose

The ACE Economy defines Attentive Cognitive Evaluation as the scarce production factor in AI-native organizations. This contract makes the model measurable, falsifiable and operable.

The objective is not to maximize AI output or minimize human time. The objective is to maximize legitimate, risk-adjusted realized value per effective ACE-second without weakening authority, evidence, learning or execution integrity.

2. Normative principle

A human attention-second counts as ACE only when all of the following are present:

1. relevant attention;
2. sufficient context;
3. competent judgment;
4. valid authority for the decision;
5. explicit responsibility binding.

Authority is a hard gate, not a discount factor.

Effective ACE = clock seconds × attention × context × judgment × responsibility ×
authority_gate

Where:

- attention, context, judgment, and responsibility range from 0 to 1;
- authority_gate is 0 or 1;
- if authority_gate = 0, effective ACE is 0 regardless of all other factors.

This prevents an unauthorized but otherwise competent decision from being treated as partially valid ACE.

3. ACE classes

Every recorded ACE interval MUST be assigned one primary class.

ACE-P — Purpose and problem framing

Human evaluation used to define the goal, problem boundary, priorities, prohibited outcomes and acceptable trade-offs.

ACE-E — Evaluation

Human evaluation used to assess quality, relevance, uncertainty, evidence, alternatives and consequences.

ACE-A — Authorization

Human evaluation used to determine whether a consequential action may proceed within valid mandate.

ACE-A is non-substitutable where human authority is required. Additional AI analysis cannot replace missing authority.

ACE-L — Learning

Human evaluation used to interpret outcomes, update decision criteria, improve workflows and preserve institutional learning.

4. Complementary synthetic unit

An AI Deliberation Second is one second of model, tool or agent activity used to reduce uncertainty, produce options, assemble evidence or prepare action.

Raw ADS MUST NOT be treated as value.

Define Effective Synthetic Deliberation:

$ESD = ADS \times usefulness \times quality \times evidence_coverage \times novelty \times cost_efficiency$

Each factor ranges from 0 to 1.

ESD exists to prevent useless machine activity from inflating apparent ACE leverage.

5. Required measurement record

Each consequential workflow MUST produce one ACE measurement record linked to the action and execution receipt.

Required fields:

```
ace_record_version: ace_measurement_v1
workflow_id: string
outcome_id: string
action_ref: string
principal_id: string
responsible_human_id: string
ace_class: ACE-P | ACE-E | ACE-A | ACE-L
clock_seconds: number
attention_score: number
context_score: number
judgment_score: number
responsibility_score: number
authority_gate: 0 | 1
mandate_ref: string | null
evidence_coverage: number
decision_surface_items: integer
material_uncertainties: array
decision: allow | modify | defer | deny | step_up | halt
clearance_ref: string | null
execution_receipt_ref: string | null
gross_value: number | null
avoided_loss: number | null
rework_cost: number | null
incident_cost: number | null
trust_impact: number | null
learning_artifact_ref: string | null
measurement_confidence: number
measurement_method: observed | estimated | sampled | reconstructed
timestamp: string
```

6. Factor definitions

Attention

Measures whether the responsible human was actively engaged with the actual decision rather than merely present.

Evidence may include:

- explicit decision interaction;
- time on the decision surface;
- interruption rate;
- acknowledgment of unresolved uncertainty;
- confirmation of the proposed action.

Attention **MUST NOT** be inferred from online status, keyboard activity or surveillance data alone.

Context completeness

Measures whether the human had the minimum sufficient current state needed for a legitimate decision.

Context should be scored against a workflow-specific checklist, not by self-assessment alone.

Minimum dimensions:

- current action;
- affected principal;
- mandate boundary;
- material alternatives;
- consequence range;
- relevant evidence;
- unresolved uncertainty;
- reversibility;
- cumulative exposure.

Judgment competence

Measures whether the human possesses the domain competence required for the specific evaluation.

Evidence may include:

- role qualification;
- delegated decision class;
- validated experience;
- prior calibration;
- specialist step-up where required.

Judgment **MUST** be measured prospectively where possible. Outcome quality alone cannot prove competence because good outcomes can arise from luck.

Responsibility binding

Measures whether responsibility for the decision and its consequence is explicit and attributable.

A named approver without ownership is insufficient.

The binding should identify:

- responsible person;
- consequence domain;
- escalation duty;
- review obligation;
- outcome follow-up.

Authority gate

Authority is valid only where identity, mandate, scope, time, cumulative exposure and delegation chain are verified.

The gate MUST be zero when:

- mandate is absent;
- mandate is expired or revoked;
- scope is exceeded;
- delegation cannot be verified;
- cumulative limits are exceeded;
- the wrong principal is bound;
- required step-up is missing.

7. Core metrics

Effective ACE-seconds

$$\text{effective_ace_seconds} = \text{clock_seconds} \times \text{attention} \times \text{context} \times \text{judgment} \times \text{responsibility} \times \text{authority_gate}$$

ACE Yield

$$\text{ACE Yield} = \text{risk-adjusted realized value} / \text{effective ACE-seconds}$$

ACE Yield MUST only be reported when the following floors are met:

- authority gate = 1;
- evidence coverage above workflow threshold;
- execution integrity above workflow threshold;
- observed or validated outcome data available;
- material incident and rework costs included.

ACE Leverage

$$\text{ACE Leverage} = \text{ESD} / \text{effective ACE-seconds}$$

High leverage with low outcome quality, low evidence coverage or high rework MUST be classified as cognitive dumping, not productivity.

ACE Utilization

$$\text{ACE Utilization} = \text{effective ACE-seconds on human-required matters} / \text{total scarce human attention seconds}$$

The objective is not 100 percent utilization. Recovery, learning, relationship work and reflection are inputs to future ACE quality.

Authorization Compression

Authorization Compression = preparation seconds avoided / ACE-A seconds consumed

Compression is valid only if material uncertainty, alternatives, mandate boundaries and consequence range remain visible.

Decision Surface Efficiency

Decision Surface Efficiency = necessary decision information / total information presented

A smaller surface is not automatically better. It must remain sufficient for a legitimate decision.

8. Guardrails against metric gaming

No ACE metric may improve solely because the organization:

- reduces review time below a validated threshold;
- suppresses evidence or uncertainty;
- reclassifies human-required decisions as automated;
- delays incident recognition;
- excludes rework or trust damage;
- assigns nominal rather than real responsibility;
- treats unauthorized decisions as partially valid;
- increases synthetic output without measuring usefulness.

Every dashboard MUST display ACE Yield together with:

- outcome quality;
- authority integrity;
- evidence coverage;
- rework rate;
- reversal rate;
- incident rate;
- unresolved uncertainty;
- ACE load concentration.

9. Learning and capital accounting

Learning MUST NOT be counted simultaneously as full current value, full capital creation and full future cost reduction.

A learning artifact may be recognized as ACE or context capital only when it is:

1. documented;
2. linked to an observed outcome;
3. reusable beyond the original decision;
4. assigned an owner;
5. reviewed for continued validity.

Suggested treatment:

- creation cost is expensed in the current period;
- validated reusable learning may be registered as non-financial ACE capital;
- future cost reduction is recognized only when observed;
- depreciation is recorded when knowledge becomes stale, inaccessible or unused.

10. Falsifiable hypotheses

The model is supported only if repeated observation confirms hypotheses such as:

1. Higher context completeness reduces rework and reversal rates.
2. Verified mandate at execution reduces unauthorized consequence and remediation cost.
3. Better decision surface efficiency reduces ACE consumption without reducing outcome quality.
4. Higher ACE utilization increases risk-adjusted value until decision load degrades ACE quality.
5. Synthetic output has negative marginal return when evaluation capacity is saturated.
6. ACE-A cannot be substituted by additional ADS where human authority is legally or institutionally required.
7. Preserved outcome evidence increases future ACE Yield through better calibration and learning retention.

A failed hypothesis **MUST** trigger revision of the model rather than reinterpretation of the data to protect it.

11. Pilot design

The first implementation **SHOULD** use three workflows with different consequence profiles:

1. low-risk reversible knowledge work;
2. medium-risk enterprise approval;
3. high-risk financial, legal, personnel or production action.

For each workflow, record at least:

- baseline human time;
- synthetic preparation time;
- effective ACE-seconds by class;

- evidence coverage;
- rework;
- decision reversal;
- execution result;
- realized value or avoided loss;
- incident cost;
- learning artifact production.

Compare:

- human-only workflow;
- AI-assisted workflow without execution governance;
- AI-assisted workflow with execution-boundary check clearance and receipts.

12. Relationship to execution-boundary check and enforcement layer

ACE determines where scarce human evaluation creates value.

execution-boundary check determines whether the required evaluation and authority are present immediately before consequential action.

enforcement layer enforces the resulting decision.

Receipts bind clearance, execution and outcome evidence.

Canonical chain:

Purpose -> Synthetic preparation -> ACE-E -> ACE-A -> execution-boundary check -> enforcement layer -> Execution -> Receipt -> Outcome -> ACE-L

The system must not confuse evaluation with authorization.

A validator may assess evidence.

A model may recommend an action.

Only the valid authority holder or a properly delegated deterministic policy may authorize the consequence.

13. Decision rule

The governing economic objective is:

Maximize legitimate, risk-adjusted realized value per effective ACE-second, subject to hard constraints on authority, evidence, execution integrity, trust and learning retention.

This is the operational definition of the ACE Economy.

14. Implementation

The contract is implemented and tested in this repository:

- `contracts/ace-measurement-v1.json` — machine-readable JSON-Schema for the `ace-measurement.v1` record.
- `src/ace-measurement.ts` — TypeScript implementation of `aceQuality`, `effectiveAceSeconds`, `productionValue`, `aceYield`, `aari`, and `computeAce`. The authority gate is enforced as a hard $0|1$ multiplier so unauthorized decisions accrue zero effective ACE. `aceYield` is bounded: weakening evidence coverage or execution integrity strictly reduces yield.
- `test/ace-measurement.test.ts` — falsifiable tests, including: invalid/empty mandate forces ACE to 0; ACE Yield cannot rise by weakening legitimacy; ACE-A is distinct from ACE-E (analysis cannot substitute authorization).

This closes the gap identified in review: the model is now measurable, falsifiable and runnable, and links to execution receipts (execution-boundary check/enforcement layer) via `executionReceiptRef` and `mandateRef`.