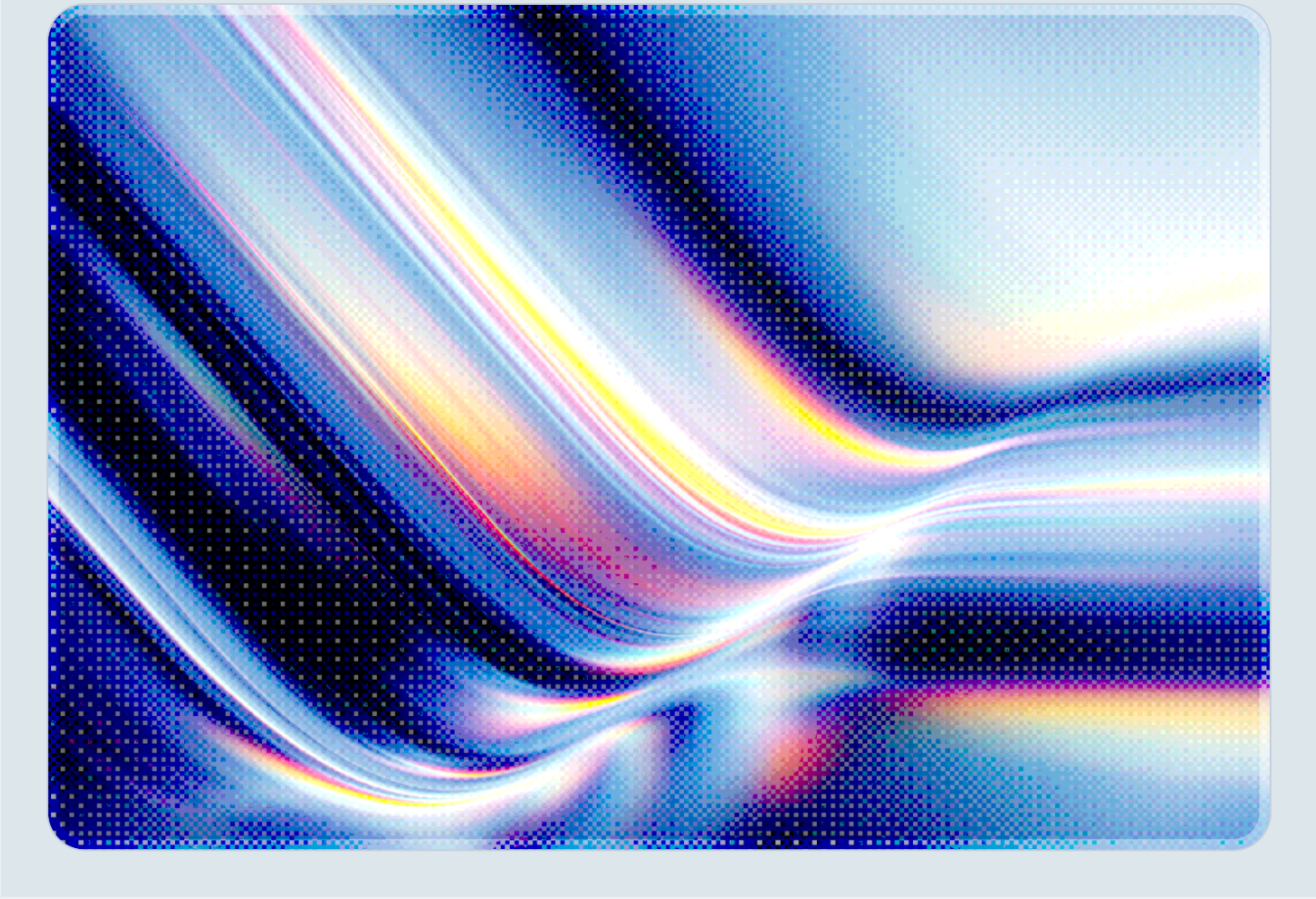
CSAI Foundation | Cloud Security Alliance

AI-Assisted CVE Enrichment: A Research Agenda and Pilot Proposal

Empirical analysis of the post-NVD enrichment landscape and a calibration pilot to test feasibility of AI-drafted, human-verified CPE and CVSS enrichment

2026-04-25

Unofficial AI-assisted Research



CSAI

CSA cloud security alliance®

Table of Contents

Executive Summary

The popular framing of NVD's recent trajectory — that it collapsed in 2024 — is not what the data shows. A per-CVE-year analysis of 263,544 records across 27 years of operational history shows that NVD did not collapse; it restructured. CVSS severity scoring, long treated as NVD's signature output, was largely transferred to CVE Numbering Authorities between 2019 and 2024. CWE classification followed the same pattern. CISA's Authorized Data Publisher program absorbed SSVC enrichment and short-turnaround analysis at scale starting in 2024. Across these fields, top-line coverage has barely moved even as NVD's authorship share has collapsed.

The remaining gap is narrower and more specific. Common Platform Enumeration applicability — the structured product identification that vulnerability scanners depend on to match CVE records to deployed software — was not picked up by any other party. Coverage fell from approximately 95 percent in 2022 to 57 percent in 2026, and the data does not show a third party stepping in. Roughly 25,000 to 30,000 CVEs per year are now published with no machine-readable product identification from any source. A secondary, methodologically distinct issue is the absence of neutral validation against CNA-supplied CVSS vectors, which historically came from NVD and now does not arrive consistently from any source.

This paper proposes that the Cloud Security Alliance respond to this situation as a **research and coordination entity** rather than as a production data operator. The proposal has three parts. First, the empirical findings establishing what the gap is and is not. Second, a tightly bounded definition of the problem to be solved, with explicit constraints on what AI-assisted enrichment can and cannot accomplish under realistic data availability conditions. Third, a Phase 0 calibration pilot — a 1,000-CVE proof of concept — designed to validate the feasibility of AI-drafted, human-verified enrichment before any operational program is committed.

The pilot is the central deliverable of this paper. Its purpose is not to declare a program ready, but to test the assumptions underlying the proposed pipeline against the actual backlog population: token consumption distributions, retrieval frequency, human review time per record, and accuracy against expert ground truth. Pilot outputs are designed to be open and independently scrutinized: a published codebase, a calibrated dataset, distributional cost measurements, and a technical report. A successful pilot enables CSA and the broader community to define what a sustainable enrichment program would require. An unsuccessful pilot — or one that surfaces unexpected variance — produces equally useful evidence for the industry conversation about what cannot be automated and where human-only enrichment must remain.

This document does not propose vendors, sponsors, or operating commitments. It does not claim production readiness. It treats the cost figures presented here as pre-pilot estimates whose primary purpose is to bound the order of magnitude and to be replaced by measured values after the pilot concludes. The

accompanying interactive cost calculator, hosted at dashboard.csai.foundation/nvd-enrichment-calculator.html, is structured around distributional modeling — p50, p80, and p95 cost estimates rather than single averages — to reflect the genuine uncertainty in pre-pilot parameters.

1. Empirical Findings

The findings in this section are drawn from analysis of 263,544 CVE records spanning 1999 through 2026, sourced from `CVEProject/cvelistV5`, `cisagov/vulnrichment`, the public NVD JSON corpus, and the FKIE-CAD NVD JSON data feeds. The analysis identifies seven operational eras of NVD activity, quantifies the post-2022 transitions in authorship share, and isolates the specific fields where the post-NVD ecosystem has and has not absorbed previously NVD-produced work.

1.1 The Seven NVD Eras

Per-CVE-year coverage analysis identifies seven distinct operational profiles in NVD's history, each defined by a different signature of which enrichment fields were produced and at what coverage. The Foundation era (2000 to 2006) ran a three-field program of CVSS v2, CWE, and CPE at near-universal coverage. The Multi-Lingual Expansion era (2007 to 2014) added a Spanish translation program that brought coverage from below 10 percent to approximately 95 percent within a single year. The CVSS v3.0 Rollout (2015 to 2018) and v3.1 Transition (2019) eras dual-scored and then triple-scored CVEs across overlapping CVSS versions. Peak NVD (2020 to 2021) is the period when all enrichment fields ran simultaneously at 85 to 95 percent coverage; this is the period against which most modern security tools were calibrated, and it lasted two years.

Silent Triage (2022 to 2023) is the era most relevant to the current planning context. Before any public crisis, NVD had quietly cut CVSS v2 authorship from 86 percent to zero, allowed Spanish coverage to drop from 94 percent to 57 percent, and reduced CWE authorship — while protecting CVSS v3.1 and CPE at 83 to 95 percent. The Visible Drawdown era (2024 to present) is the period in which the previously protected fields have also begun declining. NVD-authored CVSS v3.1 fell from 83 percent to 31 percent. NVD-authored CWE fell from 65 percent to 11 percent. NVD-authored CPE fell from 93 percent to 56 percent.

The strategic interpretation supported by this history is that the "NVD provides comprehensive enrichment" assumption was reliably true for two years out of 27, and silent cuts preceded each visible cut by approximately two years. The data does not support extrapolating from any single year as a steady state. It does support planning under the assumption that field-by-field changes are continuous and that retrospective measurement is more reliable than forward announcement as a signal.

1.2 The CVSS and CWE Handoff Was a Success

The single most counterintuitive finding in the dataset is that CVSS v3.1 coverage in 2026 is comparable to coverage in 2020. In 2020, 91.9 percent of CVEs had a v3.1 vector. In 2023, 95.6 percent did. In 2026, 88.5 percent do. The decline from peak is real but small, and it conceals a structural transition: NVD-authored v3.1 fell from 86.6 percent to 28.9 percent in the same window, while CNA-authored v3.1 rose from 18.6 percent to 58.3 percent. The handoff was not announced, but it was effective.

Year	NVD-authored CVSS v3.1	CNA-authored CVSS v3.1	Any v3.1
2020	86.6%	18.6%	91.9%
2023	82.5%	53.2%	95.6%
2026	28.9%	58.3%	88.5%

CWE classification followed the same pattern. NVD's share declined from 78 percent to 16 percent across the decade while CNA-authored CWE coverage rose from 5 percent to 75 percent. Top-line CWE coverage held essentially flat. The data shows a successful, ecosystem-wide redistribution of CVSS and CWE work from a single source to a distributed source — not the loss of those enrichment types from the public record.

The practical implication is that any tool or analytical framework that treats CVSS and CWE as "NVD data" is operating against the wrong source for the majority of recent CVEs. The authoritative producer for these fields is the CNA container in the CVE Program record, not the NVD database.

1.3 CPE Is the Specific Unfilled Gap

Across the four enrichment field categories examined in this analysis, CPE applicability is the only one for which the data does not show another party absorbing the work after NVD's retreat.

Field	2022 coverage	2026 coverage	Where the work moved
CVSS v3.1	94.7%	88.5%	CNAs (then partial CISA)
CWE	85.0%	84.0%	CNAs
SSVC	51.5%	93.0%	CISA ADP
CPE applicability	95.5%	57.4%	No third-party producer

CNA-authored CPE coverage holds at 16 to 20 percent and shows no sustained upward trend. CISA's ADP program explicitly does not produce CPE applicability enrichment within its current scope. No commercial or open-source third party publishes CPE applicability at scale into the canonical CVE record. The CPE *dictionary* — the master list of product identifiers maintained by NIST — continues to be updated, but the work that connects a given CVE to specific CPE names with version ranges is the part that has lapsed.

The downstream consequence is operational. Without CPE applicability, a vulnerability scanner cannot automatically associate a CVE with a discovered software inventory item, regardless of whether the affected product is identifiable in prose. Records from approximately 60 CNAs that do not consistently write CPE — including several large vendors and the WordPress-plugin CNA ecosystem that now dominates the CVE corpus by volume — become invisible to NVD-dependent scanner configurations from the moment of publication.

1.4 CNAs Score CVSS Differently

A separate, methodologically distinct finding is that different CNAs produce systematically different CVSS distributions for technically comparable vulnerabilities. Across CNAs with more than 200 scored CVEs, distributions of CVSS subfield values differ by amounts that exceed plausible random variation. Some CNAs score Confidentiality, Integrity, and Availability impact as High in 70 to 80 percent of records; others score the same fields as Low in 60 to 80 percent of records. Some CNAs score Attack Vector as Local in 80 percent of records, reflecting hardware-local vulnerability populations; others score Attack Vector as Network in 80 percent, reflecting web-application populations.

Empirically, in 9,844 cases where two CNAs both scored the same vulnerability, one source chose Confidentiality=Low while the other chose Confidentiality=High in 6,680 cases — a methodological pattern, not measurement noise. The effect is that "CVSS 9.8" from one CNA and "CVSS 9.8" from another CNA describe materially different vulnerability populations. For downstream consumers treating CVSS as a uniform severity number across sources, this introduces a systematic bias that is not currently audited or corrected by any neutral third party.

The relevance of this finding to the present proposal is that **independent validation of CNA-supplied CVSS vectors** — validating the score against the documented vulnerability behavior, rather than re-deriving it from scratch — is a tractable AI task and a useful neutral service. Validation produces a second data point that downstream consumers can use to reconcile cross-CNA disagreements.

1.5 The CVE Corpus Has Shifted

CVE submission volume has approximately quadrupled over the past decade while median severity has dropped by about one full CVSS point. In 2016, median CVSS was 7.5 and 14 percent of records were Critical; in 2024, median CVSS was 6.5 and 10 percent of records were Critical. The composition has shifted

from a High-dominated distribution (2016 to 2020) toward a Medium-dominated distribution (2022 onward), driven primarily by the growth of WordPress-plugin and similar small-scope web-application CNAs.

For the present proposal, the consequence is that the typical 2026 CVE is a Medium-severity web-application bug rather than a High-severity enterprise software vulnerability. An enrichment pipeline must be efficient on the typical record, not optimized only for the high-severity tail. This shapes the tier distribution assumption used in the cost model: a small fraction of records will warrant intensive Complex-tier processing, and the majority should be processable in a Simple-tier path with low compute and low reviewer time.

2. Problem Definition

The empirical findings narrow the problem space to a tractable scope. The primary problem is the production of CPE applicability statements for CVEs where neither the filing CNA nor NVD has supplied one. The secondary problem is the production of independent validation of CNA-supplied CVSS vectors. Three additional candidate enrichment types — multi-language descriptions, automated reference tagging, and CVSS v4.0 authoring — are noted but treated as out of scope for this proposal and explicitly deferred.

2.1 Primary Problem: CPE Applicability Synthesis

For approximately 25,000 to 30,000 CVEs per year, a public, machine-readable CPE applicability statement does not exist. The required output is a structured statement consisting of one or more CPE 2.3 strings drawn from the existing NIST-maintained dictionary, scoped by version range using `versionStartIncluding`, `versionEndExcluding`, and equivalent qualifiers. The input is the CVE description and any vendor advisory or release-notes URLs included in the record's reference list.

The task is bounded in three useful ways. First, the CPE *dictionary* itself is maintained by an external party and is not in scope. Second, the task is a string-construction-and-version-range-extraction problem, not a free-form reasoning problem; the universe of plausible outputs is constrained by the dictionary. Third, output correctness is verifiable: a downstream reviewer or automated checker can determine whether a generated CPE string parses, whether it matches a dictionary entry, and whether the version range is consistent with the advisory text. These properties make the task a reasonable candidate for AI-assisted drafting with human verification.

2.2 Secondary Problem: Independent CVSS Validation

For CVEs that already carry a CNA-supplied CVSS vector — approximately 95 percent of new records — the problem is to read the advisory and validate the vector against the documented behavior. The task is structurally simpler than from-scratch scoring: the AI is checking an existing answer against a rubric rather than producing an answer from first principles. Validation outputs are categorical (confirmed, partially confirmed with rationale, disagreement with proposed alternate) and accompanied by per-metric reasoning that a human reviewer can verify.

For the residual 5 percent of records that lack a CVSS vector entirely, the task reverts to from-scratch scoring and is correspondingly more expensive in both AI compute and human verification. The cost model treats from-scratch scoring as a Complex-tier path used for a small fraction of total volume.

2.3 Out of Scope

The following enrichment types are explicitly out of scope for this proposal. They are noted because they appear in adjacent program proposals and because the pilot architecture would, with additional work, support their addition. They are not part of the pilot deliverable and are not part of the cost or capacity figures presented in this paper.

Multi-language CVE descriptions are out of scope. Translation has materially different cost economics, different review requirements, and different sponsorship structures than CPE and CVSS work. Automated reference tagging is out of scope as a standalone deliverable; if the pilot produces tagged URLs as a no-cost byproduct of advisory fetching, that output is incidental rather than committed. CVSS v4.0 authoring is out of scope; v4.0 adoption remains low across the ecosystem and the use cases that would benefit most are concentrated in industrial and operational technology domains that warrant their own scoping work.

2.4 What This Proposal Does Not Claim

This proposal does not claim to replace NVD. It does not claim that AI can produce enrichment of equivalent quality to expert human enrichment across all CVE records. It does not claim that a fully automated pipeline can operate without human review. It does not claim that the cost figures in this paper will hold at operational scale; the explicit purpose of the pilot is to test those figures empirically. It does not claim that all 25,000 to 30,000 unfilled CPE records can be enriched; some fraction will lack sufficient public advisory data to support any well-formed CPE applicability statement, and the pilot is designed to measure that fraction rather than assume it away.

3. Constraints and Boundaries

A realistic enrichment program operates within constraints that the headline framing of "AI restores NVD" tends to obscure. Three categories of constraint deserve explicit treatment: data availability, coverage expectations, and the program's relationship to the work NIST continues to do.

3.1 Data Availability

AI-assisted enrichment is bounded by the quality and accessibility of the underlying advisory data. The pipeline cannot infer product identification from a CVE description that only says "a vulnerability was found in a popular product." It cannot derive a CVSS Attack Vector from an advisory that omits any description of how the vulnerability is reached. Where vendor advisories sit behind paywalls, behind authenticated portals, or behind embargoes that have not lifted at publication time, the pipeline has access only to the CVE record's prose description and the publicly resolvable subset of reference URLs.

The empirical consequence is that some fraction of the 25,000 to 30,000 unfilled CPE records will have insufficient public data to support a defensible enrichment output. The pilot is structured to measure this fraction directly: every record processed during the pilot is classified into one of three confidence categories.

The first category, **high-confidence enrichment**, corresponds to records where public data — the CVE description plus publicly accessible advisories — supports a well-formed CPE applicability statement and a validated CVSS vector with explicit per-metric reasoning. These are the records the pipeline produces full enrichment for.

The second category, **partial-confidence enrichment**, corresponds to records where public data supports some enrichment fields but not others. Examples include records where product identification is clear but version-range scoping is ambiguous, or records where the advisory describes impact but not attack preconditions. The pipeline produces what is supportable, marks what is not, and surfaces the missing-data dependencies for downstream consumers to reason about.

The third category, **insufficient-data**, corresponds to records where public data does not support any defensible enrichment output. The pipeline records the record as inspected, logs the missing-data diagnosis, and abstains from publication. Abstention with diagnosis is itself a useful output: it tells downstream consumers that no enrichment is forthcoming and explains why, which is more informative than an absent record or an enrichment hallucinated from insufficient data.

The Phase 0 pilot is designed to measure the empirical distribution across these three categories. Pre-pilot estimates are not provided, because the data does not yet exist to bound them; pilot measurement is the load-bearing step.

3.2 Coverage Expectations

The objective is **majority coverage improvement, not completeness**. The data does not support a goal of restoring 95 percent CPE coverage to the full CVE corpus through AI-assisted enrichment alone, because the underlying advisory data does not always exist in sufficient form to permit it. A realistic objective is to add structured CPE applicability to the high-confidence subset, partial enrichment to the partial-confidence subset, and explicit non-coverage notices to the insufficient-data subset, with the relative sizes of the three subsets to be determined by pilot measurement.

This positioning is materially different from a guarantee of comprehensive enrichment. It commits to transparency about what was and was not produced, on a per-record basis, with explicit confidence labels. It does not commit to a fixed coverage percentage, because that percentage depends on properties of the input data that are not under the program's control.

3.3 Relationship to NIST and the Existing Producers

CSA's role under this proposal does not displace or compete with NIST. NIST continues to maintain the CPE dictionary, to enrich its priority set of CVEs (KEV, federal software, and EO 14028 critical software), and to serve as the canonical source of CVSS specifications and authoritative scoring guidance through FIRST.org. CISA continues to operate the ADP pipeline for SSVC enrichment. CNAs continue to author CVSS, CWE, and a minority of CPE entries.

The proposed CSA work fills the gap between these existing producers — specifically, the records outside NIST's priority set, where CNAs have not supplied CPE applicability or where independent CVSS validation would benefit downstream consumers. The operational difference between CSA's proposed output and NIST's current behavior in that gap is twofold: CSA produces partial enrichment with explicit confidence labels rather than deferring the record entirely, and CSA's enrichment carries explicit provenance metadata that allows downstream consumers to reason about its reliability. The two postures are complementary rather than overlapping.

4. Proposed Pipeline Architecture

The pipeline that the Phase 0 pilot will exercise consists of seven stages. The architecture is described here in its pilot configuration; production scaling considerations are out of scope for this paper.

4.1 Ingest

A daily scheduled task pulls the CVE Program's CVE list and filters for records in `Awaiting Analysis`, `Not Scheduled`, or equivalent unenriched states. Each record is normalized into a work item carrying the CVE identifier, the prose description, the publication and modification timestamps, the reference URL list, and any CNA-supplied CVSS, CWE, or CPE entries. A deduplication check excludes records that have been processed in prior pipeline runs. A priority-set filter defers to NIST for records identified as KEV, federal software, or EO 14028 critical software.

4.2 Classification

A lightweight router examines each work item and assigns it to a complexity tier — Simple, Moderate, or Complex — based on description length, reference count, the presence of a CNA-supplied CVSS vector, the presence of identifiable product names, and the breadth of affected versions implied by the advisory text. Tier assignment governs subsequent stages: it determines model selection, retrieval-loop depth, and review queue routing.

The pilot does not assume a fixed tier distribution. The pre-pilot planning estimate is approximately 55 percent Simple, 42 percent Moderate, and 3 percent Complex, derived from a 450-record analysis of the active backlog, but the pilot measures the empirical distribution against its 1,000-record sample and refines the estimate.

4.3 Drafting

A tier-appropriate model produces the per-record draft: proposed CPE strings with version-range scoping, a CVSS validation result against any CNA-supplied vector with explicit per-metric reasoning, a CWE classification, and a reviewer-facing rationale block summarizing the evidence used. Each draft carries a calibrated confidence score per artifact.

The drafting stage uses prompt caching to reuse a stable rubric prefix — the CVSS v3.1 specification, the CWE common-weakness mappings, the CPE format constraints, the output schema, and a small set of few-shot exemplars — across many calls at the cached-prefix price tier. The pre-pilot planning estimate for cache hit rate is 90 percent; the pilot measures the actual rate.

4.4 Conditional Retrieval

The drafting stage may invoke an agentic retrieval loop when the CVE description alone is insufficient: fetching vendor advisory URLs, retrieving release notes, and consulting other publicly accessible references when present. Retrieval is gated, not unconditional. The router decides whether retrieval is warranted based on description completeness, reference availability, and tier. This gating is the primary cost-control mechanism in the pipeline, because retrieval drives the largest variance in token consumption and is the dominant term in worst-case cost scenarios.

The pre-pilot planning estimate is that retrieval is invoked on approximately 60 percent of Moderate-tier records and approximately 90 percent of Complex-tier records, with each invocation adding roughly 30,000 to 80,000 input tokens to the per-CVE budget. The pilot measures the actual frequency and cost distribution.

4.5 Self-Critique

A validation pass reviews the draft's CPE scoping against the linked vendor advisories, checks the CVSS derivation against its own stated reasoning, and challenges the CWE mapping for edge cases. Any disagreement between the draft and the self-critique on a scored artifact automatically routes the record to the consensus review queue regardless of the draft's nominal confidence score.

4.6 Human Review

Records sort into one of two human review queues based on confidence and self-critique outcome. The single-reviewer queue handles high-confidence, clean drafts in a rubber-stamp pattern: the reviewer reads the draft, verifies the cited evidence, and approves or edits. The two-reviewer consensus queue handles low-confidence drafts, drafts flagged by self-critique, and drafts of high-blast-radius records (large version ranges, widely deployed products); two independent reviewers must agree before publication.

The pre-pilot estimate of weighted-average review time is 8 minutes per record, with a per-tier range of 2 to 4 minutes for Simple, 5 to 10 minutes for Moderate, and 10 to 20 minutes for Complex, and an 80/20 split between rubber-stamp and consensus paths. The pilot measures the actual distribution. Section 5 develops the human review model in more detail.

4.7 Publish

Approved enrichments emit to the open feed and to the CVE Program ADP container for each enriched record, conditional on ADP admission being granted. Each published artifact carries a provenance envelope: model identifier and version, prompt template version and commit hash, tier assignment, retrieval-loop outcome, self-critique outcome, reviewer attestation, and a cryptographic signature. The provenance envelope allows downstream consumers to audit production parameters and to file structured challenges through a redress process.

5. Token Consumption Model

The cost of the pipeline is dominated by AI compute, which is in turn dominated by token consumption at the drafting and retrieval stages. The pre-pilot model decomposes per-CVE token usage into stages, distributions, and conditional contributions, and identifies the cost-control mechanisms that bound the worst case.

5.1 Stage-Level Decomposition

For a single CVE, the stages contribute to token usage as follows.

Base enrichment corresponds to the drafting stage running against the CVE description and the cached rubric prefix. The fresh input is small (a few thousand tokens of CVE text plus modest few-shot context) and the cached input is the rubric (approximately 10,000 tokens). Output is the structured enrichment record (1,500 to 6,000 tokens depending on tier).

Retrieval is the dominant variable contribution and is conditional. When invoked, retrieval fetches one to several reference URLs, each contributing 5,000 to 30,000 input tokens depending on advisory length and rendering. The retrieval cost multiplier reflects both the per-invocation token addition and the frequency of invocation.

Self-critique runs the draft against a validation rubric. Input tokens are roughly the size of the draft plus a critique-specific instruction block (5,000 to 15,000 tokens). Output is a critique result (a few hundred tokens).

Output formatting is a final pass that ensures the published artifact conforms to the ADP container schema. It is small (a few thousand input tokens, a few hundred output tokens) and is included in the validation multiplier rather than as a separate term.

5.2 Distributional Estimates

The pilot will measure the empirical distribution of per-CVE token usage. Pre-pilot point estimates by tier are summarized below; the distributional bands are a rough planning guide and are bracketed by the pilot's measurement objective.

Tier	p50 fresh input tokens	p80 fresh input tokens	p95 fresh input tokens	Output tokens (typical)
Simple	~10,000	~40,000	~80,000	1,500
Moderate	~50,000	~110,000	~220,000	3,000
Complex	~150,000	~250,000	~500,000	6,000

Cached tokens are approximately constant at 10,000 per call across tiers, reflecting the stable rubric prefix.

The variance between p50 and p95 is dominated by retrieval. A Simple-tier record that does not invoke retrieval lands near the p50; a Simple-tier record that invokes retrieval and pulls a long advisory lands near the p95. The same dynamic at larger absolute scale governs Moderate and Complex tiers.

5.3 Retrieval Modeling

Retrieval is the principal source of cost variance in the pipeline. Pre-pilot planning treats retrieval as conditional and gated by tier:

Tier	Retrieval frequency (estimated)	Per-invocation token cost
Simple	~20%	~10,000
Moderate	~60%	~30,000
Complex	~90%	~80,000

These estimates derive from manual inspection of a sample of 50 backlog records and are presented as priors to be updated by pilot measurement. The retrieval frequency and per-invocation token cost are independently measurable during the pilot, and the calculator allows both to be set to user-supplied values.

5.4 Cost Control Mechanisms

The pipeline is designed with four explicit cost-control mechanisms. **Tiered routing** sends Simple records to inexpensive models, reserving frontier models for the small Complex tier; this is the primary multiplier on baseline cost. **Retrieval gating** invokes retrieval only when the router concludes the CVE description is insufficient; this bounds the worst-case input token term. **Prompt caching** keeps the stable rubric prefix at cached pricing, eliminating the per-call cost of the largest static input block. **Early exit** allows the pipeline

to abstain from producing an artifact when self-critique identifies low-confidence inputs that no reviewer time can rescue; this saves both compute and reviewer minutes on records that would have ended in `insufficient-data` anyway.

5.5 Pre-Pilot Status

All token estimates in this section are pre-pilot. The point of Phase 0 is to replace these estimates with measured distributions from the actual backlog population. The cost figures presented in Section 7 should be read as ranges that span plausible values around the pre-pilot estimates; readers are encouraged to use the calculator to evaluate cost under their own assumptions.

6. Human Review Model

The human review model assumes that AI drafts are verified rather than generated by humans. This is a structurally different workload than from-scratch enrichment and produces a structurally different time profile.

6.1 Per-Tier Time Estimates

Pre-pilot estimates of reviewer time per record are decomposed by tier. The estimates derive from a combination of bottom-up decomposition of review activities (read CVE description, verify CPE strings against dictionary, verify CVSS vector against advisory, verify CWE mapping, complete attestation form) and analogy to AI-in-radiology workflows where peer-reviewed studies have measured 3 to 10 times reductions from from-scratch reading.

Tier	Min/CVE (rubber-stamp)	Min/CVE (consensus, per reviewer)	Estimated share routed to consensus
Simple	2–4	7–10	10%
Moderate	5–10	12–18	25%
Complex	10–20	20–35	60%

The weighted-average time per CVE under the pre-pilot estimate is approximately 8 minutes, calculated against the 55/42/3 tier distribution and an estimated 80/20 rubber-stamp/consensus split. The 80/20 split is a planning anchor; the pilot measures the actual rate at which self-critique and confidence scoring route records to the consensus path.

6.2 Review Workflows

Two review patterns are supported in the pilot. The **rubber-stamp** workflow has a single reviewer read the AI draft, spot-check the cited evidence, and approve or edit; consensus is not required and the path optimizes for throughput on high-confidence records. The **consensus** workflow has two independent reviewers each produce an independent verification; agreement is required for publication, and disagreement triggers a third-reviewer adjudication. The consensus workflow is reserved for low-confidence, edge-case, or high-blast-radius records.

The cost of reviewer time scales linearly with the share of records routed to consensus. A 90/10 rubber-stamp/consensus split materially reduces per-record reviewer minutes; a 60/40 split materially increases them. The pilot measures the actual split that emerges from the pipeline's confidence scoring and self-critique.

6.3 Reviewer Capacity

Reviewer capacity is a function of review time per record, the records-per-week throughput target, and the per-reviewer weekly commitment. The calculator allows readers to vary all three. For the pilot, reviewer capacity is sized around the 1,000-CVE sample and a six-month timeline; the pilot does not require sustained capacity at operational scale.

7. Cost Model

The cost figures in this section are pre-pilot. They are presented as distributions rather than point estimates, with a deliberate emphasis on the range of uncertainty.

7.1 Cost Components

Per-CVE cost has two principal components: AI compute and human review time. AI compute scales with token usage at each stage, with the cached-input price applied to the rubric prefix and full input pricing applied to fresh input and output. Human review time scales with reviewer rate (volunteer at zero monetary cost, or compensated at a defined hourly rate) and with the weighted-average minutes per record. Infrastructure (feed publishing, ADP container submission, monitoring, redress workflow) is a smaller fixed-plus-marginal term that the calculator handles separately.

7.2 Distributional Estimates

For 1,000 CVEs at pilot scale and 51,000 CVEs at illustrative steady-state scale, the pre-pilot distributional cost estimates under tiered model routing with multi-provider averaging are summarized below. The p50 estimate is the modal expected cost; the p80 estimate is the expected cost under moderately adverse conditions (higher retrieval frequency, longer advisories, lower cache hit rate, or higher tier mix); the p95 estimate is the worst-case planning bound.

Scenario	Per-CVE compute (p50)	Per-CVE compute (p80)	Per-CVE compute (p95)
Pilot mode (lower cache, exploration overhead)	~$0.40	~$1.00	~$2.50
Steady-state mode (mature pipeline, high cache)	~$0.20	~$0.50	~$1.20

Annualized to the illustrative 51,000-CVE steady-state scope, the compute bands span approximately $10,000 (p50) to $60,000 (p95) under steady-state conditions. Under pilot conditions, the same scale would span approximately $20,000 (p50) to $130,000 (p95). These figures **deliberately do not collapse to a single number**, because the pre-pilot model cannot. The pilot measures the actual distribution.

7.3 Robustness Across Variance

The cost model is presented as robust if it remains within non-profit-feasible bounds across a 3 to 10 times variance in the underlying parameters. Under that test, the steady-state compute envelope at 10 times the p50 estimate ($100,000 per year for 51,000 records) remains tractable, suggesting the program's economic feasibility is not contingent on the most optimistic parameter assumptions. The pilot exists to determine whether the actual parameter values land in the lower or upper portion of the projected range.

7.4 Pilot vs Steady-State

Two operating modes are distinguished. **Pilot mode** runs at lower cache hit rates (caching is not yet warmed in the production sense), with exploration overhead from prompt iteration, calibration retries, and gold-standard comparison runs. **Steady-state mode** assumes a mature pipeline with stable prompt templates, warmed caches, and routine record processing. The calculator exposes both modes as toggles; the pilot's principal output is to measure the gap between them.

7.5 Distinction From Final Cost

The figures in this section are not final per-CVE prices. They are pre-pilot estimates whose primary function is to bound the order of magnitude. The pilot's measurement of actual distributional cost is the basis on which any subsequent operational program would be costed. Presenting these numbers as definitive would misrepresent the state of the evidence.

8. Phase 0 Pilot Proposal

The Phase 0 calibration pilot is the central deliverable of this paper. Its purpose is to validate or revise the pre-pilot estimates of token consumption, retrieval frequency, reviewer time, and accuracy, against the actual backlog population, before any operational program is committed.

8.1 Scope

The pilot processes 1,000 CVEs. The sample is drawn from the active backlog of records in `Awaiting Analysis` and `Not Scheduled` states, published between January 2025 and March 2026, and stratified to ensure coverage across the populations the program is designed to serve. Stratification dimensions include CNA (with explicit representation from the WordPress-plugin ecosystem, large-vendor advisories, open-source-project CNAs, and small commercial vendors), description length quartiles, reference count quartiles, and CNA-supplied-CVSS-presence flag.

The 1,000-record size is chosen to produce statistically informative measurements of distributional parameters (p50, p80, p95) without committing the program to operational scale. It is large enough to surface tier distribution, retrieval frequency, and accuracy at meaningful resolution; it is small enough to be processed within a six-month timeline by a calibration-focused team.

8.2 Objectives

The pilot has four objectives, each of which is measurable and produces a public artifact.

The first objective is to **measure token consumption distributions** at every stage of the pipeline. The pilot logs per-record fresh input, cached input, output, and stage-level breakdowns. Outputs include p50, p80, and p95 distributions per tier and per stage, plus an empirical retrieval frequency and per-invocation cost distribution.

The second objective is to **measure reviewer time** with the same distributional resolution. The pilot logs per-record reviewer minutes for both rubber-stamp and consensus paths, the rate at which records route to each path, and the rate of edit, reject, and abstain decisions. Outputs include p50, p80, and p95 review times per tier and the empirical confidence-routing distribution.

The third objective is to **measure accuracy** against expert ground truth. A 100-record subset of the pilot sample is enriched independently by three domain experts without AI assistance, using the existing NVD enrichment methodology as the baseline. AI pipeline output on these 100 records is compared to the

expert-produced ground truth on CPE correctness, CVSS validation accuracy, CWE classification accuracy, and reference tagging accuracy. Outputs include accuracy rates with confidence intervals and a typology of the disagreements.

The fourth objective is to **measure the data-availability distribution** described in Section 3.1. The pilot classifies every record into high-confidence, partial-confidence, or insufficient-data categories and reports the empirical distribution. This measurement directly bounds the achievable coverage of any subsequent operational program.

8.3 Outputs

The pilot produces six public outputs. **An open-source pipeline codebase** under Apache 2.0 covers all stages from ingest through ADP submission. **One thousand published enrichment records** are submitted through the CSA ADP container, conditional on ADP admission. **A gold-standard calibration dataset** of 100 records carries expert-produced enrichments and the comparison metrics. **A technical report** quantifies token consumption, reviewer time, accuracy, and data-availability distributions at p50, p80, and p95. **An updated public cost calculator** incorporates pilot-measured parameter distributions as a calibrated preset replacing the pre-pilot estimates. **A conference presentation** at a suitable security research venue summarizes the methodology and findings for community scrutiny.

8.4 Success Criteria

Pilot success is defined against measurable thresholds rather than commitment to a follow-on program. The pilot is considered successful when it has produced reliable distributional measurements of all four objective categories — even if those measurements indicate that AI-assisted enrichment performs worse than the pre-pilot estimates predicted. A pilot that produces well-measured negative results is as informative as a pilot that produces well-measured positive results.

The pre-pilot acceptance thresholds, against which pilot results will be compared, are:

- CPE accuracy at or above 85 percent on the gold-standard subset
- CVSS validation accuracy at or above 90 percent on records with CNA-supplied vectors
- p80 per-record reviewer time at or below 12 minutes
- p80 per-record compute cost at or below $1.50 in steady-state mode
- High-confidence-or-partial-confidence classification share at or above 70 percent of the pilot sample

Results at or above these thresholds support continued program development. Results below these thresholds inform the question of which program structures, if any, are appropriate.

8.5 Timeline

The pilot runs over approximately six months. The first two months focus on infrastructure build and the gold-standard calibration run on the 100-record expert-enriched subset. The middle two months process an additional 500 records with weekly calibration checks and prompt template refinement. The final two months process the remaining 400 records, complete the gold-standard accuracy comparison, and prepare the public technical report and presentation.

8.6 Budget

The Phase 0 pilot budget totals $85,000. The breakdown is summarized below. Detail and grant-application framing are provided in Appendix B.

Line item	Amount
Pipeline infrastructure and development	$35,000
Model API compute (1,000 CVEs at pilot-mode rates with calibration retries)	$5,000
Expert reviewer compensation (133 hours at $75/hr)	$10,000
Gold-standard calibration dataset (3 reviewers × 100 records × 30 min × $75/hr)	$11,250
Program coordination (0.25 FTE for 6 months)	$12,000
Publication, open-source release, and conference presentation	$11,750
Total	**$85,000**

The budget is structured for submission to the OpenAI Cybersecurity Grant Program and to peer foundation funders whose mandates align with cybersecurity public infrastructure.

9. Path to Sustainability

The pilot is the precondition for a sustainability conversation, not a substitute for it. A successful pilot enables the security community — including CSA, CISA, the CVE Program, foundation funders, and the broader vulnerability management ecosystem — to define what a sustainable enrichment program would require. This paper does not propose specific operating commitments, sponsor structures, or funding vehicles for that subsequent program.

The data shows that any sustainable program serving the gap CSA has identified has four prerequisites. It requires non-profit stewardship — a governance posture that does not embed commercial incentives in enrichment decisions, because the value of neutrality is precisely what differentiates a public baseline from commercial enrichment products. It requires ecosystem coordination — formal participation in the CVE Program's ADP framework, alignment with CISA's existing pipeline, and explicit non-overlap with NIST's priority set. It requires sustained reviewer capacity, whether volunteer, compensated, or hybrid; the pilot measures the per-record reviewer minutes that anchor capacity planning. And it requires open licensing — the data analysis underlying this paper indicates that the security community will only adopt a baseline that is free of commercial restrictions.

These prerequisites are jointly necessary. A program meeting all four becomes a candidate for community adoption. The shape of the eventual funding and operating model is a downstream question that the pilot's results, and the community conversations around them, will inform.

10. Risks and Open Questions

This section names the principal risks to the proposed work and the open questions whose answers materially affect the research agenda.

10.1 Data Sparsity

The largest risk is that the data-availability distribution proves more skewed toward `insufficient-data` than the pre-pilot estimate assumes. If a substantial majority of unfilled CPE records lack public advisory data sufficient to support any defensible CPE applicability statement, the addressable scope of the program is correspondingly smaller. The pilot directly measures this distribution; it is a primary outcome variable, not a peripheral one.

10.2 Model Variance

The cost figures and accuracy estimates assume that the AI models used in the pipeline produce stable outputs across runs and prompt iterations. In practice, model versions update, providers deprecate older models, and prompt template changes can produce step changes in output quality. The pilot uses pinned model versions and version-controlled prompts to bound this risk, but a long-term operating program would need ongoing calibration against drift. The technical report discusses observed variance during the pilot and recommends a re-calibration cadence.

10.3 Reviewer Scaling

The pilot's reviewer capacity is sized for 1,000 records over six months. A subsequent operational program covering 51,000 records per year requires roughly 50 times the reviewer-hours of the pilot, which would have to be sourced through some combination of volunteer commitment, compensated reviewer networks, or paid staff. The pilot does not test reviewer capacity at this scale; it measures per-record reviewer time, which is the input to any subsequent capacity model. The risk of reviewer burnout under sustained operational scale is real and is a question the community conversation following pilot conclusion will need to address.

10.4 ADP Acceptance

The proposal assumes that CSA can be accepted as an Authorized Data Publisher in the CVE Program, allowing pilot output to appear inside canonical CVE records. ADP admission is a governance decision that is not under CSA's unilateral control and may take longer than the pilot's six-month timeline. The pilot is structured to produce its outputs in ADP-compatible format and to publish to the open feed regardless of admission status; the ADP path is a desirable but not load-bearing component of the pilot.

10.5 NVD's Future State

NVD's operational trajectory over the next 24 months is uncertain. Plausible scenarios include further reduction in priority-set scope, partial restoration of broader enrichment if resources are restored, transition to a CISA-led pipeline replacement, or continuation at the current reduced scope. The pilot's findings are informative across all of these scenarios, because the data-availability and accuracy results characterize what AI-assisted enrichment can do regardless of whether NVD's capacity changes. The downstream sustainability conversation will need to be re-evaluated against whatever NVD's actual posture becomes.

10.6 Open Research Questions

Several questions surfaced by the empirical findings are not directly addressed by the pilot but are appropriate research targets for future work. **Why has CISA's ADP program not absorbed CPE applicability** is a coordination question whose answer shapes the long-term role for any third-party CPE work. **Whether systematic CNA scoring divergence reflects underlying methodological differences or downstream calibration error** is an empirical question that an exploitation-outcome study could answer. **What share of the deprecated CPE dictionary entries have replacement mappings** is a data-quality question whose answer determines how much historical debt a CPE-focused program would inherit. None of these are pilot deliverables; all of them are conversations the pilot's findings would inform.

11. Conclusion

The data shows that the CVE enrichment ecosystem has restructured rather than collapsed. Most of the work NVD historically performed has been absorbed by CNAs and by CISA. The remaining specific gap — CPE applicability for records outside the priority set — is real, measurable, and has not been picked up by any existing producer. AI-assisted, human-verified enrichment is a candidate response, but its feasibility at scale rests on parameters that have not been measured in the production population.

The appropriate next step is not to declare a program ready. The appropriate next step is to test the feasibility claim. A 1,000-record calibration pilot, with explicit distributional measurement of token consumption, reviewer time, accuracy, and data availability, produces the evidence needed for the security community to decide what comes next. A successful pilot enables a substantive conversation about a sustainable program. An unsuccessful pilot produces equally useful information about which fields and which record populations are not candidates for AI assistance.

Either outcome serves the community better than continued reliance on pre-pilot estimates. The Phase 0 pilot is necessary, sufficient, and modest in its claims. It is what the data supports.

Appendix A – Cost Calculator Specification

This appendix specifies the structure and behavior of the CSA NVD Enrichment Distribution Calculator, hosted at dashboard.csai.foundation/nvd-enrichment-calculator.html. The calculator complements this paper by allowing readers to substitute their own assumptions for any pre-pilot parameter and to evaluate the resulting distributional cost outcomes.

A.1 Design Principles

The calculator is built on five design principles. Outputs are **distributional, not single-valued**: every cost, capacity, and time figure is presented as p50, p80, and p95 rather than as a single average. Assumptions are **explicit and editable**: every input parameter is visible, labeled, and adjustable, with a documented source citation in the paper. The interface supports **dual-mode operation**: a Pilot Mode reflecting expected pilot-scale variance and lower cache hit rates, and a Steady-State Mode reflecting a mature pipeline. Sensitivity analysis is **first-class**: a 2×, 5×, and 10× stress toggle allows readers to test cost robustness against parameter variance. The interface is **explainable to non-technical reviewers**: every metric carries a one-line definition, every cost driver is decomposed into its contributing terms, and a summary panel answers the three questions reviewers most often ask (per-CVE cost range, total annual cost range, primary cost drivers).

A.2 Input Parameters

The calculator accepts the following user-adjustable parameters.

Volume and scope: annual CVE volume, NVD priority set coverage, computed gap CVEs per year, pilot sample size for pilot-mode runs.

Tier distribution: percentage share of records assigned to each of the Simple, Moderate, and Complex tiers, with the constraint that the three shares sum to 100 percent.

Token usage per tier: for each tier, the p50, p80, and p95 fresh input tokens, along with cached input tokens and output tokens. Defaults reflect the pre-pilot estimates from Section 5.2.

Retrieval modeling: for each tier, the retrieval invocation frequency (percentage of records) and the per-invocation token cost.

Model pricing (multi-provider): input, cached, and output prices per million tokens, separately configurable for each tier. The calculator ships with pricing tables for representative cost-effective, mid-range, and frontier models, plus cross-provider average rows.

Validation multiplier: an overhead factor applied to total token cost to account for self-critique and output formatting passes. Default is 1.4×.

Cache hit rate: the share of cached input tokens billed at the cached price. Pilot-mode default is 70 percent; steady-state default is 90 percent.

Human review parameters: per-tier rubber-stamp minutes (range), per-tier consensus minutes per reviewer (range), and per-tier consensus routing share. Default ranges and shares reflect Section 6.1.

Reviewer cost rate: dollars per hour, with $0 representing a pure volunteer model and configurable values representing compensated review.

A.3 Distributional Output

For each input scenario, the calculator computes:

- p50, p80, and p95 per-CVE compute cost
- p50, p80, and p95 per-CVE reviewer time
- p50, p80, and p95 per-CVE total cost (compute plus compensated review)
- p50, p80, and p95 annual total cost at the configured CVE volume
- Compute-vs-human cost split, by percentage of the p50 total cost
- Cost driver decomposition, attributing percentage of p50 compute cost to base inference, retrieval, and validation
- Tier contribution decomposition, attributing percentage of p50 compute cost to each tier

A.4 Mode Toggle

The calculator exposes a Mode toggle with two settings.

Pilot Mode uses a 70 percent cache hit rate, applies a 1.5× exploration overhead multiplier on token usage, and uses the upper end of the per-tier reviewer minute ranges. This mode is intended for budget planning around the Phase 0 pilot or any early-stage operating attempt.

Steady-State Mode uses a 90 percent cache hit rate, applies a 1.0× overhead multiplier (no exploration), and uses the central per-tier reviewer minute estimates. This mode is intended for projection of mature operational cost.

A.5 Sensitivity Analysis

A separate Sensitivity panel exposes a multiplier toggle (1×, 2×, 5×, 10×) that scales the p50 token usage, retrieval frequency, and reviewer minutes. The panel reports the corresponding p50 cost outcomes under each multiplier, allowing readers to evaluate program robustness against parameter excursions.

A.6 Visualization

Visualization is structured around three views. A **distributional cost curve** shows the spread between p50, p80, and p95 per-CVE cost as a horizontal range bar with a marker at p50. A **cost drivers breakdown** shows base inference, retrieval, and validation as a stacked bar at the p50 cost level. A **tier contribution breakdown** shows Simple, Moderate, and Complex contributions to total annual cost as a stacked bar.

A.7 No Single Magic Number

The calculator does not produce a single per-CVE cost figure as its primary output. The primary output is the p50-to-p95 range, displayed as a range with an explicit note that the p95 figure represents the planning bound under adverse parameter combinations. This presentation is deliberately at odds with the convention of single-number per-CVE pricing, because pre-pilot estimates do not justify single-number precision. After pilot conclusion, the calculator will incorporate measured distributions as a calibrated preset; the range presentation will tighten, but it will remain a range rather than collapse to a single value.

Appendix B — Phase 0 Pilot Grant Application Framing

This appendix presents the Phase 0 pilot in the format suitable for submission to the OpenAI Cybersecurity Grant Program and to peer foundation funders.

Project title: AI-Assisted CVE Enrichment Calibration Pilot — A 1,000-Vulnerability Proof of Concept

Submitting organization: Cloud Security Alliance / CSAI Foundation (501(c)(3))

Grant program: OpenAI Cybersecurity Grant Program

Requested amount: $85,000

Timeline: Q3 2026 to Q4 2026 (approximately 6 months)

B.1 Problem Statement

Approximately 25,000 to 30,000 CVEs per year are published with no machine-readable Common Platform Enumeration applicability statement from any source. This makes the affected vulnerability records invisible to automated vulnerability scanners, regardless of whether the underlying products are deployed in customer environments. The work that historically produced these statements — performed by NIST's National Vulnerability Database — was formally reduced in scope in April 2026. No free, public, neutral alternative currently exists.

This is not a question of replacing NVD. Most of NVD's historical work has been absorbed by CVE Numbering Authorities and by CISA's Authorized Data Publisher program. The remaining specific gap — CPE applicability for records outside the federal priority set — is narrow but has not been picked up by any other producer.

B.2 Hypothesis

An AI-assisted, human-verified enrichment pipeline can produce CPE applicability statements and validated CVSS vectors at distributional cost per record below $1.50, with reviewer time per record below 12 minutes at the p80, and accuracy at or above 85 percent against expert ground truth. This hypothesis has not been validated at operational scale and cannot be derived from first principles; it can only be tested.

B.3 Methodology

The pilot tests the hypothesis against 1,000 CVE records drawn from the active backlog. The records are stratified by CNA, description length, reference count, and CNA-supplied-CVSS presence. For each record, the pipeline produces an AI-drafted enrichment, applies a self-critique pass, and routes the result to either a single-reviewer rubber-stamp queue or a two-reviewer consensus queue based on confidence. A 100-record gold-standard subset is independently enriched by three domain experts without AI assistance and serves as the accuracy baseline.

Token consumption, reviewer minutes, retrieval frequency, and accuracy metrics are logged per record and reported as p50, p80, and p95 distributions in the public technical report.

B.4 Budget Detail

Line item	Amount	Detail
Pipeline infrastructure and development	$35,000	Ingest, classification, drafting, retrieval, self-critique, review interface, ADP container submission, observability
Model API compute	$5,000	$5/CVE conservative budget at pilot-mode rates with calibration retries (10× the projected steady-state rate)
Expert reviewer compensation	$10,000	133 hours at $75/hr for 1,000-record review (8 min/record weighted average)
Gold-standard calibration dataset	$11,250	3 reviewers × 100 records × 30 min × $75/hr
Program coordination	$12,000	0.25 FTE × 6 months at modest non-profit rate
Publication and dissemination	$11,750	Open-source release infrastructure, technical report production, conference travel and submission
Total	**$85,000**	

The model API estimate is intentionally conservative — approximately 10× the projected steady-state rate — to absorb pilot-mode exploration costs, prompt template iteration, and re-runs. Unused model budget rolls into the open-source release line.

B.5 Public Outputs

All pilot outputs are open. The pipeline codebase is released under Apache 2.0. The 1,000 published enrichment records are submitted via the CSA ADP container under CC-BY-4.0 license. The 100-record gold-standard dataset is released under CC-BY-4.0. The technical report is published under CC-BY-4.0 and submitted to a suitable security research venue. The cost calculator at dashboard.csai.foundation is updated with measured pilot distributions as a calibrated preset, replacing the pre-pilot estimates.

B.6 Alignment with OpenAI Cybersecurity Grant Objectives

The proposal advances the OpenAI Cybersecurity Grant Program's interest in using AI to benefit cybersecurity defenders. The pipeline restores access to machine-actionable vulnerability data for the organizations — small enterprises, academic institutions, government agencies, non-profit security teams, and individual practitioners — who cannot afford commercial enrichment subscriptions and currently have no reliable access to CPE applicability and CVSS validation for the majority of published CVEs. The program uses multiple AI providers in a multi-vendor architecture that demonstrates practical interoperability rather than vendor lock-in. All outputs are open-licensed with full provenance and are available to the security community without commercial restriction. Independent third-party validation of the methodology is built into the design through the gold-standard calibration dataset and the open release of pipeline code.

B.7 Post-Grant Path

The pilot is explicitly designed to produce evidence rather than to commit a follow-on program. Successful pilot results enable subsequent community conversations about sustainable operating structures, in which CSA participates as a research and coordination entity rather than as a unilateral operator. Unsuccessful pilot results are equally informative for the security industry's understanding of where AI assistance is and is not productive in vulnerability enrichment.

Appendix C – Glossary

ADP (Authorized Data Publisher). A party authorized by the CVE Program to contribute enrichment metadata to CVE records via an ADP container in the CVE JSON 5.x schema.

AICM (AI Controls Matrix). CSA's controls framework for AI systems, extending the Cloud Controls Matrix with AI-specific control objectives.

CNA (CVE Numbering Authority). An organization authorized by MITRE to assign CVE identifiers and optionally to provide enrichment metadata including CVSS vectors and CWE classifications.

CPE (Common Platform Enumeration). A structured naming scheme for identifying software, hardware, and operating systems. Vulnerability scanners use CPE strings to match discovered assets to published CVEs.

CSAI Foundation. The Cloud Security Alliance AI Foundation, a 501(c)(3) non-profit focused on AI security and safety research.

CVSS (Common Vulnerability Scoring System). A standardized scoring framework expressing vulnerability severity. Versions 3.1 and 4.0 are in active use as of 2026.

CWE (Common Weakness Enumeration). A community-maintained taxonomy of software weakness categories.

Era 7. CSA's designation for the current NVD operational period (2024 to present), characterized by declining coverage across all previously protected enrichment fields.

KEV (Known Exploited Vulnerabilities). CISA's catalog of CVEs with confirmed active exploitation.

NVD (National Vulnerability Database). The NIST-operated database that has enriched MITRE CVE entries with CPE, CVSS, CWE, and reference metadata since 2000.

Peak NVD. CSA's designation for the 2020 to 2021 operational era when NVD produced comprehensive enrichment on 85 to 95 percent of CVEs – the two-year window against which most modern security tools were calibrated.

Silent Triage (Era 6). CSA's designation for the 2022 to 2023 period when NVD silently cut CVSS v2 authoring and reduced CWE and Spanish coverage two years before the visible public backlog crisis.

SSVC (Stakeholder-Specific Vulnerability Categorization). A CISA-maintained framework for prioritizing vulnerability response based on structured decision points, supplied via the ADP program.

References

[1] NIST. "NIST Updates NVD Operations to Address Record CVE Growth." NIST News, April 2026.

[2] Help Net Security. "NIST admits defeat on NVD backlog, will enrich only highest-risk CVEs going forward." Help Net Security, April 16, 2026.

[3] The Hacker News. "NIST Limits CVE Enrichment After 263% Surge in Vulnerability Submissions." The Hacker News, April 2026.

[4] Infosecurity Magazine. "NIST Drops NVD Enrichment for Pre-March 2026 Vulnerabilities." Infosecurity Magazine, 2026.

[5] Socket.dev. "NIST Officially Stops Enriching Most CVEs as Vulnerability Volume Surges." Socket Blog, 2026.

[6] SecureWorld. "The NVD Course Correction: Navigating NIST's Strategic Pivot for 2026." SecureWorld, 2026.

[7] VulnCheck. "Enhancing Access to NIST NVD Data: Introducing CPE Enrichment in VulnCheck NVD++." VulnCheck Blog, 2024.

[8] Cloud Security Alliance. "A Vulnerability Management Crisis: The Issues with CVE." CSA Blog, November 21, 2024.

[9] FIRST.org. "Common Vulnerability Scoring System (CVSS)." Forum of Incident Response and Security Teams, 2026.

[10] FIRST.org. "Exploit Prediction Scoring System (EPSS)." Forum of Incident Response and Security Teams, 2026.

[11] SecurityWeek. "NIST Still Struggling to Clear Vulnerability Submissions Backlog in NVD." SecurityWeek, 2026.

[12] Cybersecurity Dive. "NIST limits vulnerability analysis as CVE backlog swells." Cybersecurity Dive, 2026.

[13] Jerry Gamblin. "2025 CVE Data Review." JerryGamblin.com, January 1, 2026.

[14] CISA. "Stakeholder-Specific Vulnerability Categorization (SSVC)." Cybersecurity and Infrastructure Security Agency, 2026.

[15] CISA. "Known Exploited Vulnerabilities Catalog." Cybersecurity and Infrastructure Security Agency, continuously updated.

[16] CISA. "Vulnrichment Project." CISA on GitHub, continuously updated.

[17] MITRE. "CVE List – Common Vulnerabilities and Exposures." The MITRE Corporation, continuously updated.

[18] MITRE. "Common Weakness Enumeration (CWE)." The MITRE Corporation, continuously updated.

[19] MITRE. "CVE JSON 5 Schema and ADP Container Specification." The CVE Project on GitHub, 2026.

[20] NIST. "Common Platform Enumeration (CPE) Dictionary." National Vulnerability Database, continuously updated.

[21] Anthropic. "Prompt caching with Claude." Anthropic Documentation, 2026.

[22] OpenAI. "Cybersecurity Grant Program." OpenAI, 2026.

[23] Cloud Security Alliance. "AI Controls Matrix (AICM)." CSA Research, 2026.

[24] Cloud Security Alliance. "STAR Program." CSA, continuously updated.

[25] Open Source Vulnerabilities. "OSV Schema Specification." Open Source Security Foundation, 2026.