

Research Article

Certificate-Gated Clearance in Change-Impact Analysis: A Fail-Closed Architecture

Ankur Garg¹, Jeffrey Esposito¹

1. Rockland Trust, United States

Organizations that operate under regulatory or policy oversight must decide, for every change to a governing instrument, which of their governed artifacts are affected. Most such decisions are negative: the change does not apply. Automating negative decisions is attractive because they dominate review volume, but an incorrect negative decision as a *false clearance* removes an obligation from review without anyone noticing. Prediction-based methods score or explain a negative decision after making it; they do not establish that the decision is admissible.

We present RCII², a knowledge-based decision architecture in which a non-impact clearance is permitted only when the system constructs a machine-verifiable certificate that an independent verifier accepts. Clearance grounds are partitioned into *certifying* grounds, which are logically sufficient and checkable against frozen source records, and *suggestive* grounds, which may inform review but can never authorize clearance. Every certificate is rule-bound, version-bound, hash-bound, and subject to a conflict veto, complete dependency enumeration, and an authority-bounded evidence path. Cases without an accepted certificate fail closed and escalate.

On a frozen synthetic corpus of 2,020 change artifact decisions, RCII² cleared 850 cases with zero observed false clearances (false-clearance rate 0.0000, 95% CI [0.0000, 0.0083]), non-impact precision 1.0000, and safe coverage 0.4208. All 850 certificates were accepted by the independent verifier, and all 600 tampered certificates across twelve mutation classes were rejected. A supervised baseline clearing the identical number of cases produced a false-clearance rate of 0.1217.

Correspondence: papers@team.qeios.com — Qeios will forward to the authors

1. Introduction

1.1. *The negative decision problem*

An organization subject to external regulation maintains a body of *governed artifacts*: policies, procedures, controls, system configurations, and governed records. When a governing instrument changes, each artifact must be assessed for impact. In practice, the overwhelming majority of assessments conclude that the change does not apply.

This asymmetry creates an automation opportunity and a risk. The opportunity is that negative decisions dominate volume, so automating them removes most of the work. The risk is asymmetric in the opposite direction. A false positive that results in escalating an unaffected artifact costs reviewer time, and the cost is bounded and linear. A false clearance automatically clearing an artifact that was in fact affected silently removes an obligation from the review queue. Nobody looks at it again. The error surfaces later, if at all, through examination findings or an incident.

Because the two mistakes have very different consequences, overall accuracy is not enough to judge the system. It must be evaluated on the false-clearance rate at a stated coverage, and it must make the negative decision *defensible* to a party who was not involved in producing it.

1.2. *Why prediction is not enough*

Most existing approaches predict an outcome. A classifier, retrieval model, or rule set assigns a score or label, and another system records or explains the result. That creates three problems for negative decisions.

First, a confidence score reflects the model's internal state, not the underlying evidence. It may not be preserved, and a reviewer or examiner cannot verify it against a source record.

Second, explanations come after the decision. Explainable AI (XAI) methods describe why a model produced an output, but they do not prove that the decision was permitted by the evidence^{[1][2][3]}. The explanation also comes from the same system, or from a system built around it.

Third, prediction systems usually must produce an answer. Every case receives a label whether or not the evidence is strong enough to support one. Selective classification allows abstention, but conventional approaches usually base that decision on a confidence score or learned selection function rather than independently verifiable case evidence^{[4][5][6]}.

1.3. *The proposed principle*

RCII² uses a different standard:

A case can be cleared as non-impacting only when the system produces a compact, self-contained certificate that an independent verifier can check against frozen source records and public rule definitions.

The system does not make the decision first and explain it afterward. It must produce verifiable evidence before the clearance is allowed. When it cannot, the case is escalated. Escalation is the safe default, not a system failure.

This approach draws on certifying algorithms and proof-carrying code, where one component produces a witness and a separate component verifies it^{[21][8]}. The key difference here is that only some change-impact cases can produce a valid witness, so the ability to construct one becomes the rule for when the system may decide and when it must abstain.

1.4. *Contributions*

1. A certificate-gated architecture that allows a non-impact clearance only after independent verification, rather than relying on model confidence.
2. A clear separation between certifying and suggestive grounds, so suggestive evidence can support review but can never authorize clearance.
3. A defined certificate format and a fifteen-rule verifier that works without access to the decision engine, model state, or confidence scores.
4. A controlled evaluation on a frozen synthetic corpus of 2,020 decisions, comparing RCII² with ungated rules and a supervised baseline at matched coverage, including a controlled distribution-shift test.
5. A tamper-resistance test covering twelve mutation classes.
6. A reproducible laboratory package containing the frozen corpus, rules, decision ledgers, certificates, and independent metric reconstruction.

The contribution is a fail-closed, independently verified, proof-carrying system for non-impact decisions.

1.5. Scope of the claim

This study is a controlled proof of concept using a synthetic corpus. It shows that the mechanism works as designed and can be reproduced, but it does not yet prove production performance or guarantee zero future false clearances.

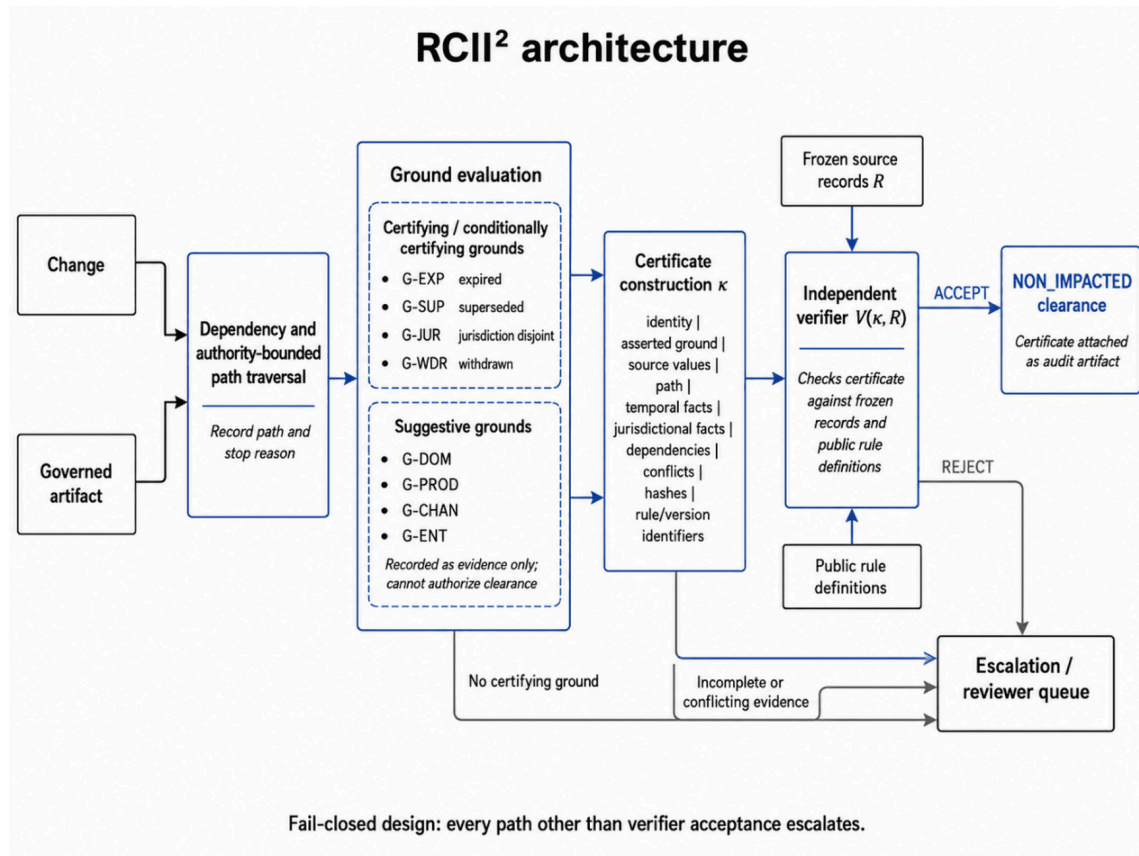


Figure 1. RCII² architecture. A change and a candidate governed artifact enter ground evaluation. Certifying grounds may trigger certificate construction; suggestive grounds are recorded as evidence only. Constructed certificates pass to an independent verifier. Clearance is emitted only on verifier acceptance; all other paths escalate.

2. Related Work

We compare RCII² with seven related areas of research. The key difference is that other methods produce or explain a decision, while RCII² requires independent verification before the decision is allowed.

2.1. Change-impact analysis

Change-impact analysis (CIA) began in software engineering as a way to identify which parts of a system may be affected by a change, often using dependency graphs, program slicing, or traceability links^{[9][10][11]}. Regulatory and policy change management uses related formal, rule-based, and process-oriented methods for governed documents and obligations^{[12][13][14][15]}. These approaches are good at finding what might be affected, but they do not define how to prove that something is not affected or require independently verifiable evidence for that conclusion.

2.2. Prediction-only classification

Supervised classifiers use learned features to assign labels in legal, compliance-adjacent, and impact-analysis tasks^{[16][17][18]}. They depend on labeled training data, may perform poorly when the data changes, and do not inherently produce evidence that an independent reviewer can verify. Arms D, E, and F represent this type of approach.

2.3. Explainable AI (XAI)

Explainable AI methods such as feature attribution, surrogate models, concept-based explanations, and counterfactual explanations help to describe why a model produced a result^{[1][2][3]}. Recent KBS work continues to improve explainability through ensemble XAI, concept bottlenecks, and ontology-enriched causal reasoning^{[19][20][21]}. These methods may make a decision easier to understand, but they do not determine whether the system was allowed to make it. RCII² certificates are different because they must satisfy defined evidence requirements before the clearance is released.

2.4. Rules engines and knowledge-based systems

Production rule systems and business rules engines evaluate conditions and trigger actions^{[22][23]}. Current KBS research also continues to develop formal rule extraction and logical-reasoning structures, including three-way rule extraction from Boolean formal contexts^[24]. RCII² also uses rules, but a conventional rules engine may allow any satisfied condition to authorize a decision. Our Arm A represents that approach because any negative ground can end the analysis. RCII² differs by separating grounds according to whether they are logically sufficient and by requiring independent verification before a clearance is allowed.

2.5. Provenance and audit logging

Provenance systems record how a result was produced^{[25][26]}, and W3C PROV provides a standard model for representing entities, activities, agents, and their derivations^{[27][28]}. Recent KBS work continues to use knowledge-based structures to support reliable provenance and intent capture^{[29][30]}. Provenance records help reconstruct what happened, but they do not necessarily control whether the decision is allowed. RCII² certificates are created before clearance and must be approved by an independent verifier before the clearance can occur.

2.6. Selective classification and abstention

Selective classification allows a model to abstain when its confidence is too low or when a learned selection function declines the case^{[31][4][5][6]}. Conformal prediction also supports calibrated prediction sets and abstention with formal coverage guarantees^{[32][33][34]}. Recent KBS work illustrates a parallel emphasis on uncertainty-aware and open-set reasoning^[35]. These are the closest related methods, but the trigger is different. Conventional selective methods abstain based on a score, set size, or learned selection rule, while RCII² abstains when it cannot construct verifiable evidence. A score depends on how well the model separates cases. A certificate depends on whether the required evidence actually exists.

2.7. Proof-carrying code, certifying algorithms, and formal verification

Proof-carrying code requires code to include proof that it follows a safety policy, which the receiving system checks independently^[8]. Certifying algorithms produce a result together with evidence that a simpler checker can verify^{[36][7][37]}. Formal methods use mathematical reasoning to establish whether a system satisfies specified properties, while assurance cases organize structured evidence and argument for human review^{[38][39][40]}.

RCII² uses the same producer-and-checker model. The difference is that only some change-impact cases can produce a valid certificate. When no certificate can be built, the system abstains and escalates the case. Recent work has begun to move beyond confidence-only abstention: Khosravani^[41] predicts only when all hypotheses consistent with stated constraints force agreement, and Xu^[42] separates verdict generation from commitment authorization when evidence is structurally conflicting. RCII² is therefore not claimed to originate evidence-structure-based abstention in general. To our knowledge, its narrower contribution is the combination of certificate constructibility as the clearance gate, independent

verification against frozen records, and a structurally enforced distinction between certifying and suggestive evidence for negative change-impact determinations.

3. Problem Definition and Research Questions

3.1. Definitions

Let C denote a change to a governing instrument, A a governed artifact version, and R a frozen corpus of source records.

Governed artifact. A versioned organizational object subject to a governing instrument: policy, procedure, control, system configuration, or governed record. Each version carries temporal validity, jurisdictional scope, authority linkage, and applicability attributes across product, channel, entity, customer, and transaction dimensions.

Change-impact decision. A determination, for the pair (C, A) , of one of three states: - **IMPACTED**: the change alters what A requires; - **NON_IMPACTED**: the change cannot alter A ; - **INDETERMINATE**: neither can be established.

INDETERMINATE is the default. The other two require justification.

Clearance ground. A defined condition, with specific evidence requirements, that supports a finding of non-impact.

Certifying ground. A condition is certifying only when it logically proves non-impact and can be checked directly from structured source data without interpreting free text. Both requirements must be met.

Suggestive ground. A condition is suggestive when it may be relevant but does not fully prove non-impact or cannot be verified directly from structured data. It may help prioritize review, but it can never authorize clearance.

Certificate. A structured record that identifies the case, states the clearance ground, cites the supporting source data, records the relevant path, scope, dependencies, conflicts, and dates, and includes the hashes, rules, and version details needed for independent verification.

Verifier. A function $V(\kappa, R) \rightarrow \{\text{ACCEPT}, \text{REJECT}\}$, executing without access to engine state, model state, or confidence scores, which re-evaluates the terminating predicate directly against R .

For example, if a certificate claims that an artifact is non-impacting because it expired before the change took effect, the verifier checks the artifact's end date and the change's effective date in the frozen records.

It accepts the certificate only if those dates support the claim.

Clearance. A finding of NON_IMPACTED. Clearance is allowed only when the independent verifier accepts the certificate; $V(\kappa, R) = \text{ACCEPT}$.

False clearance. A case the system clears as non-impacting even though the gold label is IMPACTED.

$$\text{FCR} = \frac{\text{false clearances}}{\text{all gold-IMPACTED cases}}$$

For example, if 5 of 100 truly impacted cases are incorrectly cleared, the false-clearance rate is 5%.

Non-impact precision. The share of cleared cases that are truly NON_IMPACTED.

$$\text{Non-impact precision} = \frac{\text{cleared cases with gold label NON_IMPACTED}}{\text{all cleared cases}}$$

For example, if 90 of 100 cleared cases are truly non-impacting, non-impact precision is 90%.

Safe coverage. The share of all decisions that are correctly cleared as NON_IMPACTED and supported by an accepted certificate.

$$\text{Safe coverage} = \frac{\text{cleared} \wedge \text{gold NON_IMPACTED} \wedge V = \text{ACCEPT}}{\text{all decisions}}$$

All three conditions must be met. A correct non-impact decision without an accepted certificate does not count as a safe clearance.

Escalation rate. The share of all cases sent for further review because the system could not safely clear them.

$$\text{Escalation rate} = \frac{\text{INDETERMINATE cases}}{\text{all decisions}}$$

3.2. Central hypothesis

H. Certificate-gated non-impact clearance can prevent false clearances while still clearing a useful share of cases.

False-clearance rate (FCR) and coverage must be reported together. A system can achieve zero false clearances by clearing nothing, and it can achieve full coverage by clearing everything. Neither result is useful on its own.

3.3. Research questions

- **RQ1.** Can certificate gating prevent false non-impact clearances without supervised training data? • **RQ2.** Can it still clear a useful share of cases rather than escalating everything? • **RQ3.** Can independent verification reject altered, incomplete, conflicting, or fabricated evidence? • **RQ4.** At the same coverage, does certificate gating produce fewer false clearances than prediction-only methods? • **RQ5.** Does certificate-gated clearance remain stable under a controlled distribution shift? • **RQ6.** Which parts of the architecture are essential to the result?

3.4. Falsification conditions

Three failure conditions were defined before the evaluation:

- **F1.** The false-clearance rate exceeds 5%, or the upper bound of the 95% confidence interval exceeds 10%.
- **F2.** Safe coverage falls below 5%.
- **F3.** Certificate acceptance falls below 98%, or tamper rejection falls below 95%.

F2 prevents a meaningless result. A system can achieve a false-clearance rate of zero simply by clearing nothing, but such a system has no practical value.

4. RCII² Method

4.1. Decision procedure

RCII² uses the following deterministic decision procedure:

1. Traverse the dependency graph from C toward A along authority-bounded edges. Record the path and stop reason.
2. Evaluate certifying grounds in fixed deterministic order. On the first satisfied ground:
 1. verify machine-decidability;
 2. verify complete dependency enumeration;
 3. construct certificate κ ;
 4. submit κ to the independent verifier V ;
 5. if $V(\kappa, R) = ACCEPT$, emit **NON_IMPACTED**; otherwise, escalate.

3. If no certifying ground produces an accepted certificate, evaluate suggestive grounds and record them as evidence. Never terminate on a suggestive ground.
4. Emit **INDETERMINATE** with the unresolved predicate set.

No stochastic component appears on the decision path.

4.2. Clearance grounds

The frozen ground taxonomy contains eight grounds grouped into three classes.

Ground	Class	Condition
G-EXP	CERTIFYING	The artifact version expired before the change took effect.
G-SUP	CERTIFYING	The artifact version was replaced by a later version.
G-JUR	CERTIFYING	The artifact and the change apply to different jurisdictions.
G-WDR	CONDITIONALLY CERTIFYING	The governing instrument was withdrawn.
G-DOM	SUGGESTIVE	The business domains do not appear to match.
G-PROD	SUGGESTIVE	The product scopes do not overlap.
G-CHAN	SUGGESTIVE	The channel scopes do not overlap, excluding cases marked ALL.
G-ENT	SUGGESTIVE	The entity scopes do not overlap.

Table 1. Clearance-ground taxonomy and authorization status.

The certifying grounds can be checked directly from a small number of structured fields. For example, a reviewer can verify that an artifact expired on 26 January 2023 and that the change took effect on 24 September 2024. A statement such as “the business domains do not appear to overlap” requires interpretation, so it is only suggestive.

G-WDR is conditionally certifying because withdrawal does not always end an obligation. A savings clause may preserve some requirements, so that condition must be checked and recorded when the certificate is created.

4.3. Certificate structure

Each certificate contains:

- **Case identity:** the change, artifact, artifact version, certificate, and run identifiers.
- **Asserted ground:** the ground used to support clearance and its classification.
- **Rule binding:** the rule, rule version, and certificate schema version.
- **Predicate and evidence:** the condition being tested and the source values used to support it.
- **Authority-bounded path:** the approved dependency path, including its depth and stop reason.
- **Temporal facts:** the change effective date and the artifact's validity period.
- **Jurisdictional facts:** the jurisdictions that apply to the change and the artifact.
- **Applicability facts:** the product, channel, entity, customer, and transaction dimensions evaluated.
- **Dependency enumeration:** the number of dependencies reviewed and whether the set is complete.
- **Suggestive evidence:** relevant grounds that were recorded but did not authorize clearance.
- **Conflicting evidence:** any records that contradict the claimed ground.
- **Integrity information:** the certificate content hash and the corpus manifest hash.
- **Disposition:** the final decision.

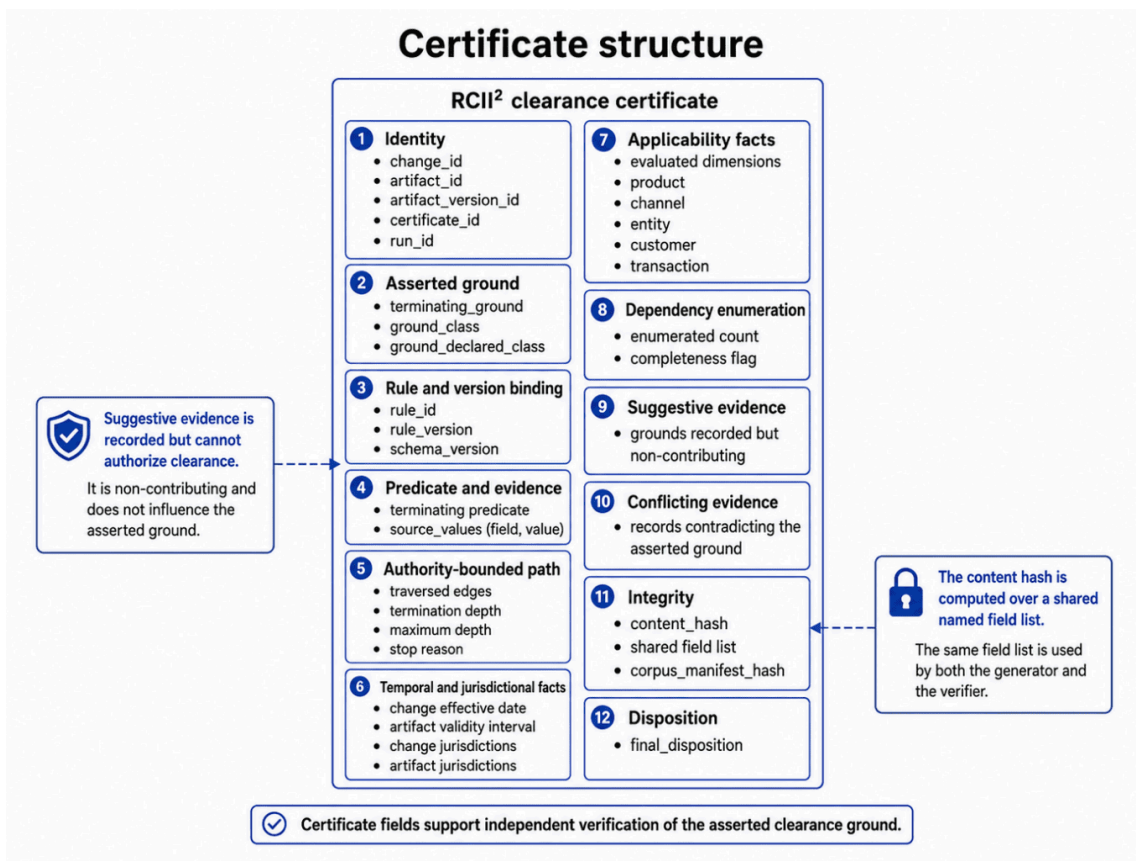


Figure 2. Certificate structure. The figure groups certificate fields by function, including identity, rule and version binding, evidence, dependency path, completeness, conflicts, integrity, and final disposition.

Two details are especially important. Suggestive evidence is recorded but clearly marked as non-authorizing, so the verifier can confirm that it did not produce the clearance. The content hash is also calculated over a fixed field list shared by the certificate generator and verifier, so any difference between the two is detected rather than ignored.

4.4. Independent verification

The verifier receives only the certificate, the frozen corpus, and the public rule definitions. It does not use the decision engine, supervised models, or confidence scores.

It applies fifteen checks:

Rule	Check
V-01	The certificate schema version is valid.
V-02	The certificate identifier has the correct format.
V-03	The run identifier matches the expected run.
V-04	The certificate content hash is valid.
V-05	The asserted ground appears in the frozen taxonomy.
V-06	The ground is classified as CERTIFYING or CONDITIONALLY CERTIFYING.
V-07	The stated disposition is NON_IMPACTED.
V-08	The predicate is present and well formed.
V-09	The cited source values are present.
V-10	Dependency enumeration is complete.
V-11	The authority-bounded path is valid.
V-12	The temporal facts are consistent.
V-13	The jurisdictional facts are consistent.
V-14	The corpus manifest hash matches.
V-15	No conflicting evidence shows that the case is IMPACTED.

Table 2. Verification Rules

The central principle is simple: the certificate generator cannot approve its own result.

V-06 prevents suggestive grounds from authorizing clearance. V-15 blocks clearance when conflicting evidence exists. V-10 requires all relevant dependencies to be included. V-11 requires the evidence path to use links that are authoritative, valid at the relevant time, and applicable to the case.

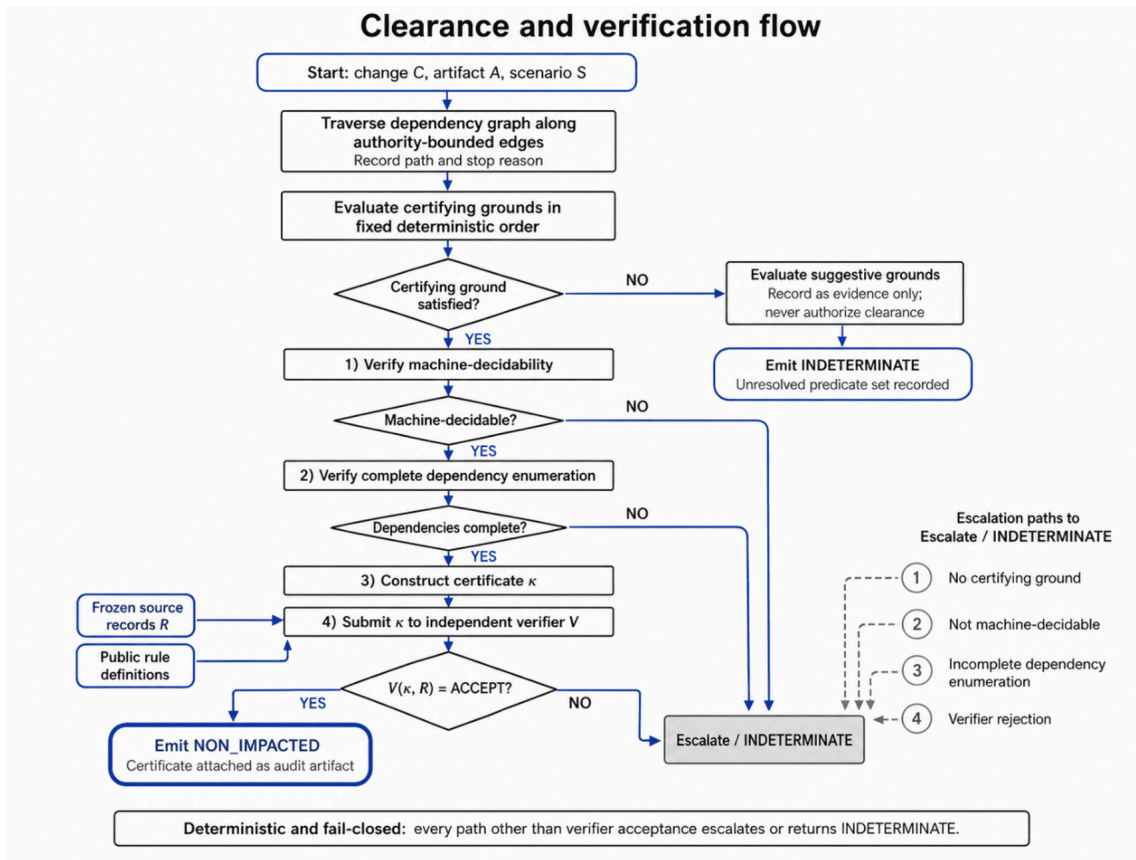


Figure 3. Clearance and verification flow. Deterministic and fail-closed: any path other than verifier acceptance results in INDETERMINATE or escalation.

4.5. Fail-closed escalation

Every case is escalated unless the verifier accepts the certificate. This remains true even when a certifying ground is found. No case can be cleared without an accepted certificate, and no confidence score can override that requirement.

A certificate defect therefore creates additional review work rather than an unnoticed false clearance.

4.6. Where learned models remain useful

Learned models can still play an important role in RCII². They may rank escalated cases, suggest possible grounds for a rule author to formalize, or retrieve related artifacts.

The limit: a model score may support the process, but it cannot authorize the final clearance because a score is not independently verifiable evidence.

5. Experimental Design

5.1. Synthetic corpus

The corpus models governed artifacts such as policies, procedures, controls, systems, and governed records. Each case includes structured information about version, time, authority, jurisdiction, product, channel, customer, and transaction scope.

The corpus is entirely synthetic. No Rockland Trust documents, customer information, or production data were used.

Six partitions were created:

Partition	Cases	Purpose
dev	600	Rule development and supervised training
val	400	Model selection and calibration
final_r	2,020	Primary evaluation
final_a	600	Adversarial evaluation
final_s	400	Distribution-shift evaluation
final_u	300	Unseen-ground evaluation

Table 3. Synthetic corpus partitions and their purposes.

Unless stated otherwise, all results in Section 8 use final_r.

The final_r partition contains:

- 1,400 **NON_IMPACTED** cases
- 460 **IMPACTED** cases
- 160 **INDETERMINATE** cases

These cases are divided into seven strata:

Stratum	Cases	Description
NL_DISPOSITIVE	700	Non-impacted with a certifying ground
NL_SUGGESTIVE	350	Non-impacted with only a suggestive ground
NL_NO_GROUND	200	Non-impacted with no ground
NL_CONDITIONAL	150	Non-impacted with a conditionally certifying ground
IMP_NO_GROUND	230	Impacted with no ground
IMP_SUGGESTIVE	230	Impacted with a suggestive ground
INDETERMINATE	160	Cases that cannot be resolved from the available evidence

Table 4. Gold-label strata in the final_r evaluation set.

The most important adversarial stratum is **IMP_SUGGESTIVE**. In these cases, suggestive evidence appears to support non-impact even though the artifact is actually impacted. Any system that allows suggestive evidence to authorize clearance will make false clearances here.

5.2. Experimental arms

The study compared six approaches:

Arm	Description
A	Ungated rules, where any negative ground may authorize clearance
B	A permissive baseline that clears nearly every case
C	Certificate-gated RCII² using no labeled training data
D	A supervised baseline adjusted to clear the same number of cases as Arm C
E	A supervised baseline tested with label budgets of 0, 10, 25, 50, 100, 250, 500, and 600
F	A supervised baseline trained on one distribution and evaluated on the shifted final_s partition

Table 5. Experimental arms used to compare certificate-gated clearance with ungated and supervised baselines.

Arm C uses no labeled training data. Arms D, E, and F use the same structured features available to the ground predicates.

5.3. Ablations

Thirteen ablation variants were used to test the contribution of individual grounds, certificate requirements, and decision rules. Table 6 summarizes each variant.

Ablation	Change
AB-1	Remove G-JUR
AB-2	Remove all certifying grounds
AB-3	Remove G-SUP
AB-4	Remove G-EXP
AB-5	Remove G-WDR
AB-6	Promote suggestive grounds to certifying
AB-7	Remove the dependency-enumeration requirement
AB-8a–d	Remove selected certificate field requirements
AB-9	Invert the decision rule
AB-10	Apply random clearance at matched coverage

Table 6. Ablation variants used to test the contribution of individual grounds and verification safeguards.

5.4. Tamper testing

Twelve mutation classes were tested. Each class was applied to 50 valid certificates, producing 600 tampered certificates in total.

Mutation class	Description
ALTERED_CITATION	Changes a cited source value
CHANGED_DATE	Changes a relevant date
FABRICATED_AUTHORITY_EDGE	Adds an authority link that does not exist
NON_AUTHORITATIVE_EDGE	Uses a dependency link that is not authoritative
OMITTED_DEPENDENCY	Removes a required dependency
WRONG_RULE_VERSION	Substitutes an incorrect rule version
CORRUPTED_HASH	Alters the certificate hash
INCOMPLETE_ENUMERATION	Marks an incomplete dependency set as complete
SUGGESTIVE_AS_DISPOSITIVE	Uses a suggestive ground as if it could authorize clearance
CERTIFICATE_WRONG_CASE	Attaches the certificate to the wrong case
CONFLICTING_SOURCE_RECORDS	Introduces evidence that contradicts the clearance
STALE_CORPUS_MANIFEST	Uses an outdated corpus manifest hash

Table 7. Certificate mutation classes used in the tamper-resistance evaluation.

5.5. Reproducibility protocol

The laboratory was implemented in Python 3.12 using only the standard library and no external dependencies.

Each component used its own fixed random seed:

- corpus: 42
- supervised baseline: 43
- tamper testing: 44
- ablations: 45

Each seed was applied through its own `random.Random(seed)` instance so that components did not share global random state.

All JSONL outputs used a consistent format with sorted keys, and records were ordered by scenario identifier. The full laboratory was run twice, and the deterministic scientific outputs were compared byte for byte.

6. Results

6.1. Primary result

Arm C cleared 850 of 2,020 cases with zero observed false clearances. Its false-clearance rate was 0.0000, with a 95% Wilson confidence interval of [0.0000, 0.0083]. Non-impact precision was 1.0000, and safe coverage was 0.4208.

Arm	Cleared	False clearances	FCR	NI precision	Coverage
A, ungated rules	979	213	0.4630	0.7007	0.4847
B, permissive baseline	2,019	460	1.0000	0.6934	0.9995
C, certificate-gated RCII ²	850	0	0.0000	1.0000	0.4208
D, supervised matched coverage	850	56	0.1217	0.8965	0.4208

Table 8. Primary results on final_r (n = 2,020; 460 gold-IMPACTED cases). False-clearance rate and coverage are reported together.

Gold label	Cleared	Escalated	Total
NON_IMPACTED	850	550	1,400
IMPACTED	0	460	460
INDETERMINATE	0	160	160
Total	850	1,170	2,020

Table 9. Arm C dispositions by gold label.

The 850 clearances were supported by four terminating grounds:

Ground	Clearances
G-SUP	246
G-JUR	230
G-EXP	224
G-WDR	150

Table 10. Terminating grounds

None of the pre-registered failure conditions was triggered. F1 was not triggered because the false-clearance rate remained below 5% and the upper confidence bound remained below 10%. F2 was not triggered because safe coverage remained above 5%. F3 was not triggered because certificate acceptance remained above 98% and tamper rejection remained above 95%.

6.2. Clearance Quality at Matched Coverage

The experiment shows that RCII² correctly enforced certificate-gated clearance and selected only certifiable cases, while a supervised baseline at the same coverage also cleared impacted and unresolved cases.

Arm C cleared all 700 **NI_DISPOSITIVE** cases and all 150 **NI_CONDITIONAL** cases, for a total of 850 clearances. It cleared no cases from the suggestive, unsupported, impacted, or unresolved strata.

This supports two clear findings. First, RCII² enforced the certificate-gated decision rule exactly as intended and did not allow unsupported cases through the clearance gate. Second, at the same coverage of 0.4208, the supervised baseline also cleared 850 cases, but 56 of those came from **IMP_NO_GROUND** and were false clearances. It also cleared 43 **NI_NO_GROUND** cases and 32 **INDETERMINATE** cases.

The result therefore shows that, within this controlled synthetic corpus, RCII² selected only certifiable cases while the supervised baseline did not. The 0.4208 coverage rate should not be treated as an estimate

of production performance because it reflects the proportion of certifiable cases built into the corpus. Production coverage will depend on the real prevalence of certifying grounds in operational data.

6.3. Certificate integrity

All 850 valid certificates were accepted, and no case was cleared without an accepted certificate.

All 600 tested certificate mutations were rejected. Because each mutation changed hash-bound content, every case triggered V-04, the certificate-integrity check. Several class-specific rules also fired as expected, including V-10 for incomplete dependencies, V-11 for invalid authority paths, V-14 for stale corpus manifests, and V-15 for conflicting evidence.

The results therefore show that the verifier detected every tested post-generation mutation. They do not yet show how the class-specific rules perform when a compromised generator alters a certificate before the hash is created. A future laboratory cycle should test hash-preserving or pre-hash mutations to evaluate that scenario directly.

Measure	Result
Certificates issued	850
Certificates accepted by the independent verifier	850 / 850 (1.0000)
Cleared cases without a certificate	0
Certificates without a corresponding clearance	0
Tampered certificates rejected	600 / 600 (1.0000)

Table 11. Certificate issuance, verification, and tamper-test results

6.4. Diagnostic label-deprivation experiment

The supervised baseline produced no clearances with zero labels and cleared the entire corpus at every nonzero label budget. Its false-clearance rate was therefore 1.0000 at every tested budget from 10 to 600 labels.

This is a diagnostic result, not a valid comparison of supervised performance under label scarcity. The baseline appears to have collapsed to a degenerate decision rule. The result is not used to support the

main comparison with RCII².

Further evaluation requires review of class weighting, feature representation, decision thresholds, and calibration before the supervised label-deprivation experiment is repeated.

Label budget	Cleared	FCR
0	0	Not applicable
10	2,020	1.0000
25	2,020	1.0000
50	2,020	1.0000
100	2,020	1.0000
250	2,020	1.0000
500	2,020	1.0000
600	2,020	1.0000
Arm C, 0 labels	850	0.0000

Table 12. Supervised baseline performance under different label budgets.

6.5. Distribution shift performance

Arm	Partition	Cleared	FCR
C, certificate-gated	final_s	180 / 400	0.0000
E, supervised	final_s	400 / 400	1.0000

Table 13. Performance under the controlled distribution shift.

Under the tested shift, Arm C cleared 180 of 400 cases, for coverage of 0.4500, with zero observed false

clearances. Arm F cleared all 400 cases and produced a false-clearance rate of 1.0000.

The difference reflects how the two approaches make decisions. Arm C checks explicit predicates against the current record, so a change in the data distribution affects which cases qualify but does not change what the predicates mean. Arm F relies on a decision boundary learned from the original distribution, and that boundary did not hold under the tested shift.

This result applies only to the single controlled shift evaluated here and should not be treated as evidence of general robustness across all production shifts.

6.6. Ablations

Ablation	Description	Delta	Status
AB-1	Remove G-JUR	230	Decision-level
AB-2	Remove all certifying grounds	850	Decision-level
AB-3	Remove G-SUP	246	Decision-level
AB-4	Remove G-EXP	224	Decision-level
AB-5	Remove G-WDR	150	Decision-level
AB-6	Promote suggestive grounds to certifying	1,169	Decision-level
AB-7	Remove dependency-enumeration requirement	0	Verification-level
AB-8a–d	Remove selected certificate field requirements	0	Verification-level
AB-9	Invert the decision rule	2,020	Decision-level
AB-10	Apply random clearance at matched coverage	1,019	Decision-level

Table 14. Ablation results on final_r (n = 2,020). Delta is the number of decisions that differ from Arm C.

Eight ablations changed the final decisions. Removing each certifying ground changed exactly the number of cases cleared by that ground: 230 for G-JUR, 246 for G-SUP, 224 for G-EXP, and 150 for G-WDR. This shows that each ground contributed independently.

AB-6 produced the most important result. Promoting suggestive grounds to certifying changed 1,169 decisions and allowed the type of over-clearance seen in Arm A, including 213 false clearances. This confirms that the separation between certifying and suggestive grounds is essential.

AB-10 also changed 1,019 decisions. This shows that RCII² is not simply clearing an arbitrary group of 850 cases. It is selecting a specific set based on certifying evidence.

AB-7 and AB-8a–d did not change which cases were selected for clearance. Their role is different. Certificate fields do not determine which cases qualify; they determine whether a proposed clearance survives verification. Their effect is measured by the tamper suite, which rejects mutations targeting each of these fields.

6.7. Reproducibility

After the final package correction, two independent laboratory runs produced 30 of 30 byte-identical deterministic scientific outputs. Wall-clock execution time was stored separately in a nondeterministic execution log and was not included in the comparison.

All published metrics were also reconstructed independently from the frozen prediction ledgers and gold-label files. Every reconstructed value matched the canonical results.

One major reproducibility defect was found and corrected during the laboratory cycle. AB-10 originally used Python's built-in `hash()` function to derive per-case seeds. Because that function can vary across processes, the same run produced different outputs. It was replaced with a SHA-256-derived seed, after which the outputs were byte-identical across runs.

7. Discussion

7.1. Proof and confidence are different quantities

A confidence score reflects the model's internal state. A certificate reflects the evidence.

The difference matters when the system is wrong. A confidence-based system may clear a case incorrectly without revealing the mistake. A certificate-gated system cannot clear the case unless it can produce evidence that passes independent verification. When it cannot, the case is escalated for review.

This is what makes the architecture fail-closed. Escalation is the default, and no confidence threshold can override it.

7.2. Abstention as a design property

Arm C escalated 1,170 of 2,020 cases, or 57.9%. That includes 550 cases that were truly non-impacting but could not be supported by an accepted certificate.

This is intentional. RCII² gives up some automation in exchange for safer clearance. The cost is 550 additional reviews. The benefit is that all 460 impacted cases were escalated.

Whether that trade is worthwhile depends on the institution's cost of manual review compared with the risk of missing an obligation. This experiment does not answer that business question.

7.3. Implications for knowledge-based systems

The approach may apply beyond regulatory change. It is relevant wherever a negative decision removes something from human attention, the cost of a missed case is high, and the supporting evidence can be checked from structured records.

The key knowledge-engineering contribution is the ground taxonomy. Grounds are classified by whether they logically support non-impact and whether they can be checked directly from structured data. They are not classified only by how useful they are for prediction.

The AB-6 result shows that this distinction is essential. When suggestive grounds were allowed to authorize clearance, false clearances increased.

7.4. Where learned components fit

Learned models can support RCII² by ranking escalated cases, suggesting possible grounds for human formalization, or retrieving related artifacts for review.

They may assist the process, but they may not authorize the final clearance. A model score is not independently verifiable evidence.

8. Threats to Validity and Limitations

This study is a controlled proof of concept. Its limits are stated directly below.

8.1. The corpus is synthetic

The corpus was entirely generated. No Rockland Trust documents, customer information, or production data were used.

The synthetic artifacts model policies, procedures, controls, systems, and governed records using structured fields. Real artifacts are more complex. Jurisdiction may be implied rather than stated, version history may branch, withdrawals may preserve some obligations, and applicability may be buried in free text.

The study therefore shows that RCII² works as designed in the laboratory. It does not yet show how it will perform on real operational data.

8.2. Coverage reflects the corpus design

Arm C cleared all 700 `NI_DISPOSITIVE` cases and all 150 `NI_CONDITIONAL` cases. Its coverage of 0.4208 therefore reflects the share of the synthetic corpus designed to be certifiable.

This should not be treated as an estimate of production coverage. Real coverage will depend on how often certifying grounds appear in operational data.

8.3. Zero observed error does not mean zero risk

RCII² produced zero observed false clearances across 460 gold-IMPACTED cases. The upper bound of the 95% confidence interval was 0.0083.

This result does not prove that the true false-clearance rate is zero. It also does not show that zero false clearances will occur on other corpora or in production.

8.4. The corpus and ground taxonomy were developed together

The clearance-ground taxonomy and synthetic corpus were created within the same laboratory design process. The corpus was intentionally built to include certifying, suggestive, conflicting, and no-ground cases.

The experiment therefore tests whether RCII² correctly enforces the designed evidence structure. It does not show that the same structure will appear naturally in independently collected data.

A future study should use a corpus designed or labeled independently by researchers who were not involved in defining the clearance grounds.

8.5. The supervised label-deprivation baseline needs further work

As shown in Section 8.4, the supervised baseline cleared the entire corpus at every nonzero label budget.

This is a degenerate result and may reflect the baseline configuration rather than the limits of supervised learning. The label-deprivation comparison is therefore treated as diagnostic only.

A fairer comparison requires further review of class weighting, feature representation, thresholds, and calibration.

8.6. Single corpus, single shift tested

The primary results come from one synthetic corpus generated by one process. Only one controlled distribution shift was evaluated.

The study has not yet been replicated across independent corpora or multiple shift conditions.

8.7. Verifier soundness is tested, not proven

The verifier rejected all 600 tested certificate mutations.

This is evidence that it detected the tested post-generation changes. It is not a formal proof of verifier soundness. Because every mutation changed hash-bound content, every case also triggered the content-hash check.

Future testing should include hash-preserving and pre-hash mutations to exercise the class-specific verifier rules more directly.

8.8. No production deployment or workflow integration

RCII² has not yet been integrated into an operational workflow.

The study does not measure reviewer time, workload reduction, implementation cost, or production coverage.

9. Practical Implications

9.1. Deployment model

RCII² would sit before human review. Cases with accepted certificates would be cleared and stored with the certificate. All other cases would go to a reviewer.

The system would never remove a case from human attention without leaving behind evidence that can be checked.

9.2. The certificate as the audit record

For a regulated institution, the certificate is more valuable than the clearance itself.

If an examiner asks why an obligation was not reviewed, the institution can provide a record tied to specific source fields and rules. That record can be checked independently. A confidence score cannot provide the same level of support.

9.3. Prerequisites

RCII² requires a governed corpus in machine-readable form. This includes version history, effective dates, jurisdiction, dependency links, and authority relationships.

Creating and maintaining that corpus is likely to be the largest cost of adoption. This study does not estimate that cost. The reasoning and verification layer is comparatively lightweight.

9.4. Ongoing governance

The clearance grounds must be reviewed over time. A ground that is logically sufficient today may need to be reconsidered when regulations, policies, or business structures change.

The organization should also monitor which grounds are producing clearances. A major shift in that pattern may change both coverage and risk.

9.5. What this study does not support

This study does not support production deployment yet.

Before deployment, RCII² must be tested on realistic operational data, compared with current human review, evaluated for implementation and governance cost, and reviewed by compliance officers, auditors, and examiners to confirm that the certificates are useful in practice.

10. Conclusion

We presented RCII², a knowledge-based decision architecture that allows a non-impact clearance only when an independent verifier accepts a machine-verifiable certificate. Certifying grounds must logically support non-impact and be checkable from structured data. Suggestive grounds may support review, but they cannot authorize clearance. Every other case is escalated.

On a frozen synthetic corpus of 2,020 decisions, RCII² cleared all 850 cases designed to satisfy certifying or conditionally certifying grounds. It did not clear any impacted or uncertifiable cases. This produced zero observed false clearances, with an FCR of 0.0000 and a 95% confidence interval of [0.0000, 0.0083], non-impact precision of 1.0000, and coverage of 0.4208. All 850 valid certificates were accepted, and all 600 tested post-generation mutations were rejected. A supervised baseline clearing the same number of cases produced an FCR of 0.1217. None of the pre-registered failure conditions was triggered.

These results show that RCII² correctly enforced the clearance policy, preserved certificate integrity, and produced reproducible results within the designed synthetic corpus. The coverage rate reflects the corpus design and should not be treated as an estimate of production performance. Zero observed error also does not mean zero future risk. External validation will require an independently designed, realistically distributed corpus, a human baseline, and evidence that the certificates are useful to the people who must rely on them.

The central claim is that the admissibility rule matters. Requiring verifiable evidence before a negative decision changes the failure mode. When RCII² cannot prove the clearance, it escalates the case instead of clearing it silently. For decisions that remove obligations from human attention, that difference is the point.

Statements and Declarations

Funding

This research received no external funding.

Potential Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. A patent disclosure related to the method described in this paper is under preparation.

Ethics

This study involved no human participants, human data, animal subjects, or activities requiring institutional ethics review. The corpus was entirely synthetic.

Data Availability

The complete laboratory package, rcii2-cycle2-b1-1.0.0 with run identifier B1-RUN1, includes the corpus generator, six frozen data partitions, gold labels, rule definitions, certificate schema, decision engine, supervised baselines, independent verifier, thirteen ablations, twelve tamper classes, fourteen prediction ledgers, 850 certificates, derived metrics, two reproducibility runs, reconstruction scripts, manifests, checkpoint reports, and the defect log.

The laboratory runs on Python 3.12 or later using only the standard library. It can be rerun with a single command, and all reported metrics can be reconstructed independently from the frozen artifacts.

CRedit Authorship Contribution

Jeffrey Esposito: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing, original draft, Writing, review and editing, Visualization.

Ankur Garg: Conceptualization, Methodology, Supervision, Writing, review and editing, Visualization.

Both authors reviewed and approved the final manuscript.

Use of Generative AI

During the preparation of this manuscript, the authors used OpenAI's ChatGPT to assist with language refinement, structural organization, and manuscript preparation. All scientific concepts, theoretical contributions, analyses, interpretations, and final editorial decisions were developed, verified, and approved by the authors, who take full responsibility for the content of this publication.

References

1. ^a ^bAdadi A, Berrada M (2018). "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)." *IEEE Access*. 6:52138–52160. doi:[10.1109/ACCESS.2018.2870052](https://doi.org/10.1109/ACCESS.2018.2870052).
2. ^a ^bBarredo Arrieta A, Díaz-Rodríguez N, Del Ser J, et al (2020). "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges Toward Responsible AI." *Inf Fusion*. 58:82–115. doi:[10.1016/j.inffus.2019.12.012](https://doi.org/10.1016/j.inffus.2019.12.012).
3. ^a ^bGuidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D (2018). "A Survey of Methods for Explaining Black Box Models." *ACM Comput Surv*. 51(5):93. doi:[10.1145/3236009](https://doi.org/10.1145/3236009).

4. ^a ^bEl-Yaniv R, Wiener Y (2010). "On the Foundations of Noise-Free Selective Classification." *J Mach Learn Res.* **11**:1605–1641.
5. ^a ^bGeifman Y, El-Yaniv R (2017). "Selective Classification for Deep Neural Networks." In: *Advances in Neural Information Processing Systems* 30. p. 4878–4887.
6. ^a ^bGeifman Y, El-Yaniv R (2019). "SelectiveNet: A Deep Neural Network with an Integrated Reject Option." In: *Proceedings of the 36th International Conference on Machine Learning*. PMLR 97. p. 2151–2159.
7. ^a ^bMcConnell RM, Mehlhorn K, Näher S, Schweitzer P (2011). "Certifying Algorithms." *Comput Sci Rev.* **5**(2): 119–161. doi:[10.1016/j.cosrev.2010.09.009](https://doi.org/10.1016/j.cosrev.2010.09.009).
8. ^a ^bNecula GC (1997). "Proof-Carrying Code." In: *Proceedings of POPL '97*. p. 106–119. doi:[10.1145/263699.263712](https://doi.org/10.1145/263699.263712).
9. ^ΔBohner SA (2002). "Software Change Impacts—An Evolving Perspective." In: *Proceedings of the International Conference on Software Maintenance*. p. 263–272.
10. ^ΔBohner SA, Arnold RS (1996). "An Introduction to Software Change Impact Analysis." In: *Software Change Impact Analysis*. IEEE Computer Society Press. p. 1–26.
11. ^ΔLi B, Sun X, Leung H, Zhang S (2013). "A Survey of Code-Based Change Impact Analysis Techniques." *Softw Test Verif Reliab.* **23**(8):613–646. doi:[10.1002/stvr.1475](https://doi.org/10.1002/stvr.1475).
12. ^ΔGovernatori G, Rotolo A (2010). "Norm Compliance in Business Process Modeling." In: *RuleML 2010*. LNCS 6403. Springer. p. 194–209. doi:[10.1007/978-3-642-16289-3_17](https://doi.org/10.1007/978-3-642-16289-3_17).
13. ^ΔHashmi M, Governatori G (2018). "Norms Modeling Constructs of Business Process Compliance Management Frameworks: A Conceptual Evaluation." *Artif Intell Law.* **26**(3):251–305. doi:[10.1007/s10506-017-9215-8](https://doi.org/10.1007/s10506-017-9215-8).
14. ^ΔHashmi M, Governatori G, Wynn MT (2016). "Normative Requirements for Regulatory Compliance: An Abstract Formal Framework." *Inf Syst Front.* **18**(3):429–455. doi:[10.1007/s10796-015-9558-1](https://doi.org/10.1007/s10796-015-9558-1).
15. ^ΔHashmi M, Governatori G, Lam HP, Wynn MT (2018). "Are We Done with Business Process Compliance: State of the Art and Challenges Ahead." *Knowl Inf Syst.* **57**(1):79–133. doi:[10.1007/s10115-017-1142-1](https://doi.org/10.1007/s10115-017-1142-1).
16. ^ΔAlarfaj FK (2026). "Focusing on Legal Cases: Automatic Classification of Legal Documents with Sentence Embeddings and Deep Learning Models." *PLoS One.* **21**(6):e0350673. doi:[10.1371/journal.pone.0350673](https://doi.org/10.1371/journal.pone.0350673).
17. ^ΔDuffy W, O'Connell E, McCarroll N, et al (2025). "Evaluating Rule-Based and Generative Data Augmentation Techniques for Legal Document Classification." *Knowl Inf Syst.* **67**:7825–7846. doi:[10.1007/s10115-025-02454-x](https://doi.org/10.1007/s10115-025-02454-x).
18. ^ΔZeni N, Kiyavitskaya N, Mich L, Cordy JR, Mylopoulos J (2015). "GaiusT: Supporting the Extraction of Rights and Obligations for Regulatory Compliance." *Requirements Eng.* **20**(1):1–22. doi:[10.1007/s00766-013-0181-](https://doi.org/10.1007/s00766-013-0181-)

8.

19. [△]Ahmed R, Sumesh EP (2026). "An Intelligent Phishing Email Detection System Using Ensemble Methods and Explainable AI." *Knowledge-Based Syst.* 337:115388. doi:[10.1016/j.knosys.2026.115388](https://doi.org/10.1016/j.knosys.2026.115388).
20. [△]Ahn D, Kim S, Ko BC (2026). "Energy-Ensemble Concept Bottleneck Models for Enhancing Interpretability and Accuracy in Concept-Based Learning." *Knowledge-Based Syst.* 341:115821. doi:[10.1016/j.knosys.2026.115821](https://doi.org/10.1016/j.knosys.2026.115821).
21. [△]Pham TTT, Nguyen TTD (2026). "Causal Reasoning over Ontology-Enriched Graphs for Interpretable Medical AI." *Knowledge-Based Syst.* 341:115803. doi:[10.1016/j.knosys.2026.115803](https://doi.org/10.1016/j.knosys.2026.115803).
22. [△]Forgy CL (1982). "Rete: A Fast Algorithm for the Many Pattern/Many Object Pattern Match Problem." *Artif Intell.* 19(1):17–37. doi:[10.1016/0004-3702\(82\)90020-0](https://doi.org/10.1016/0004-3702(82)90020-0).
23. [△]Hayes-Roth F (1985). "Rule-Based Systems." *Commun ACM.* 28(9):921–932. doi:[10.1145/4284.4286](https://doi.org/10.1145/4284.4286).
24. [△]Niu DY, Mi JS (2026). "Reductions of Three-Way Concept Lattices and Three-Way Rules Extraction Based on Boolean Formal Contexts." *Knowledge-Based Syst.* 341:115757. doi:[10.1016/j.knosys.2026.115757](https://doi.org/10.1016/j.knosys.2026.115757).
25. [△]Buneman P, Khanna S, Tan WC (2001). "Why and Where: A Characterization of Data Provenance." In: *Database Theory—ICDT 2001*. LNCS 1973. Springer. p. 316–330. doi:[10.1007/3-540-44503-X_20](https://doi.org/10.1007/3-540-44503-X_20).
26. [△]Herschel M, Diestelkämper R, Ben Lahmar H (2017). "A Survey on Provenance: What For? What Form? What From?" *VLDB J.* 26:881–906. doi:[10.1007/s00778-017-0486-1](https://doi.org/10.1007/s00778-017-0486-1).
27. [△]Moreau L, Missier P, editors (2013). "PROV-DM: The PROV Data Model." W3C Recommendation.
28. [△]Moreau L, Groth P, Cheney J, Lebo T, Miles S (2015). "The Rationale of PROV." *J Web Semant.* 35:235–257. doi:[10.1016/j.websem.2015.04.001](https://doi.org/10.1016/j.websem.2015.04.001).
29. [△]Alsulaimani S, Zhao M, Hussain FK (2026). "Fractional Digital Asset Ownership—An Intelligent Fractional NFT-Based Approach to the Reliable Provenance of Digital Asset Co-Ownership." *Knowledge-Based Syst.* 341:115631. doi:[10.1016/j.knosys.2026.115631](https://doi.org/10.1016/j.knosys.2026.115631).
30. [△]Pons G, Bilalli B, Queralt A (2026). "Capturing and Anticipating User Intents in Data Analytics via Knowledge Graphs." *Knowledge-Based Syst.* 341:115835. doi:[10.1016/j.knosys.2026.115835](https://doi.org/10.1016/j.knosys.2026.115835).
31. [△]Chow CK (1970). "On Optimum Recognition Error and Reject Tradeoff." *IEEE Trans Inf Theory.* 16(1):41–46. doi:[10.1109/TIT.1970.1054406](https://doi.org/10.1109/TIT.1970.1054406).
32. [△]Angelopoulos AN, Bates S (2023). "Conformal Prediction: A Gentle Introduction." *Found Trends Mach Learn.* 16(4):494–591. doi:[10.1561/22000000101](https://doi.org/10.1561/22000000101).
33. [△]Shafer G, Vovk V (2008). "A Tutorial on Conformal Prediction." *J Mach Learn Res.* 9:371–421.

34. [△]Vovk V, Gammerman A, Shafer G (2005). *Algorithmic Learning in a Random World*. Springer. doi:[10.1007/b106715](https://doi.org/10.1007/b106715).
35. [△]Fu H, Wang Q, Gong C, Su Z (2026). "Incremental Open-Set Evidential Reasoning for Uncertainty-Aware Fault Diagnosis in Rotating Machinery." *Knowledge-Based Syst.* **344**:116090. doi:[10.1016/j.knosys.2026.116090](https://doi.org/10.1016/j.knosys.2026.116090).
36. [△]Alkassar E, Böhme S, Mehlhorn K, Rizkallah C (2011). "Verification of Certifying Computations." In: *Computer Aided Verification*. LNCS 6806. Springer. p. 67–82. doi:[10.1007/978-3-642-22110-1_7](https://doi.org/10.1007/978-3-642-22110-1_7).
37. [△]Namjoshi KS (2001). "Certifying Model Checkers." In: *Computer Aided Verification*. LNCS 2102. Springer. p. 2–13. doi:[10.1007/3-540-44585-4_2](https://doi.org/10.1007/3-540-44585-4_2).
38. [△]Assurance Case Working Group (2021). "Goal Structuring Notation Community Standard, Version 3." SCSC-141C.
39. [△]Newcomb A, Ochoa O (2026). "Formal Methods for Safety-Critical Machine Learning: A Systematic Literature Review." *Front Artif Intell.* **9**:1749956. doi:[10.3389/frai.2026.1749956](https://doi.org/10.3389/frai.2026.1749956).
40. [△]Wei R, Kelly TP, Dai X, Zhao S, Hawkins R (2019). "Model Based System Assurance Using the Structured Assurance Case Metamodel." *J Syst Softw.* **154**:211–233. doi:[10.1016/j.jss.2019.05.013](https://doi.org/10.1016/j.jss.2019.05.013).
41. [△]Khosravani M (2026). "Selective Prediction from Agreement: A Lipschitz-Consistent Version Space Approach." *arXiv preprint arXiv:2605.02611*.
42. [△]Xu H (2026). "Cherry-Pick Override: Unsafe Directional Commitment in LLM Judges Under Mixed Evidence." *arXiv preprint arXiv:2606.07834*.

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.