AMEF: An Evidence-Centric Framework for Multi-source Reasoning and Safe Abstention in Smart Parking

Ferry Sugianto*, Mir'atul Khusna Mufida
Informatics Engineering Study Program
Politeknik Negeri Batam
*Email: ferry.7312411001@students.polibatam.ac.id

***Abstract*—Smart parking systems increasingly combine heterogeneous information sources, including computer vision, human reports, parking officers, and external systems. These sources may differ in reliability, freshness, coverage, uncertainty, and provenance, making direct aggregation insufficient for operational reasoning. This paper presents the Adaptive Multi-source Evidence Framework (AMEF), an evidence-centric framework that separates Observation, Evidence, Estimate, and Knowledge, while assigning distinct responsibilities to Qualification, Evaluation, Reasoning, and Learning. The framework explicitly supports Safe Abstention when available evidence is insufficient for a substantive estimate. AMEF is specialized into a ParkEase domain profile and implemented as a reference system supporting multi-source observation intake, evidence qualification and evaluation, reasoning episodes, Safe Abstention, and traceability. The implementation is evaluated using an external seeded conformance test suite comprising 29 scenario definitions across five scenario families and 290 replay runs. All runs satisfy the specified policy assertions. The final batch produces 160 Estimates, 120 Safe Abstentions, and 10 expected configuration failures. Among the 160 Estimates, 120 agree with the synthetic ground truth and 40 do not, yielding 75% diagnostic correctness. These results demonstrate behavioral conformance of the evaluated implementation while also exposing limitations in the current reasoning and provenance method. The diagnostic result is limited to the constructed synthetic scenarios and is not interpreted as field accuracy or superiority over alternative fusion methods.**

***Index Terms*—Adaptive Multi-source Evidence Framework, AMEF, smart parking, multi-source information, evidence lifecycle, Safe Abstention, traceability, reference implementation, conformance testing, synthetic testing.**

I. Introduction

Smart parking systems increasingly rely on heterogeneous information sources to estimate parking occupancy and support operational monitoring. Computer-vision and learning-based approaches have been widely used for parking occupancy detection and surveillance [1], [2], [3], [4], [5]. Computer-vision models, human users, parking officers, and external systems may provide information about the same parking zone while differing in modality, freshness, reliability, coverage, uncertainty, and provenance. Consequently, an operational system cannot safely assume that every incoming source output is equivalent to a final decision input.

Existing smart-parking research has addressed perception, occupancy detection, prediction, sensor fusion, and deployment [5], [1], [2], [3], [4], [6]. Research on multi-source evidence fusion has additionally addressed credibility, uncertainty, robustness, and conflicting information [7], [8], [9], [10]. These approaches are important at the sensing, prediction, and fusion layers, but a broader operational question remains: how should heterogeneous observations be qualified, evaluated, combined, traced, and converted into an operational outcome when evidence is incomplete or conflicting?

This paper proposes the Adaptive Multi-source Evidence Framework (AMEF), an evidence-centric framework that explicitly separates four epistemic representations—Observation, Evidence, Estimate, and Knowledge—and four process responsibilities—Qualification, Evaluation, Reasoning, and Learning. The framework also defines Safe Abstention as an explicit outcome when the available Evidence does not satisfy configured sufficiency conditions, drawing conceptually on the reject-option and selective-classification literature [11], [12], [13], [14].

AMEF is intentionally defined as a framework rather than as a single fusion algorithm. This distinction allows domain-specific evaluation and reasoning methods to be instantiated and versioned without changing the framework's conceptual lifecycle. In this work, AMEF is specialized into the ParkEase AMEF Domain Profile and realized as a ParkEase Reference Implementation for smart parking. The executed evaluation is limited to the Operational Loop; the Learning Loop remains within the design scope.

The contributions of this paper are threefold:

1) AMEF introduces an evidence-centric lifecycle that separates Observation, Evidence, Estimate, and Knowledge, with Qualification, Evaluation, Reasoning, and Learning as distinct responsibilities.
2) ParkEase provides a concrete domain instantiation and reference implementation of the AMEF Operational Loop, including Safe Abstention and end-to-end traceability.
3) An external seeded conformance test suite evaluates the implementation using 29 scenario definitions and 290 replay runs, explicitly separating policy conformance from diagnostic comparison against synthetic ground truth.

II. Related Work

A. Smart Parking Perception and Occupancy Estimation

Smart-parking research has explored computer vision, deep learning, edge artificial intelligence, sensor fusion, and occupancy prediction [5], [1], [2], [3], [4], [6]. These studies primarily address perception, detection, prediction, or deployment. In the AMEF position, such outputs are treated as Observations entering a broader evidence lifecycle rather than as final operational decisions.

B. Multi-source Evidence Fusion

Multi-source fusion research addresses uncertainty, credibility, conflict, and evidence combination [7], [8], [9], [10]. Such methods provide useful methodological foundations for evaluating and combining heterogeneous information. AMEF addresses a broader lifecycle concern: how information moves from Observation into qualified Evidence, how Evidence is evaluated in context, how contributions are used for Reasoning, and how insufficient Evidence can result in explicit Safe Abstention.

C. Traceability and Explainability

Traceability and explainability research examines how AI processes and outcomes can be inspected or communicated [15], [16], [17]. AMEF addresses traceability from an operational perspective by maintaining an auditable relationship among Producer, Observation, Evidence, Evaluation, Contribution, and Reasoning Episode. This audit chain is complementary to, rather than a replacement for, human-facing explanation mechanisms.

D. Conformance Testing and Design Science

The research treats AMEF and its related artifacts as a design-science-oriented framework, domain profile, reference implementation, and conformance test suite. Automated conformance-testing approaches provide a methodological precedent for externally checking specified system behavior [18]. External black-box testing is used to evaluate observable behavior without depending on internal implementation state.

E. Research Gap

Despite these advances, the reviewed studies address different parts of the problem space rather than providing a unified operational evidence lifecycle for heterogeneous information. Smart-parking studies primarily focus on perception, detection, prediction, or deployment [5], [1], [2], [3], [4]; evidence-fusion studies focus on mathematical combination, credibility, uncertainty, and conflict [7], [8], [9], [10]; traceability and explainability studies focus on auditability, explanation, and user trust [15], [16], [17]; and conformance-testing studies focus on behavioral verification [18]. Data-lifecycle perspectives also emphasize the management and reuse of data across stages [19]. These strands leave a gap between heterogeneous observations and an explicitly qualified, evaluated, traceable, and uncertainty-aware operational outcome. AMEF addresses this gap by defining an evidence-centric lifecycle and by making Safe Abstention and end-to-end traceability explicit framework capabilities.

TABLE I
Positioning of AMEF Against Related Work

Area	Representative Work	Primary Focus	AMEF Position
Smart parking	CV, forecasting, edge-cloud studies	Detection, prediction, deployment	Consumes outputs as Observations
Evidence fusion	Multisource fusion studies	Fusion, uncertainty, conflict	Defines lifecycle and method boundary
Traceability	Traceability/explainability studies	Auditability and explanation	Evidence-to-outcome audit chain
Framework/DSR	Framework and systems studies	Artifacts and evaluation	AMEF as framework artifact
Conformance	Automated testing studies	Specified behavior	External black-box conformance
Reject option	Reject-option studies	Abstention/rejection	Safe Abstention as explicit outcome

III. AMEF Framework

A. Framework Overview

The Adaptive Multi-source Evidence Framework (AMEF) is an evidence-centric framework for organizing heterogeneous information before it is used to produce an operational estimate. Rather than treating every incoming source output as an equivalent decision input, AMEF distinguishes four epistemic representations: Observation, Evidence, Estimate, and Knowledge. These representations are processed through four distinct responsibilities: Qualification, Evaluation, Reasoning, and Learning.

AMEF is a framework rather than a single fusion algorithm. Domain-specific evaluation and reasoning methods can therefore be instantiated and versioned without changing the framework's conceptual structure. In ParkEase, these methods are realized through a versioned implementation method.

B. Epistemic Representations

Observation represents an incoming claim or measurement produced by a source. An Observation may originate from a human user, parking officer, computer-vision model, or external system. It retains source and contextual information but is not automatically treated as trusted Evidence.

Evidence represents an Observation that has passed the Qualification boundary and is eligible for downstream Evaluation. Evidence remains associated with provenance and contextual attributes so that its contribution can be evaluated rather than treated as an unconditional fact.

Estimate is a substantive outcome produced by Reasoning from evaluated Evidence. It represents the system's current operational conclusion for the target context.

Knowledge represents information produced by the Learning Loop from validated experience. Knowledge is distinct from an instantaneous Estimate because it is intended to

influence future operation rather than the current episode alone.

C. Process Responsibilities

Qualification determines whether an Observation satisfies structural and contextual requirements for use as Evidence. **Evaluation** assesses qualified Evidence using contextual properties such as freshness, reliability, coverage, uncertainty, and source relationships. **Reasoning** combines evaluated contributions and determines the operational outcome. **Learning** forms Knowledge from validated experience and is conceptually separated from the immediate operational decision.

The executed scope is the AMEF Operational Loop:

$$Observation \rightarrow Qualification \rightarrow Evidence \rightarrow Evaluation \rightarrow Reasoning \rightarrow \{Estimate, Safe\ Abstention\}. \tag{1}$$

The Learning Loop is retained as part of the framework design but is outside the executed experimental scope.

D. Safe Abstention

Safe Abstention is an explicit operational outcome rather than an implementation failure. This design is conceptually related to reject-option classification, in which a classifier may defer or reject a prediction when confidence or expected risk does not justify a definitive decision [11], [12], [13], [14]. It allows the system to refrain from producing a substantive Estimate when available Evidence does not satisfy configured sufficiency conditions. In ParkEase, abstention may result from conditions involving insufficient effective support, high uncertainty, inadequate evidence diversity, absence of usable current Evidence, or other configured reasoning constraints.

E. Evaluation and Reasoning Method Boundary

The ParkEase reference implementation uses a versioned baseline method. Freshness is represented by

$$f(age) = \begin{cases} 1, & age \leq 10, \\ e^{-0.03(age-10)}, & age > 10. \end{cases} \tag{2}$$

Contextual reliability is represented as

$$R_c = \mathrm{clamp}(R_b M, 0, 1), \tag{3}$$

and potential influence is represented as

$$I_p = f R_c C U, \tag{4}$$

where R_b is baseline reliability, M is a contextual mode factor, C represents coverage, and U represents the uncertainty-related term used by the implementation.

Dependency and conflict adjustments are subsequently applied:

$$I_e = I_p D F_c, \tag{5}$$

where D is the configured dependency factor and F_c is the conflict factor.

These equations are implementation-specific parameters of the ParkEase baseline method and are not universal AMEF rules.

TABLE II
OBSERVATION SOURCES IN PARKEASE

Producer	Role in ParkEase
HUMAN_USER	User-generated parking observations or reports.
PARKING_OFFICER	Operational observations produced by parking personnel.
CV_MODEL	Computer-vision-derived parking occupancy observations.
EXTERNAL_SYSTEM	Observations originating from external information systems.

For candidate state S_i, aggregate contribution can be represented as

$$C(S_i) = \sum_j I_{ij}, \tag{6}$$

followed by normalized support

$$P(S_i) = \frac{C(S_i)}{\sum_k C(S_k)}. \tag{7}$$

Entropy can be used as one of the reasoning diagnostics:

$$H = -\sum_i P(S_i) \log P(S_i). \tag{8}$$

F. Traceability

AMEF maintains the traceability chain

$$Producer \rightarrow Observation \rightarrow Evidence \rightarrow Evaluation \rightarrow Contributio \tag{9}$$

This chain permits an Estimate or Safe Abstention to be traced back to the observations and producers participating in the reasoning episode.

IV. PARKEASE REFERENCE IMPLEMENTATION

A. Domain Instantiation

ParkEase is a domain instantiation of AMEF for smart parking management. The implementation maps abstract AMEF concepts to parking-zone observations, evidence records, evaluation results, reasoning episodes, and operational outcomes. ParkEase is not treated as the definition of AMEF; it provides a concrete reference realization through which the framework's Operational Loop can be executed and evaluated.

B. Observation Sources

C. Parking State Representation

ParkEase uses candidate parking states $S1$–$S4$. The state mapping is domain-specific and uses active capacity together with vehicle-count observations. Candidate states are treated as hypotheses during the reasoning episode rather than as observations themselves.

D. Observation-to-Evidence Processing

An incoming observation does not directly enter the final reasoning result. It first passes through Qualification. Qualified observations become Evidence and retain provenance and contextual attributes. Evaluation then produces contribution-related information used by Reasoning.

TABLE III
CONFORMANCE SCENARIO FAMILIES

Family	Evaluation Purpose
Aligned	Normal cases where source observations support a consistent candidate state.
Conflict	Contradictory observations and conflict-handling behavior.
Temporal	Freshness, aging, and temporal reasoning constraints.
Insufficient Evidence	Explicit Safe Abstention and controlled insufficiency conditions.
Adversarial	Edge cases involving misleading, coordinated, or structurally challenging evidence.

E. Reasoning Episode

A reasoning episode aggregates evaluated contributions associated with a target Parking Zone. Contributions are grouped by candidate state and normalized to form a support distribution. The implementation then considers the highest-supported candidate together with decision criteria including effective support, the margin between leading candidates, and normalized entropy. Depending on configured thresholds, the episode produces either an Estimate or Safe Abstention.

F. Traceability Implementation

ParkEase preserves provenance across the operational pipeline. An observation is associated with its producer and subsequently linked to the evidence record, evaluation result, contribution, and reasoning episode in which it participates. This provides an end-to-end audit path for examining how an Estimate or Safe Abstention was produced.

V. CONFORMANCE EVALUATION

A. Evaluation Objective

The evaluation investigates whether the ParkEase Reference Implementation exhibits behavior specified by the AMEF/ParkEase operational policy under controlled and reproducible scenarios. The evaluation is therefore primarily a conformance assessment rather than a field-accuracy study. A separate diagnostic comparison against synthetic ground truth is used to identify limitations in the current evaluation and reasoning method.

B. External Test Architecture

The test harness is external to the ParkEase implementation and interacts with the system through REST APIs. Scenario constraints and synthetic ground truth are maintained outside the system under test. This separation allows the test harness to evaluate observable behavior without relying on internal implementation state.

C. Scenario Design

The final conformance batch contains 29 scenario definitions grouped into five scenario families: Aligned, Conflict, Temporal, Insufficient Evidence, and Adversarial. Each scenario is executed for ten episodes using derived seeds from a base seed of 42, resulting in 290 replay runs.

TABLE IV
OVERALL EXPERIMENTAL OUTCOMES

Outcome	Runs	Proportion
Estimate	160	55.2%
Safe Abstention	120	41.4%
Expected configuration failure	10	3.4%
Total	290	100%

D. Evaluation Oracles and Metrics

The policy oracle determines conformance using the expected outcome and configured policy assertions for each scenario. For Estimate cases, the selected candidate state is checked against the expected state. For Safe Abstention cases, the required abstention conditions and reason-code requirements are checked according to scenario policy.

The diagnostic oracle is independent of the policy assertion and compares selected Estimates against synthetic ground truth maintained by the external test harness. It is used to identify potentially incorrect outcomes and diagnose limitations in the implementation method. It does not redefine policy conformance.

Policy conformance is calculated as

$$PC = \frac{N_{\text{policy-pass}}}{N_{\text{runs}}} \times 100\%. \tag{10}$$

Estimate diagnostic correctness is calculated only for runs producing substantive Estimates:

$$EC = \frac{N_{\text{truth-correct}}}{N_{\text{Estimate}}} \times 100\%. \tag{11}$$

Synthetic ground truth is not provided to ParkEase during reasoning. It is maintained by the external harness and used only after execution for diagnostic comparison.

VI. RESULTS

A. Overall Conformance

The final conformance batch consists of 290 replay runs across 29 scenario definitions. All runs satisfied the applicable policy assertions, resulting in an overall policy conformance rate of

$$PC = \frac{290}{290} \times 100\% = 100\%. \tag{12}$$

This result indicates behavioral conformance under the tested scenarios; it does not indicate 100% prediction accuracy or field-level correctness.

B. Overall Outcome Distribution

Across the 290 runs, the implementation produced 160 Estimates and 120 Safe Abstentions. Ten runs corresponded to the expected configuration-failure condition in INS-05. The distribution demonstrates that the operational loop includes an explicit non-estimation path when evidence conditions are insufficient.

TABLE V
CONFORMANCE AND DIAGNOSTIC MEASURES

Measure	Num.	Den.	Result	Interpretation
Policy Conformance	290	290	100%	Configured policy behavior
Estimate Correctness	120	160	75%	Synthetic diagnostic match
Safe Abstention	120	290	41.4%	Runs without substantive Estimate

C. Diagnostic Correctness

Among the 160 runs that produced an Estimate, 120 matched the synthetic ground truth and 40 did not:

$$EC = \frac{120}{160} \times 100\% = 75\%. \tag{13}$$

This result is interpreted only as a diagnostic measurement against the constructed scenarios. It is not a field-accuracy estimate and does not establish superiority over alternative reasoning or evidence-fusion methods.

D. Diagnostic Discrepancy Concentration

The 40 diagnostic discrepancies were not uniformly distributed across the scenario suite. They were concentrated in selected conflict and adversarial scenarios, particularly CON-02, ADV-03, ADV-04, and ADV-08. These cases identify limitations in the current implementation method and provenance model rather than non-conformance of the implementation to its configured policy.

The observed issues include conflict weighting, dependency recognition, strong-single-source behavior, and dependency handling based on artifact references. These findings motivate further method-level investigation rather than a change to the framework-level conformance interpretation.

E. Expected Configuration Failures

The ten expected configuration-failure runs were associated with the intentionally incomplete INS-05 configuration. These cases are reported separately from Estimates and Safe Abstentions because the configured preconditions prevented normal evaluation. They are therefore not interpreted as reasoning failures or prediction errors.

VII. DISCUSSION

A. Policy Conformance Does Not Imply Accuracy

The 100% policy conformance rate demonstrates that the evaluated ParkEase implementation consistently satisfied the specified behavioral assertions across the 290 replay runs. This provides evidence that the AMEF operational policy could be translated into observable and externally testable system behavior. However, policy conformance should not be interpreted as prediction accuracy. The diagnostic analysis shows that 40 of the 160 produced Estimates disagreed with synthetic ground truth, yielding 75% diagnostic correctness.

B. Framework Behavior and Reasoning Quality

The separation between conformance and diagnostic correctness supports the distinction between AMEF and the versioned ParkEase evaluation/reasoning method. AMEF defines the lifecycle and responsibilities through which observations become evidence and operational outcomes, whereas the specific scoring, dependency, conflict, and decision mechanisms are implementation choices. Diagnostic discrepancies therefore identify areas for improving a particular instantiation without by themselves invalidating the framework abstraction.

C. Diagnostic Errors Reveal Method Limitations

The concentration of diagnostic discrepancies in selected conflict and adversarial scenarios provides more information than the aggregate 75% rate alone. The observed cases indicate sensitivity to particular evidence configurations, including conflicting contributions, coordinated or dependent sources, and situations in which a strong individual source can disproportionately influence the outcome. Future method revisions should therefore investigate richer representations of source dependence and conflict.

D. Safe Abstention

The 120 Safe Abstention outcomes demonstrate that the operational loop is not required to produce a substantive Estimate for every observation set. The reasoning stage can terminate without an Estimate when configured evidence sufficiency conditions are not met. The present evaluation does not establish that Safe Abstention improves real-world decision outcomes because no field deployment or human decision study was conducted. Its demonstrated contribution is narrower: abstention can be represented as an explicit and testable operational outcome.

E. Traceability

Traceability distinguishes the AMEF approach from a purely numerical fusion pipeline. Prior work on explainability, causability, and AI traceability highlights the importance of making AI-supported outcomes inspectable and understandable to users [15], [16], [17]. Each reasoning episode can be associated with producer, observation, evidence, evaluation, and contribution records. This structure provides an audit path for examining how an Estimate or Safe Abstention was produced. In the current study, traceability is demonstrated as an implementation capability rather than evaluated through a separate human-centered explainability study.

F. Implications

The results suggest that an evidence-centric architecture can provide a boundary between heterogeneous sensing systems and downstream operational reasoning. Producers can be changed or added without requiring the framework to redefine the epistemic status of their outputs. Evaluation and reasoning methods can also be versioned independently from the framework lifecycle, facilitating controlled experimentation with alternative reasoning methods.

VIII. Threats to Validity and Limitations

Synthetic evaluation: The evaluation uses controlled synthetic scenarios rather than field observations collected from a deployed parking environment. Results therefore establish behavior under defined test conditions and cannot be directly interpreted as real-world parking accuracy.

Limited domain scope: The reference implementation is instantiated for smart parking and uses a specific Parking Zone state model. Transferability to other operational domains was not empirically evaluated.

Baseline reasoning method: Diagnostic results are tied to the versioned ParkEase baseline evaluation and reasoning method. Alternative fusion or reasoning algorithms were not evaluated as controlled baselines, so comparative superiority cannot be established.

Learning Loop: The Learning Loop is part of the AMEF design scope but was outside the executed experimental scope. The present results therefore evaluate the Operational Loop rather than adaptive learning behavior over time.

Synthetic ground truth: Diagnostic comparison relies on synthetic ground truth defined by the external test harness. Although the ground truth is isolated from the system under test, its representativeness is constrained by scenario design.

External validation and implementation independence: The study does not include independent external validation or an independently developed implementation/conformance suite. The reference implementation remains a research prototype rather than a claim of production readiness.

IX. Conclusion

This paper presented AMEF, an evidence-centric framework for structuring heterogeneous information in operational reasoning systems. AMEF distinguishes Observation, Evidence, Estimate, and Knowledge, and assigns explicit responsibilities to Qualification, Evaluation, Reasoning, and Learning. The framework also treats Safe Abstention and traceability as explicit operational capabilities.

AMEF was instantiated in ParkEase as a reference implementation for smart parking and evaluated through an externally controlled conformance test suite. The final evaluation comprised 29 scenario definitions and 290 replay runs. All tested runs satisfied their applicable policy assertions, resulting in 100% policy conformance. The implementation produced 160 Estimates and 120 Safe Abstentions, while ten runs corresponded to an expected configuration-failure condition. Among the 160 Estimates, 120 agreed with the synthetic ground truth and 40 did not, resulting in 75% diagnostic correctness.

The results demonstrate that policy conformance and diagnostic correctness provide complementary forms of evidence. The former establishes that the evaluated implementation follows specified operational behavior, while the latter exposes limitations in the current reasoning method under selected conflict and adversarial evidence configurations. Future work will investigate external validation, cross-domain instantiation, independent implementations and conformance suites, richer dependency and conflict modeling, field validation, configuration versioning, and evaluation of the AMEF Learning Loop.

References

[1] G. Amato, F. Carrara, F. Falchi, C. Gennaro, C. Meghini, and C. Vairo, "Deep learning for decentralized parking lot occupancy detection," *Expert Systems with Applications*, vol. 72, pp. 327–334, 2017.

[2] J. M. Carrera García, J. Recas Piorno, and M. Guijarro Mata-García, "Expert system design for vacant parking space location using automatic learning and artificial vision," *Multimedia Tools and Applications*, vol. 81, no. 27, pp. 38 661–38 683, 2022.

[3] R. Ke, Y. Zhuang, Z. Pu, and Y. Wang, "A smart, efficient, and reliable parking surveillance system with edge artificial intelligence on iot devices," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 8, pp. 4962–4974, 2021.

[4] S. Y. Siddiqui, M. A. Khan, S. Abbas, and F. Khan, "Smart occupancy detection for road traffic parking using deep extreme learning machine," *Journal of King Saud University–Computer and Information Sciences*, vol. 34, no. 3, pp. 727–733, 2022.

[5] J. K. Suhr and H. G. Jung, "Sensor fusion-based vacant parking slot detection and tracking," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 1, pp. 21–36, 2014.

[6] S. Zimmermann, C. Schulz, A. Hein *et al.*, "Why specialized service ecosystems emerge—the case of smart parking in germany," *Information Systems Frontiers*, vol. 27, no. 2, pp. 585–604, 2025.

[7] Z. Wang and F. Xiao, "An improved multi-source data fusion method based on the belief entropy and divergence measure," *Entropy*, vol. 21, no. 6, p. 611, 2019.

[8] B. Shen, Y. Liu, and J.-S. Fu, "An integrated model for robust multisensor data fusion," *Sensors*, vol. 14, no. 10, pp. 19 669–19 686, 2014.

[9] X. Xiang, K. Li, B. Huang, and Y. Cao, "A multi-sensor data-fusion method based on cloud model and improved evidence theory," *Sensors*, vol. 22, no. 15, p. 5902, 2022.

[10] J.-B. Yang, J. Liu, J. Wang, H.-W. Sii, and H.-W. Wang, "Belief rule-base inference methodology using the evidential reasoning approach—rimer," *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, vol. 36, no. 2, pp. 266–285, 2006.

[11] C. K. Chow, "On optimum recognition error and reject tradeoff," *IEEE Transactions on Information Theory*, vol. 16, no. 1, pp. 41–46, 1970.

[12] Y. Geifman and R. El-Yaniv, "Selective classification for deep neural networks," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[13] ——, "Selectivenet: A deep neural network with an integrated reject option," in *Proceedings of Machine Learning Research*, vol. 97, 2019, pp. 4876–4884.

[14] V. Franc, D. Průša, and V. Voráček, "Optimal strategies for reject option classifiers," *Journal of Machine Learning Research*, vol. 24, 2023.

[15] T. Schrills and T. Franke, "How do users experience traceability of ai systems? examining subjective information processing awareness in automated insulin delivery systems," *ACM Transactions on Interactive Intelligent Systems*, vol. 13, no. 4, pp. 1–34, 2023.

[16] D. Shin, "The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable ai," *International Journal of Human-Computer Studies*, vol. 146, p. 102551, 2021.

[17] D. Kim, Y. Song, S. Kim, S. Lee, Y. Wu, J. Shin, and D. Lee, "How should the results of artificial intelligence be explained to users? research on consumer preferences in user-centered explainable artificial intelligence," *Technological Forecasting and Social Change*, vol. 188, p. 122343, 2023.

[18] J. Hwang, A. Aziz, N. Sung, A. Ahmad, F. Le Gall, and J. Song, "Autocon-iot: Automated and scalable online conformance testing for iot applications," *IEEE Access*, vol. 8, pp. 43 111–43 121, 2020.

[19] M. K. Mufida, A. Snoun, J. Sarkis, A. Ait El-Cadi, T. Delot, and M. Trépanier, "Circular economy practices for data management," *Engineering Proceedings*, vol. 97, no. 1, p. 34, Jun. 2025.