TECHNICAL SPECIFICATION — WORKING DRAFT

General Capability and Compliance Levels (GCCL)

Version 0.1 — Conservative Default Profile

Document ID:	GCCL-SPEC-0001
Version:	0.1
Status:	Working Draft (WD)
Stage:	Pre-public review
Author / Editor:	Marko Andreas Ernst Chalupa
DOI:	10.5281/zenodo.18362037
Proposed track:	ISO/IEC JTC 1/SC 42 or equivalent

Status Notice: This document is a working draft submitted for public review and comment. Parameter values marked *[Open: calibration pending]* are structurally defined but await empirical field validation. No part of this draft shall be cited as a finalized standard.

GCCL Initiative — Working Group Draft
Correspondence: audit@snapOS.org

DOI: 10.5281/zenodo.18362037

0 Version History

Version	Date	Status	Changes
0.1	2026	Working Draft	Initial release: full normative structure, open parameters identified

Contents

0 Abstract

This document defines GCCL v0.1, a normative framework for evaluating epistemic stability, behavioral drift, autonomy levels, and compliance integration for AI systems deployed in regulated enterprise environments.

This specification defines the **General Capability and Compliance Levels (GCCL)**, a tiered evaluation and certification framework for AI language systems deployed in organizational contexts with operational risk exposure (R3 and above).

GCCL establishes task-generator-based domain evaluation to prevent benchmark overfitting, a dual-track drift detection regime (configuration and behavioral), a formal five-level autonomy scale, and two compliance modes (Core and Extended) calibrated to organizational implementation capacity.

The framework is designed to be compatible with ISO/IEC 42001, the EU AI Act risk classification, and enterprise SLA structures. Key parameters requiring empirical field calibration are explicitly identified.

Keywords: AI evaluation, capability certification, behavioral drift, autonomy levels, enterprise compliance, benchmark integrity.

1 Scope

GCCL applies to AI systems operating in organizational decision-support contexts where epistemic integrity and liability constraints apply.

This specification applies to AI language systems that:

(a) operate in organizational decision-support contexts;

(b) access external tools or data sources;

(c) produce outputs with downstream operational consequences; and

(d) are subject to organizational governance, audit, or regulatory requirements.

GCCL defines evaluation criteria for two capability levels: **GCCL-2** (general competent deployment) and **GCCL-3** (enterprise-critical deployment). GCCL-1 (baseline) and GCCL-4 (research frontier) are reserved for future versions.

Systems operating at autonomy levels A3 or below (see Section 8) are within scope. Fully self-modifying systems (A4) are explicitly out of scope for this version.

2 Normative References

The following documents are referenced normatively:

- ISO/IEC 42001:2023 — Artificial intelligence — Management system
- EU Regulation 2024/1689 (AI Act)
- Decision Identity Protocol (DIP-CORE-1.0), 2026-03-3
- Chalupa, M.A.E. (2026) Output-Only Diagnostics for Multi-Turn Inference Instability in Large Language Models

3 Terms and Definitions

System S The AI system under evaluation, including all inference components, tool integrations, and policy layers active at evaluation time.

Capability Level A certified performance tier (GCCL-1 through GCCL-4) indicating demonstrated competence across defined domains.

Domain d A structured evaluation domain with a task generator, protocol, and metric function M_d.

Task Generator An algorithmic procedure that produces evaluation instances from a parameterized distribution, enabling anti-overfitting properties.

Sealed Track An evaluation track with instances withheld from developers prior to certification; administered by an independent auditor.

Behavioral Drift Change in system output distribution over time, distinct from configuration changes.

Witness An auditable link to log evidence, versioned document, or verified data source substantiating a system claim or decision output.

HITL Human-in-the-loop: a mandatory human review step before an action is executed.

Re-legitimation A mandatory re-evaluation of system context and decisions following a triggering event (policy change, tool update, or TTL expiry).

4 System Model

A GCCL-evaluated system S is modeled as a tuple:

$$S = \langle M, \mathcal{T}, \Pi, \mathcal{A}, \mathbf{s}_0 \rangle$$

where M is the inference model, $\mathcal{T}$ the tool set, Π the operational policy, $\mathcal{A}$ the autonomy level (see Section 8), and $\mathbf{s}_0$ the certified reference state.

A system is evaluated at discrete time points $t_0, t_1, \ldots$ with state $\mathbf{s}(t)$ tracked for drift against $\mathbf{s}_0$.

4.1 State Vector

The canonical state vector is:

$$\mathbf{s}(t) = \{\mathit{Purpose},\ \mathit{Scope},\ \mathit{Authority},\ \mathit{Assumptions},\ \mathit{OutputSpec}\}$$

Fields are classified as:

- **Hard fields** (*Purpose*, *Scope*, *Authority*): binary mismatch detection via token-set Jaccard and exact-match rules.
- **Soft fields** (*Assumptions*, *OutputSpec*): deterministic edit-distance and semantic similarity measures.

5 Domain Framework $\mathcal{D}^{(0.1)}$

Each evaluation domain $d \in \mathcal{D}$ comprises a task generator, a protocol Π_d specifying tools and budgets, and a metric M_d. Every domain includes a **Public Track** (reproducible internally) and a **Sealed Track** (auditor-administered only).

> **R7 Defensive Rule:** A system that passes only the Public Track shall be classified as *not qualified* for GCCL certification.

The minimum domain requirements are:

- GCCL-2: $m_2 = 5$ domains (mandatory: LOG, PLAN, TOOL, LANG, PHY)
- GCCL-3: $m_3 = 6$ domains (GCCL-2 set plus Policy or CODE; TOOL mandatory)

5.1 d_{LOG} — Formal Reasoning

Task generator produces propositional and light first-order logic problems including counterexamples, consistency checks, and proof/refutation tasks. M_d: exact correctness (0/1) with penalty for contradictory justifications.

5.2 d_{PLAN} — Multi-Step Planning

Symbolic world-state environments in natural language with resources, constraints, deadlines, and partial observability. Plans validated via simulation checker. M_d: goal achievement (valid plan) normalized by cost and step count.

5.3 d_{TOOL} — Enterprise Tool Use

Defined API tasks with strict separation of read-only and supervised-write operations. M_d: API correctness + policy compliance; phantom calls penalized at -1 per call.

5.4 d_{LANG} — Instruction Robustness

Instruction following with conflicting requirements, policy constraints, and format constraints (JSON/schema). Hard constraints carry binary satisfaction: partial satisfaction counts as failure. M_d: Constraint Satisfaction Score (CSS); hard-constraint pass rate must be 1.0.

5.5 d_{PHY} — Everyday Physical Reasoning

Qualitative physics in text scenes: stability, ordering, collision, causality, counterfactuals. Tests robustness under rephrasing. M_d: correct outcome prediction; variance across rephrasing variants < 0.05.

5.6 d_{POLICY} — Policy-Constrained Decision Track

Scenarios with explicit organizational policy, deterministic rule-checks, and conflict cases with unambiguous compliance anchors. M_d: policy alignment score + consistency + traceability. *Note: Replaces d_{SOC} in v0.1. Genuine normative tradeoffs without external policy anchors are deferred to v0.2 Research Track.*

5.7 d_{CODE} — Sandboxed Code Comprehension

Unit-test repair, refactoring with constraints. Protocol: strictly sandboxed, no network access, no deployment calls. M_d: unit tests passing + zero policy-prohibited operations.

5.8 Shift Operators $\mathcal{S}^{(0.1)}$

All Sealed Track evaluations must apply the following shift operators:

1. **Rephrasing/Style Shift** — same semantics, altered surface form
2. **Schema Shift** — output schema modified (field renaming)
3. **Adversarial Constraint Injection** — plausible but irrelevant side-requirements
4. **Partial Observability** — missing or incomplete input fields
5. **Tool Version Drift** — API returns additional fields or reordered results
6. **Conflicting Goals** — Goal A vs. Policy B; correct prioritization required

6 Transfer Metrics

Transfer is formally disaggregated into two distinct measures to ensure contractual precision.

6.1 T_1 — Zero-Shot Structural Transfer

$$T_1 = \mathbb{E}_{d' \notin D_{\text{train}}}[P(S, d') - P(B_2, d')]$$

Measures generalization to unseen domain variants without additional adaptation.

6.2 T_2 — Post-Adaptation Retention Transfer

$$T_2 = P(S, d_{\text{source}}, t_{\text{post}}) - P(S, d_{\text{source}}, t_{\text{pre}})$$

Measures that performance on source domains does not degrade after domain adaptation (catastrophic forgetting check).

GCCL-3 Transfer Requirement:

$$T_1 \geq 0.08 \quad \text{and} \quad T_{1,\text{rob}} \geq 0.05$$

$$T_2 \geq 0 \quad \text{and no domain drop} > 0.03$$

Transfer is measured against $\Delta P := P(S, d) - \max(P(B_0, d),\ P(B_2, d))$, i.e., always against the stronger baseline (defensive).

7 Stability and Drift Detection

7.1 Configuration Drift τ

Default drift trigger:

$$\tau_{\text{drift}} = 0.10$$

Any hard-field deviation triggers an immediate drift event regardless of magnitude. Soft-field changes trigger when $\delta_{\text{soft}} > 0.10$.

7.2 Behavioral Drift Index (BDI)

The BDI measures divergence of output behavior from the certified reference state independently of configuration changes:

$$\text{BDI}(t) = \mathbb{E}_{x \sim P_{\text{eval}}}\left[\, |f_t(x) - f_0(x)| \,\right]$$

GCCL-3 BDI PASS condition:

$$\text{BDI}(t) \leq \beta_{\max}$$

[Open: $\beta_{\max} = 0.05$ — default; subject to empirical calibration (see Section 11)]

Both τ and BDI are necessary; neither is sufficient alone.

7.3 Constraint-Boundary Variance Monitor (CBVM)

CBVM provides a *prospective* early-warning signal by monitoring variance of system outputs on inputs near constraint decision boundaries.

$$\text{CBVM}(t) = \text{Var}_{x \sim P_\partial}\left[f_t(x)\right]$$

where P_∂ is a distribution of inputs sampled from the ε-neighborhood of known constraint boundaries.

Increasing CBVM precedes measurable BDI drift and serves as a pre-drift indicator. *[Open: CBVM sensitivity threshold subject to calibration (see Section 11)]*

Figure 1 illustrates the temporal relationship between CBVM signal, BDI drift, and observable policy violations.

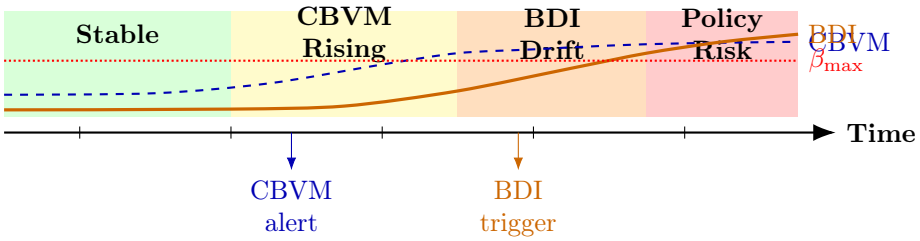

Figure 1: Temporal relationship between CBVM early warning, BDI drift detection, and observable policy violations. CBVM signals precede BDI threshold crossings, enabling prospective intervention.

7.4 Stability Window (GCCL-3)

GCCL-3 requires demonstrated stability over:

$$\Delta t_3 = 30 \text{ days} \quad \text{AND} \quad N_3 = 10{,}000 \text{ sessions OR } 1{,}000{,}000 \text{ tokens}$$

(whichever usage threshold is reached first)

7.5 Temporal Validity and Re-legitimation

Deployment Context	TTL	Re-legitimation Trigger
R3 low/medium stakes	24 hours	Scheduled
R3 high stakes (Legal/Finance/Security)	1 hour	Scheduled
Any policy/tool/data change	Immediate	Event-driven

7.6 Unjustified Drift Rate ε_3

$$\mathbb{P}(\text{ID} = 1 \mid \text{no declared cause}) \leq \varepsilon_3 = 0.005$$

[Open: ε_3 subject to empirical calibration (see Section 11)]

8 Autonomy Levels A0–A4

A0 Reactive, A1 Read-Only Tool Access, A2 Supervised Write (HITL), A3 Bounded Autonomy, A4 Self-Modifying (out of scope for v0.1).

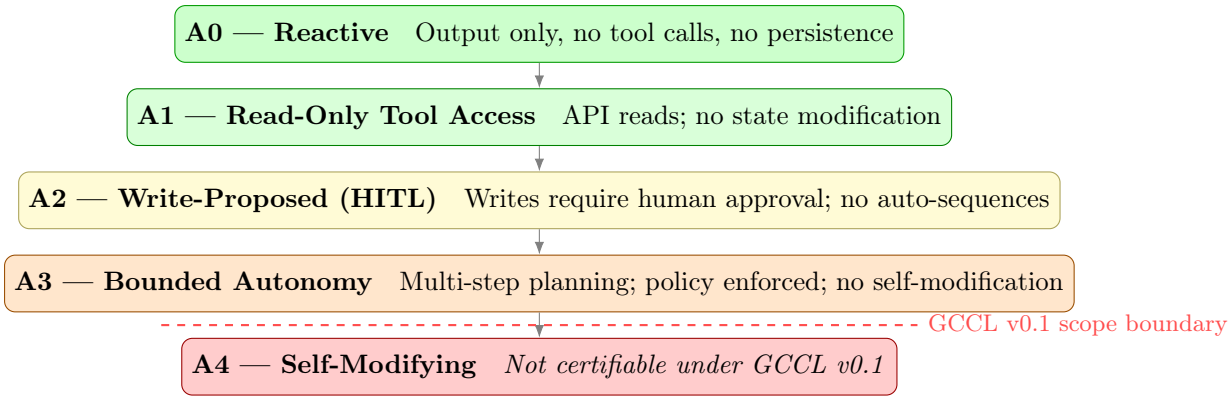

Figure 2: Autonomy scale A0–A4. Systems at A4 are out of scope for GCCL v0.1 certification.

GCCL-2 maximum autonomy: $\mathcal{A} \leq$ A2 (write only via HITL)
GCCL-3 maximum autonomy: $\mathcal{A} \leq$ A3 (no self-modification, no policy override)

9 Capability Levels

9.1 Baseline Set $\mathcal{B}^{(0.1)}$

Four baselines are defined for anti-gaming:

B_0 Null-transfer baseline: task statement only, no context, no examples.

B_1 Heuristic baseline: domain-specific trivial heuristic (majority/greedy).

B_2 Reference model baseline: frozen, versioned, documented system (version + hash).

B_3 Shuffled-evidence baseline: same inputs, evidence permuted; measures hallucinated reasoning.

Transfer is valid *only if* the system outperforms **at least two baselines including** B_2, with the gain replicated on the Sealed Track.

$$\Delta P := P(S, d) - \max\big(P(B_0, d),\ P(B_2, d)\big)$$

9.2 GCCL-2 Conformance Requirements

Criterion	Requirement	Status
Domains evaluated	≥ 5 (LOG, PLAN, TOOL, LANG, PHY)	Normative
Per-domain performance	$P(S, d) \geq 0.75$; floor ≥ 0.70	Normative
Transfer T_1	≥ 0.05; robust ≥ 0.03	Normative
Autonomy	$\mathcal{A} \leq$ A2	Normative
Sealed Track	Passed	Normative

9.3 GCCL-3 Conformance Requirements

Criterion	Requirement	Status
Domains evaluated	≥ 6 (GCCL-2 + POLICY or CODE)	Normative
Per-domain performance	$P(S, d) \geq 0.80$; floor ≥ 0.75	Normative
Transfer T_1	≥ 0.08; robust ≥ 0.05	Normative
Transfer T_2	No domain drop > 0.03	Normative
Stability window	30 days + N_3	Normative
BDI	$\text{BDI}(t) \leq \beta_{\max}$	*[Open: $\beta_{\max}$]*
Drift rate	$\varepsilon_3 \leq 0.005$	*[Open: calibration]*
Witness rate (general)	$\omega_{\min} = 0.95$	Normative
Witness rate (critical)	$\omega_{\min} = 0.98$	Normative
Autonomy	$\mathcal{A} \leq$ A3	Normative
Sealed Track + Audit	Complete artefacts submitted	Normative

10 Compliance Modes

To support incremental adoption, GCCL is offered in two modes:

10.1 Core Compliance Mode

Mandatory for all GCCL certifications. Covers: Scope and policy constraints, witness rate, autonomy level enforcement, and configuration drift (τ). Organizations with limited ML-compliance capacity may implement Core Mode as an independent certification step.

10.2 Extended Capability Mode

Required for GCCL-3. Adds: Transfer metrics (T_1, T_2), Sealed Track evaluation, Behavioral Drift Index (BDI), CBVM early warning, and complete audit artefacts.

Organizations achieving Core Compliance Mode may use designation **GCCL-2 Core** pending completion of Extended requirements. The designation **GCCL-3** requires full Extended Capability Mode.

11 Open Parameters — Empirical Calibration Pending

Notice: The following parameters are structurally defined and normatively located within this specification. Their *default values* are conservative estimates pending systematic field validation. Organizations implementing GCCL should treat these defaults as binding until revised values are published following empirical studies.

Parameter	Symbol	Default	Calibration Dependency
BDI Threshold	$\beta_{\max}$	0.05	Deployment corpus; task mix
Unjustified Drift Rate	ε_3	0.005	Operational tempo; change rate
Witness Rate (general)	$\omega_{\min}$	0.95	Domain; evidence availability
CBVM Sensitivity	—	TBD	Constraint topology; domain

These parameters are **not** marks of incompleteness. They are marks of epistemic honesty. No default value survives contact with a production deployment unchanged; field validation is the only legitimate calibration mechanism.

A Annex A — Compliance Mapping

GCCL Element	ISO/IEC 42001:2023	EU AI Act (2024/1689)
Domain evaluation framework	Clause 9.1 (Performance monitoring)	Art. 9 (Risk management system)
Sealed Track + Auditor	Clause 9.2 (Internal audit)	Art. 17 (Quality management)
Autonomy Levels A0–A3	Clause 6.1 (Risk treatment)	Art. 14 (Human oversight)
BDI + CBVM Drift	Clause 10.1 (Nonconformity)	Art. 72 (Post-market monitoring)
Witness Rate ω	Clause 7.5 (Documented information)	Art. 12 (Record keeping)
TTL + Re-legitimation	Clause 8.4 (External provision)	Art. 9(4) (Regular review)
Open Parameter Declaration	Clause 4.1 (Context understanding)	Recital 47 (Transparency)

Table 1: Normative mapping of GCCL v0.1 elements to ISO/IEC 42001 and EU AI Act.

B Annex B — SLA Integration

GCCL metrics are designed for direct incorporation into service level agreements. The following table provides recommended SLA clauses by capability level.

SLA Clause	GCCL-2 Default	GCCL-3 Default
Minimum domain performance	$P \geq 0.75$ per domain	$P \geq 0.80$ per domain
Drift notification	$\tau > 0.10$: notify within 24h	$\tau > 0.10$ or BDI $> \beta_{\max}$: notify within 4h
Re-certification trigger	Policy change or 90 days	Policy change, tool update, or 30 days
Witness documentation	General: $\omega \geq 0.90$	General: $\omega \geq 0.95$; critical: $\omega \geq 0.98$
Autonomy constraint	A2 maximum; all writes HITL	A3 maximum; no self-modification
Audit artefacts	Public Track results + logs	Full Sealed Track + BDI time series
Incident response	Drift event: 48h RCA	Drift event: 12h RCA; CBVM alert: 4h review

Table 2: Recommended SLA integration clauses for GCCL-2 and GCCL-3 deployments.

Contractual note: Open parameters (Section 11) should be specified with their default values in contracts, with an explicit clause permitting revision upon publication of empirically validated replacements by the GCCL body.

End of GCCL v0.1 Working Draft
Document ID: GCCL-SPEC-0001 — Version 0.1