AI Safety and Anti-Fabrication Policy

Effective Date: April 22, 2026
Version: v1.0.0
Last Updated: April 22, 2026
Published at: capstonehorizon.com/legal/ai-safety-antifab
Current rule version: v2, shipped April 21, 2026, commit 54e0bde
Contact: security@capstonehorizon.com

Why This Matters

Federal proposals carry serious legal weight. A proposal that claims the contractor has personnel it doesn't have, past performance it didn't earn, security clearances it doesn't hold, or language fluency that nobody on the team has is not just a bad proposal -- it can be the basis for a False Claims Act violation under 31 U.S.C. § 3729 et seq. In federal procurement, the signed certifications on an offer carry personal and organizational liability.

AI-assisted writing tools are useful for drafting. They are also capable of fabricating plausible-sounding details if not constrained. DeployReady was built with this risk at the center of the design. Our proposal engine operates under a set of anti-fabrication rules that are enforced in code, not just policy.

This document explains how those rules work, what the system does when it detects a gap between your profile and a solicitation's requirements, and what your responsibility is as the human reviewer before you sign and submit anything.

The v2 Anti-Fabrication Rules

The rules live in `server/proposal-engine-v2.ts` at lines 582-648 (commit 54e0bde). There are 9 numbered rules. Here is each one in plain English:

Rule 1: Only your company profile is the source of truth.
Every proposal generation request injects your complete `CompanyProfileForProposal` block as the authoritative fact set. The proposal engine may only use facts from that block. It cannot consult its own training data, infer from industry norms, or extrapolate from context clues. If a fact is not in your profile, it does not appear in the proposal.

Rule 2: No invented personnel.
The engine cannot create names, titles, roles, years of experience, education, or qualifications for any person. If a solicitation requires a key personnel resume and no one in your profile matches, the engine flags the gap -- it does not invent a person.

Rule 3: No invented past performance.
The engine cannot fabricate contract numbers, dollar values, client names, period of performance, or performance narratives. Past performance entries in the proposal must come verbatim or in summary from the past performance records in your company profile. If there is no matching past performance, the engine says so.

Rule 4: No invented clearances.
The engine cannot claim that your organization, any employee, or any proposed subcontractor holds a security clearance that is not explicitly listed in the company profile. If the solicitation requires cleared personnel and none are in your profile, the gap is surfaced before generation.

Rule 5: No invented language fluency.
The engine cannot claim foreign language proficiency for any individual unless that language is explicitly listed in the person's profile record. No "likely speaks Spanish" inferences based on a name or location.

Rule 6: No invented quantitative claims.
The engine cannot generate performance metrics, throughput numbers, cost figures, staffing ratios, or any other quantitative claim that is not drawn from your profile data. "We achieved a 99.8% on-time delivery rate" requires that number to be in your profile.

Rule 7: Restricted compliance matrix vocabulary.
When generating a compliance matrix, the engine is limited to three labels:

- `ADDRESSED VIA RECRUITMENT` -- the requirement is not currently met, and the proposed path is to hire for it.
- `PARTIALLY COMPLIANT` -- the company meets some but not all aspects of the requirement.
- `WILL BE COMPLIANT AT PHASE-IN` -- the company commits to meeting the requirement by contract start or a defined milestone.

The label `ADDRESSED VIA TEAMING` is only permitted if a named teaming partner appears in the company profile. The engine cannot use this label to imply a teaming relationship that is not documented.

Rule 8: No narrative from skill tags.
The engine cannot construct a past performance story or capability narrative from a list of skills. If your profile lists "Agile, Scrum, Jira" under capabilities, the engine cannot write "We successfully led a $5M Agile transformation for a federal agency" -- that is a fabricated narrative. Skills inform what can be described; they do not authorize inventing the description.

Rule 9: Default gap framing is "will recruit."
When the engine identifies a requirement gap -- a solicitation requirement that your current profile does not satisfy -- the default output framing is that your company will recruit qualified candidates to fill the gap before contract start. This is the lowest-risk framing: it makes no false claims and preserves optionality. You can edit this framing if you have a different plan (subcontracting, teaming), but the engine does not make that claim on your behalf without your input.

RAG Grounding Explained

RAG stands for Retrieval-Augmented Generation. In DeployReady's context, it means that the proposal engine does not operate from the LLM's training data alone. Before generation, the system retrieves your complete company profile and injects it as a structured block in the prompt. The LLM is instructed to treat this block as the only authoritative source of facts.

This is why you must keep your company profile accurate and current. The quality of the output is a direct function of the quality of your input. A profile with outdated past performance, missing certifications, or stale personnel records will produce a proposal that reflects those gaps -- and the anti-fabrication rules will prevent the engine from filling them in.

Capability Gap Detection

Before the proposal generation engine runs, DeployReady compares your company profile against the solicitation's explicit requirements. If it finds a requirement your profile does not satisfy -- a required clearance level, a specific language, a technical certification, or a past performance requirement in a specific NAICS code -- it surfaces these gaps to you on screen before you proceed.

You can then:

1. Update your profile to add the information if it is accurate.
2. Proceed with "will recruit" framing for personnel or capability gaps.
3. Note a teaming or subcontracting approach in your profile so the engine can reference it.
4. Decide not to bid if the gaps are too large to bridge credibly.

The gap detection step is a human decision point by design. The engine does not bypass it.

LLM Provider Data Handling

DeployReady uses three LLM providers via API: OpenAI, Anthropic, and xAI (Grok). When a proposal generation request runs, relevant content from your company profile and the solicitation is included in the API prompt.

Training on your data:

- OpenAI: Per OpenAI's API data usage policies, inputs to the API are not used to train OpenAI models by default.
- Anthropic: Per Anthropic's privacy policy, API inputs are not used to train Anthropic's models by default.
- xAI: xAI's current API data use terms should be reviewed directly at x.ai. Capstone monitors each provider's terms and will update this document if policies change.

Cross-tenant sharing: Capstone does not share prompts between tenants. Each generation request is isolated to the requesting tenant's session and profile data.

Prompt retention: Capstone does not retain LLM prompt content beyond what each provider requires for their own API operations. Capstone does not build a training dataset from customer prompts.

What DeployReady Does NOT Do

- Train any Capstone-internal model on your company profile, solicitation uploads, or generated proposals.
- Share your prompts, profile data, or outputs with any other DeployReady customer.
- Use your data for purposes other than providing the Service.
- Claim compliance labels in generated proposals beyond the three permitted vocabulary terms.
- Substitute the AI's knowledge of "typical" contractors for your actual profile.

Customer Responsibility

The anti-fabrication rules significantly reduce the risk of a proposal with invented facts. They do not eliminate all risk, and they are not a substitute for your review.

Before submitting any proposal generated with DeployReady, you are responsible for:

1. Verifying that all factual claims are accurate. Read every section for accuracy against your actual capabilities and records.
2. Confirming that all personnel listed exist and are available. Key personnel representations in proposals are binding.
3. Confirming that all past performance citations are accurate. Verify contract numbers, dollar values, and client names.
4. Ensuring compliance with the specific solicitation's instructions. The proposal engine produces a draft; format, page limits, and section ordering requirements must be verified against the solicitation.
5. Reviewing the compliance matrix. Confirm that every label is accurate and defensible.

By submitting a proposal, you -- not Capstone -- certify the accuracy of its contents. DeployReady is a drafting tool. The final product and its accuracy is your responsibility.

Version History

Version	Date	Description
v1	Pre-April 2026	Initial anti-fabrication rules
v2	April 21, 2026	9 numbered rules enforced in `server/proposal-engine-v2.ts` lines 582-648, commit 54e0bde
v3	Planned	Next iteration as real-world usage surfaces new failure modes

Current version: v2 rules, shipped April 21, 2026 at commit 54e0bde. Next iteration (v3) is planned as real-world usage surfaces new failure modes.

Contact

security@capstonehorizon.com
Capstone Horizon LLC, 7782 Rose Rock Way, Las Vegas, NV 89113
capstonehorizon.com