*sustainability*

MDPI

*Article*

# Enhancing Transportation Supply Chain Resilience Through HR Practices: A Task-Typed Reasoning Module-Enhanced RAG Expert System Framework

**Jun Ren * and Omolomo Odunayo Tobora**

School of Engineering and Built Environment, Liverpool John Moores University, Liverpool L3 3AF, UK; o.o.tobora@2023.ljmu.ac.uk

* Correspondence: j.ren@ljmu.ac.uk

**Abstract**

The transportation sector faces growing supply chain disruptions driven by workforce shortages, digital skills gaps, and weak cultural readiness for risk. Human Resource Management (HRM) offers a credible path toward stronger supply chain resilience (SCR), yet existing decision support tools lack the formal reasoning and auditability that systematic HR risk assessment requires. This paper proposes a Task-Typed Reasoning Module-Enhanced Retrieval-Augmented Generation (RAG) Expert System framework for HR-driven risk assessment in transportation supply chains. The framework employs a six-layer architecture integrating four core components: a Task-Aware RAG layer for knowledge extraction from heterogeneous HR documents, a Task-Routed Evidence Extractor for risk factor identification and source reliability scoring, a Task-Based Reasoning Core applying weighted Mamdani fuzzy inference, and a Task-Guided Synthesis Module using Dempster–Shafer evidential reasoning for risk profiling. Task-Typed Reasoning Modules (TTRMs) organise reasoning into reusable, adaptable units covering Likelihood, Impact, Vulnerability, Mitigation, and Prioritisation, refined through an expert-gated feedback loop. Applied to an illustrative UK transportation logistics case study, the framework ranked HR-driven workforce risks, quantified uncertainty through belief-plausibility intervals, and generated traceable recommendations linked to four HRM-SCR dimensions. As a proof-of-concept demonstration, the framework provides a conceptual foundation for HR risk management in transportation, extending expert system reasoning through adaptive AI to offer mathematically grounded, traceable outputs for practitioners and researchers; empirical validation in live organisational settings remains a necessary next step. By formalising the human dimension of resilience, the framework contributes toward socially, economically, and environmentally sustainable transportation supply chains, aligning HR risk management with the sustainable development objectives examined in this study.

**Keywords:** transportation supply chain resilience; strategic HR practices; logistics and operations management; AI-based expert systems; retrieval-augmented generation; task-typed reasoning modules; sustainable human resource management



Academic Editor: Bernardo Nicoletti

Received: 8 July 2026
Revised: 23 July 2026
Accepted: 29 July 2026
Published: 7 August 2026



## 1. Introduction

The global transportation sector moves goods across industries and geographies. It also remains deeply vulnerable to disruptions from geopolitical tensions, pandemics, extreme weather events, and technological change [1,2]. These disruptions do not stay

contained. A shortage of qualified drivers can paralyse distribution networks across multiple industries at once, as the chronic Heavy Goods Vehicle (HGV) driver shortage in the UK demonstrated in the years following Brexit and the COVID-19 pandemic [3,4]. Researchers and practitioners increasingly recognise the sector's strategic importance to supply chain resilience (SCR). Yet most existing resilience frameworks still concentrate on structural and logistical dimensions: network redundancy, inventory buffering, and supplier diversification. The human dimension receives comparatively little attention [5].

Human resource management (HRM) practices are an underexplored lever for SCR. Workforce capability, adaptability, engagement, and leadership succession all shape how well an organisation can anticipate, absorb, and recover from disruptions [6,7]. Recent empirical work confirms that targeted HRM interventions, including cross-functional training, flexible work arrangements, engagement initiatives, risk culture development, and digital technology adoption, measurably improve workforce agility and SCR outcomes [8]. The evidence is there. What is missing is a systematic, computationally rigorous method for assessing and prioritising HR-driven risks in transportation supply chains. HR risk factors are marked by linguistic ambiguity, incomplete measurement, and conflicting expert opinion. These are precisely the conditions where classical probabilistic methods fall short, and where formal uncertainty reasoning is most needed.

The human dimension of resilience also carries a direct sustainability dimension, one that existing SCR research treats largely implicitly. Workforce shortages, skills gaps, and low engagement are not only operational vulnerabilities; they are symptoms of eroding social sustainability in transportation employment, where precarious staffing and burnout undermine the workforce capability, adaptability, and engagement that resilience depends on [6,7]. The economic consequences compound this: firms that fail to invest in workforce development absorb the recurring costs of turnover, understaffing, and disruption rather than the one-time cost of prevention. Disruptions are also frequently resolved through rerouted freight, expedited shipping, and idle fleet capacity, so the human dimension of resilience carries an environmental cost as well, one rarely made explicit in frameworks that treat HR risk as a purely internal management concern rather than a lever for sustainable transportation operations. Addressing HR-driven risk systematically, as this paper sets out to do, therefore serves social, economic, and environmental sustainability jointly, not resilience in isolation.

Expert systems have a long tradition in risk assessment for complex socio-technical systems. They offer the reasoning transparency that purely data-driven approaches cannot [9]. The catch is a well-known knowledge acquisition bottleneck. Encoding domain knowledge manually is resource-intensive, brittle when the domain shifts, and unable to draw on the large volumes of unstructured knowledge locked in operational documents, training manuals, workforce plans, and incident reports [10]. Retrieval-Augmented Generation (RAG) addresses this directly. It allows large language models (LLMs) to draw on dynamically retrieved, domain-specific knowledge at inference time, which reduces, though does not eliminate, hallucination and improves factual grounding [11]; grounding lowers the incidence of unsupported claims but does not itself certify that a given retrieved claim is accurate, a distinction the framework's evidence-reliability scoring (Formal BPA Construction, Section 3.6) is designed to preserve rather than collapse. Existing RAG systems, however, lack the mathematical rigour, including fuzzy membership functions, evidential combination rules, and structured uncertainty quantification, that safety-critical and compliance-sensitive risk assessment domains require.

This paper addresses that gap. It proposes a Task-Typed Reasoning Module-Enhanced RAG Expert System framework designed specifically for HR-driven risk assessment

in transportation supply chain resilience. The framework makes four contributions to the knowledge.

A narrower question remains even once RAG and formal fuzzy-evidential reasoning are combined: why organise that reasoning around Task-Typed Reasoning Modules rather than existing modular AI architectures, for example multi-agent LLM systems with fixed tool sets, or classical modular expert systems with independently maintained rule blocks? Both alternatives fall short in ways specific to HR risk assessment. Multi-agent LLM architectures decompose tasks but typically leave each agent's internal reasoning opaque and its performance unbenchmarked against expert judgement, which is difficult to reconcile with compliance-sensitive HR decisions. Modular expert systems decompose rule bases by module but still require each module's rules to be manually authored and revised by a knowledge engineer whenever evidence patterns shift, reproducing the same knowledge-acquisition bottleneck the RAG layer is designed to remove. Task-Typed Reasoning Modules address both gaps at once: each TTRM is a self-contained, benchmarked reasoning unit, which addresses the opacity problem, whose inference template weights are learned from expert-validated outcomes rather than hand-authored, which addresses the maintenance problem, while remaining subject to expert sign-off before any change is committed (Section 3.10). Neither alternative architecture offers this combination of modular decomposition, empirical benchmarking, and expert-gated learning together, which motivates TTRMs as a distinct organising principle rather than an existing modular AI or expert-system pattern adopted unmodified.

First, it provides what is, to the best of the authors' knowledge, the first formal integration of Task-Aware RAG with weighted Mamdani fuzzy inference and Dempster–Shafer evidential reasoning for HR risk assessment in transportation. This means the system can process unstructured HR documents and produce formally grounded, traceable risk profiles with explicit uncertainty quantification.

Second, it introduces Task-Typed Reasoning Modules (TTRMs) as the organising principle for adaptive expert system reasoning. Unlike static rule bases, TTRMs are reusable, version-controlled, benchmarked reasoning units that evolve through an expert-gated feedback loop. The organising architecture is domain-transferable: the same five-TTRM structure (Likelihood, Impact, Vulnerability, Mitigation, Prioritisation) applies whether the risk factor is driver turnover, digital skills gaps, or leadership succession. Transferring the framework to a new domain, however, requires reconfiguring each TTRM's decision patterns, fuzzification parameters, and inference template weights against domain-specific evidence; swapping the input document corpus alone is not sufficient.

Third, it operationalises the four HRM-SCR dimensions identified in the supply chain resilience literature (Talent Management and Workforce Capability (TMGWC), Workforce Flexibility and Employee Engagement (WFE), Risk Management and Culture (RMC), and Technology and Communication (COTEC)) [8] as structured evidence categories within the Task-Routed Evidence Extractor, directly connecting HRM theory to computational risk assessment practice.

Fourth, it demonstrates through a transportation logistics case study that the framework produces risk rankings, uncertainty intervals, and full traceability chains that support both managerial decision-making and regulatory compliance.

This paper builds directly on two companion works. Tobora et al. [8] established the four HRM-SCR dimensions (TMGWC, WFE, RMC, COTEC) that structure the evidence categories used here. Ren, Jenkinson, and Tobora [12] demonstrated a RAG-augmented expert system for offshore safety using metadata-filtered retrieval to produce auditable explanations for an existing rule-based system. The present paper extends both: it operationalises the HRM-SCR dimensions as formal evidence categories within a computational

pipeline, and it advances the RAG-expert system approach by deploying retrieval earlier in the reasoning process, before formal inference rather than after it, and by replacing the static rule base with adaptive Task-Typed Reasoning Modules.

The remainder of this paper is organised as follows. Section 2 reviews the literature on HRM practices in supply chain resilience, RAG architectures, fuzzy-evidential reasoning, and modular AI reasoning. Section 3 presents the complete framework specification. Section 4 applies the framework to a case study in UK transportation logistics. Section 5 discusses results and implications. Section 6 addresses limitations and future directions. Section 7 concludes.

## 2. Literature Review

### *2.1. HRM Practices and Supply Chain Resilience in Transportation*

HRM practices in logistics and transportation are genuinely difficult. They require simultaneous attention to recruiting, retaining, and developing a workforce that can handle the shifting demands of supply chain resilience [4]. The broader question of HRM's contribution to organisational performance is well-covered in the literature [13]. Its direct impact on SCR within transportation, however, is far less examined. Existing research tends to treat HR as secondary to structural and technological resilience strategies [5,14]; organisational resilience more broadly, meanwhile, has its own established theoretical base [15].

Supply chain resilience is typically defined as a supply chain's capacity to adapt to change. This covers both reactive measures to withstand disruptions and proactive strategies for recovery [2,5,15]. A meaningful gap sits at the intersection of HRM practices and SCR. Despite solid evidence on the importance of employee training, engagement, and adaptability during crises, studies rarely connect HRM practices to SCR outcomes in any direct or quantifiable way [14,16,17]. In transportation, this gap matters more than in many other sectors. Workforce characteristics, driver skills, shift flexibility, and engagement levels directly determine whether operations can continue under disruption.

Tobora et al. [8] offer a framework that connects HRM practices to SCR outcomes across four dimensions: (1) Talent Management and Workforce Capability (TMGWC), covering talent acquisition, training and development, and leadership succession; (2) Workforce Flexibility and Employee Engagement (WFE), addressing flexible work arrangements, wellbeing, health and safety, and motivation; (3) Risk Management and Culture (RMC), incorporating diversity, risk culture development, and proactive change management; and (4) Technology and Communication (COTEC), covering digital skills, communication platforms, and collaborative tools. This gives a theoretically grounded and practically usable structure for understanding how HRM practices collectively support SCR in transportation.

Talent management is one of the most studied components of organisational resilience. Cross-functional training equips employees to cover multiple roles during disruptions, reducing dependency on any individual or team [18–20]. Leadership development supports continuity and sound decision-making in a crisis [7]. Workforce flexibility and employee engagement are equally important. Flexible work arrangements have shown real value in maintaining operational continuity under pressure [21,22], and employee recognition programmes alongside open communication channels help sustain engagement when disruptions create uncertainty [23].

A risk-aware culture helps organisations spot vulnerabilities before they grow, embedding resilience into everyday practice [24]. Technology and communication act as enabling dimensions. Digitalisation, predictive analytics, and real-time collaboration tools give employees the information they need to respond quickly and reduce the impact of disruptions [25,26]. At first glance the existing literature seems to cover these dimensions

well. Yet no framework currently provides a formal, computationally rigorous method for assessing and prioritising HR-driven risks in transportation supply chains. That gap motivates the integrated fuzzy-evidential RAG approach proposed here.

### *2.2. Retrieval-Augmented Generation in Expert Systems*

Retrieval-Augmented Generation marks a significant change in how LLM applications work. Models can now condition their outputs on external knowledge retrieved at inference time [27]. This approach addresses key limitations of parametric knowledge storage, including factual inconsistency and domain inflexibility, while reducing, though not eliminating, hallucination through grounding in source documents [11]. Recent developments in RAG architectures have explored hybrid retrieval strategies that combine dense semantic search with sparse lexical matching [11]. Graph-based RAG approaches structure domain knowledge as graphs for dynamic retrieval, enabling multi-hop reasoning across document boundaries [28]. Knowledge Graph-extended RAG systems integrate structured knowledge representations with unstructured text, providing more explainable question answering [29]. Despite these advances, combining RAG with formal expert system reasoning remains underexplored. Current RAG systems do well at information extraction and synthesis, but they lack the mathematical rigour that safety-critical applications require, where every decision must be traceable and auditable [10]. Ren et al. [12] demonstrated one approach: using metadata-filtered RAG to produce auditable explanations for an offshore safety expert system. That study used RAG as an explanatory tool after the main reasoning process. The present paper uses the same mechanism differently. RAG is deployed earlier in the pipeline, drawing in relevant knowledge before the formal reasoning begins. This difference in placement matters substantively, not merely procedurally: post hoc explanatory RAG, as in Ren et al. [12], cannot influence which risk factors are identified or how they are weighted, because the expert system's reasoning is already fixed by the time RAG output arrives; pre-inference RAG, as used here, shapes the evidence base the reasoning core itself operates on, which raises the evidence-reliability and hallucination-containment requirements addressed in the Formal BPA Construction step (Section 3.6), but also allows the system to surface risk factors and evidence patterns that a fixed rule base was not designed to look for. Section 2.5 develops this and related comparisons, against fuzzy-evidential risk models without RAG and HRM-SCR frameworks without a computational reasoning engine, into a structured comparison with the present framework (Table 1). A closely related application of the same skill-enhanced approach to fuzzy-evidential reasoning has since been developed for offshore accident analysis, combining fuzzy inference, Dempster–Shafer evidential reasoning, and expert-gated skill adaptation in a parallel domain [30].

### *2.3. Fuzzy-Evidential Reasoning Approaches*

Fuzzy logic provides a mathematical framework for reasoning with imprecise and uncertain information, which makes it well-suited to risk assessment where linguistic variables dominate [31]. Mamdani fuzzy inference systems are widely used in decision support, applying IF-THEN rules with membership functions to map crisp inputs to fuzzy outputs [32]. Dempster–Shafer (D-S) theory extends classical probability by representing ignorance and conflict between information sources explicitly [33,34]. Its ability to distinguish between belief and plausibility makes it particularly valuable when evidence is incomplete or contested [35–37]. Related work on fuzzy Bayesian networks has extended this tradition to offshore risk analysis [38]. Combining fuzzy logic with D-S evidential reasoning offers an effective approach for handling multiple sources of uncertainty: fuzzy membership values feed into belief mass function construction, while D-S theory provides a principled method for aggregating conflicting evidence.

**Table 1.** Structured comparison of the proposed framework against related approaches identified in the literature review.

| Approach | Knowledge Acquisition | Formal Uncertainty and Adaptivity | Traceability/Governance | Applied Domain |
|---|---|---|---|---|
| Static fuzzy-evidential expert systems | Manual rule elicitation by a knowledge engineer | Formal (fuzzy + Dempster–Shafer) but fixed at design time; no learning from outcomes | Rule-level traceability only; no expert sign-off record | Offshore/industrial safety [38,39] |
| LLM/RAG-only systems | Automated retrieval from unstructured documents | Narrative or point outputs; no belief-plausibility intervals; no benchmarked learning | Source citation only; no formal decision audit trail | General question-answering [13,30] |
| RAG-augmented expert system (explanatory RAG) | Automated retrieval applied after reasoning is fixed, for explanation only | Formal, inherited from the underlying expert system; no template-weight adaptation | Explanation-level traceability; RAG cannot alter what evidence entered the reasoning | Offshore safety [14] |
| HRM-SCR conceptual framework | Not applicable; conceptual, not computational | Not applicable; no formal reasoning engine | Not applicable | Transportation HRM-SCR dimensions [10] |
| TTRM-Enhanced RAG Expert System (this paper) | Automated retrieval feeding formal reasoning directly, before inference (Section 3.6) | Formal (fuzzy + D-S) and adaptive: expert-gated template-weight learning with TTRM versioning | Full trace: source → evidence → TTRM → template → output, with named expert sign-off | HR risk in UK transportation; illustrative case study only, not yet validated across other domains or organisations |

### *2.4. Modular AI Reasoning and Adaptive Expert Systems*

The concept of modular, task-specific reasoning in AI has attracted growing attention alongside the development of agent frameworks that break complex tasks into reusable capability components. The LLM-Modulo framework shows how language models can support planning by generating candidates while external verifiers check constraint satisfaction [39]. Decision Flow extends this by guiding models to reason over structured representations, enabling more transparent and auditable decision processes [40]. Clinical decision support systems have moved in a similar direction, shifting from monolithic expert systems toward modular architectures that combine specialised reasoning components [41]. This paper operationalises this modular principle as Task-Typed Reasoning Modules (TTRMs): reusable, version-controlled reasoning units organised by assessment task type rather than by risk domain, refined through an expert-gated feedback loop. Skill-based architectures of this kind allow AI systems to adjust their reasoning based on task context and accumulated experience, which static rule-based systems cannot. The TTRMs proposed here carry this principle into the domain of HR risk assessment with explicit mathematical formulation, quantified uncertainty, and end-to-end traceability.

### *2.5. Research Gap and Contribution*

The literature reveals three gaps:

- No integration of RAG with formal expert system reasoning for SCR. Existing SCR frameworks rely on manual knowledge acquisition. Existing RAG applications focus on question-answering, not on feeding structured evidence into formal reasoning systems.
- LLM-only approaches lack formal reasoning. Using an LLM directly for risk assessment cannot provide the uncertainty quantification, evidence fusion, or mathematical auditability that fuzzy inference and evidential reasoning offer.
- Expert-system-only approaches lack scalable knowledge acquisition. Traditional expert systems require all knowledge to be manually elicited and encoded. This process does not scale to the growing volume of relevant documents, data, and expert opinions.

This paper fills these gaps by proposing a framework that combines TTRMs and RAG for scalable knowledge acquisition with fuzzy inference and evidential reasoning for formal SCR assessment. The framework's five-TTRM reasoning architecture is domain-transferable in principle; applying it to a new risk domain requires re-eliciting decision

patterns, fuzzification boundaries, and template weights from that domain's own evidence base, not merely substituting the input document corpus.

Table 1 positions the proposed framework against the closest related approaches identified above, making the comparison explicit rather than leaving it to be inferred from the narrative review. The comparison is deliberately critical of the proposed framework as well as of prior work: the last column records what each approach does not provide, including the present one.

## 3. The Proposed Framework

### *3.1. Notation*

Table 2 defines the notation used throughout the framework specification in this section.

**Table 2.** Principal notation.

| Symbol | Type | Definition |
|---|---|---|
| S = (t, P, F, T, B) | Tuple | A TTRM: decision type, patterns, fuzzification config, templates, benchmarks |
| t | Categorical | Decision type in {LA, IA, VA, ME, RP} |
| P = {p1, . . ., pn} | Set | Decision patterns: evidence-to-output mappings |
| F = {(lj, μj)} | Set | Fuzzification config: linguistic terms with parametric membership functions |
| T = {($T^I$, $w^I$)} | Set | Inference templates with learned weights $w^I \in [0, 1]$ |
| B = {η, d, η($T^I$), v, date} | Tuple | Performance benchmarks: agreement rate, deviation, per-template agreement rates, version, version date |
| Ω | Set | TTRM Repository: the full set of active TTRMs |
| Ω_t | Set | TTRMs applicable at assessment stage t |
| f(S, E(f)) | Function | TTRM-evidence compatibility: $f = \alpha \cdot$type_match $+ (1 - \alpha) \cdot$coverage |
| q(f, S) | Query | Task-aware query for risk factor f under active TTRM S |
| E(f) | Structure | Evidence profile for risk factor f extracted by the grounded LLM |
| μ_l($x_0$) | Scalar ∈ [0, 1] | Fuzzy membership of input $x_0$ in linguistic term l |
| r | Scalar ∈ (0, 1] | Source reliability coefficient used in BPA construction |
| m(A) | Function | Basic probability assignment (BPA) on subset $A \subseteq \Theta$ |
| Θ | Set | Frame of discernment: {lj: j = 1, . . ., K} |
| K | Scalar ∈ [0, 1) | Dempster conflict factor: total mass assigned to empty set |
| Bel(A) | Scalar ∈ [0, 1] | Belief: lower bound of support for A |
| Pl(A) | Scalar ∈ [0, 1] | Plausibility: upper bound of support for A |
| $\varphi^I$ | Scalar | Firing strength of template $T^I$ prior to weighting (output of fuzzy inference engine) |
| $\tilde{a}^I$ | Scalar | Weighted firing strength of template $T^I$: $\tilde{a}^I = w^I \times \varphi^I$ |
| η(S) | Scalar ∈ [0, 1] | Expert agreement rate for TTRM S |
| η($T^I$) | Scalar ∈ [0, 1] | Expert agreement rate for inference template $T^I$ within TTRM S |
| d(S) | Scalar ≥ 0 | Mean deviation of TTRM S outputs from expert decisions |
| λ | Scalar ∈ (0, 1] | Learning rate for template weight update |
| $\bar{\eta}$ | Scalar | Mean per-template agreement rate (reference for weight update) |

### *3.2. Framework Overview*

The proposed framework integrates Retrieval-Augmented Generation (RAG) with a fuzzy-evidential expert system for industrial risk assessment, enhanced by Task-Typed

Reasoning Modules (TTRMs) that replace the traditional static rule base with adaptive, learned reasoning units. Figure 1 illustrates the architecture and components. Table 3 summarises the framework. The proposed framework has six layers with TTRMs embedded across Layers 2 to 5, and an expert-gated feedback loop connecting the output back to the reasoning core:

- Layer 1 (Input Layer): heterogeneous domain knowledge sources including documents, operational data, and expert judgement.
- Layer 2 (Task-Aware RAG Layer): retriever, vector database, and grounded LLM with TTRM-driven query formulation for targeted knowledge extraction.
- Layer 3 (Task-Routed Evidence Extractor): bridge between RAG outputs and expert system inputs, with a Task Router that selects the appropriate TTRM, assigns source reliability coefficients, and applies the Expert Validation Checkpoint before proceeding.
- Layer 4 (Task-Based Reasoning Core): TTRM Repository (replacing the static rule base), TTRM-driven fuzzification, formal BPA construction, and weighted Mamdani fuzzy inference.
- Layer 5 (Task-Guided Synthesis Module): Dempster–Shafer evidential reasoning with TTRM-informed weighting.
- Layer 6 (Output Layer): risk rankings, risk profiles, traceability, and TTRM trace.
- Expert-Gated Feedback Loop: captures decision outcomes, benchmarks TTRM performance, and refines TTRMs subject to expert approval.

**Table 3.** Framework layer summary.

| Layer | Name | Key Components |
|---|---|---|
| 1 | Input Sources | Documents and Manuals, Operational Data (telemetry), Expert Judgement |
| 2 | Task-Aware RAG Layer | Task-Aware Query Formulator, Vector Database, Retriever, Grounded LLM |
| 3 | Task-Routed Evidence Extractor | Risk Factor Extraction, Linguistic Variable Assignment, Reliability Coefficient Assignment, Task Router, Expert Validation Checkpoint |
| 4 | Task-Based Reasoning Core | TTRM Repository, Fuzzification Module, BPA Construction, Weighted Mamdani Inference Engine |
| 5 | Task-Guided Synthesis Module | TTRM-Informed Evidence Weighting, Dempster–Shafer Combination, Bel/Pl Computation |
| 6 | Output + Feedback Loop | Risk Rankings, Risk Profiles, TTRM Trace, Expert-Gated Benchmarking → TTRM Repository (L4) |

The design philosophy extends the original RAG-expert system principle with a third element: TTRMs handle knowledge adaptation. RAG solves the input bottleneck. The expert system provides mathematical rigour. TTRMs ensure the reasoning evolves as evidence accumulates.

### *3.3. Task-Typed Reasoning Modules*

#### 3.3.1. TTRM Definition

A Task-Typed Reasoning Module (TTRM) is a reusable, task-specific reasoning unit defined as a tuple $S = (t, P, F, T, B)$, where $t \in \{LA, IA, VA, ME, RP\}$ is the task type (Likelihood Assessment, Impact Assessment, Vulnerability Assessment, Mitigation Effectiveness, Risk Prioritisation); $P = \{p1, \ldots, pn\}$ is the set of decision patterns; $F = \{(lj, \mu j)\}$ is the fuzzification configuration; $T = \{(T^I, w^I)\}$ is the set of inference templates with learned weights; and $B = \{\eta, d, \eta(T^I), v, date\}$ is the performance benchmark, where $\eta$ is the expert agreement

rate, d is the mean deviation, $\eta(T^I)$ is the per-template agreement rate, v is the version identifier, and date is the version date. Unlike static IF-THEN rules, TTRMs are learned from accumulated decision outcomes and refined through an expert-gated feedback loop.

The modular principle draws on recent work in structured AI reasoning: the LLM-Modulo approach of separating candidate generation from constraint verification [39], and the Decision Flow framework's use of structured representations for auditable reasoning [40]. TTRMs carry this principle into risk assessment, where each module encapsulates a single assessment task type and can be applied, updated, and transferred independently of domain-specific rule content.

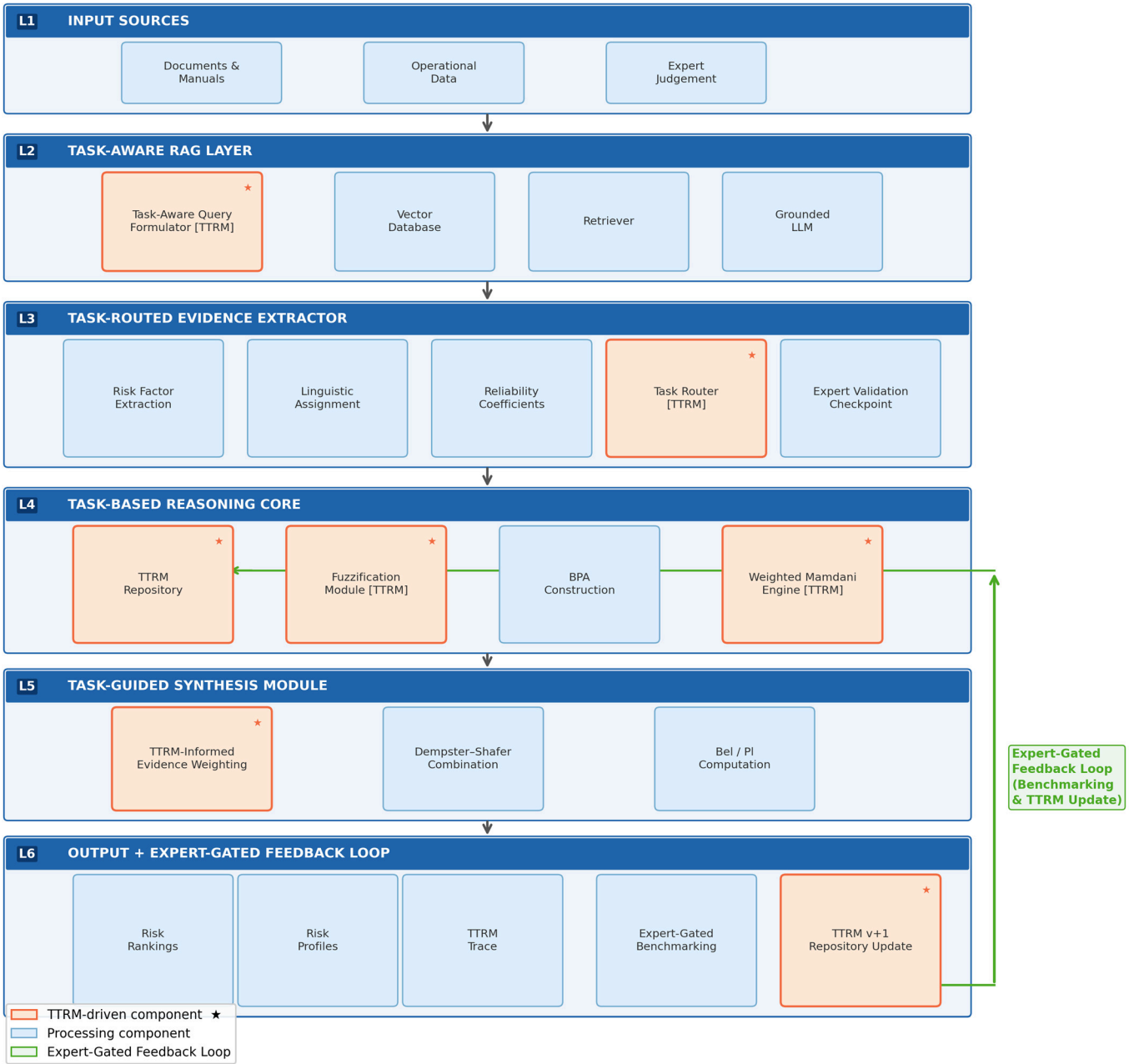


**Figure 1.** The Task-Typed Reasoning Module-Enhanced RAG Expert System Fframework. Six processing layers (L1–L6) are shown as horizontal bands. TTRM-driven components (gold, ★) are embedded across Layers 2–5, directing query formulation, evidence routing, fuzzy inference, and evidential synthesis. The Expert-Gated Feedback Loop (right) benchmarks per-TTRM and per-template performance and proposes weight updates before committing a new TTRM version to the Repository in Layer 4.

#### 3.3.2. Task-Typed Module Organisation

TTRMs are organised by decision type, not by risk category. This keeps the reasoning architecture domain-transferable: the same five-TTRM structure applies whether the risk domain is healthcare, offshore safety, or logistics. What is domain-specific is each TTRM's content, decision patterns, fuzzification boundaries, and inference template weights, which must be re-elicited and re-calibrated from domain evidence when the framework is applied outside HR risk assessment; the architecture generalises; the calibrated content does not. The framework defines five core TTRMs, as detailed in Table 4.

**Table 4.** Task-Typed Reasoning Module organisation.

| TTRM | Decision Question | Input Evidence Type | Output |
|---|---|---|---|
| Likelihood Assessment | How probable is this risk event? | Frequency data, incident reports, historical trends | Fuzzy likelihood score with belief interval |
| Impact Assessment | How severe are the consequences? | Case studies, loss records, severity benchmarks | Fuzzy impact score with belief interval |
| Vulnerability Assessment | How exposed is the system? | Control effectiveness data, gap analyses, audit findings | Fuzzy vulnerability score with belief interval |
| Mitigation Effectiveness | How well does this strategy address this risk? | Strategy evaluations, implementation records | Fuzzy effectiveness rating |
| Risk Prioritisation | Which risks need attention first? | Combined likelihood/impact/ vulnerability scores | Ranked risk list with uncertainty intervals |

#### 3.3.3. TTRMs Versus Static Rule Bases

The TTRM approach does not abandon fuzzy IF-THEN inference. Rather, it organises inference into task-type modules whose rule structures (inference templates) carry learned weights that are refined over time, and whose parameters are adapted to the specific decision context. Table 5 summarises the key differences.

**Table 5.** TTRM Repository versus static rule base.

| Property | Static Rule Base | TTRM Repository |
|---|---|---|
| Creation | Hand-crafted by knowledge engineers | Learned from RAG-extracted evidence + expert validation |
| Maintenance | Manual rewriting when domain evolves | Expert-gated feedback loop updates incrementally |
| Granularity | Individual IF-THEN rules (e.g., 245 rules) | Decision patterns per task type |
| Selection | All matching rules fire simultaneously | Task Router selects appropriate TTRM |
| Adaptability | Fixed once encoded | Evolves via benchmarked outcomes; version-controlled |
| Domain transfer | Rules rewritten for each domain | TTRMs transfer; only input evidence changes |

### *3.4. Input Layer*

- Documents and Manuals: Technical specifications, standard operating procedures, safety data sheets, regulatory guidelines, and historical incident reports.
- Operational Data: Quantitative and semi-quantitative records from operations, including sensor telemetry, production metrics, quality control measurements, and supply chain performance indicators.
- Expert Judgement: Qualitative assessments from domain practitioners, captured through structured questionnaires, semi-structured interviews, or assessment forms.

These three input types are heterogeneous and largely unstructured. Converting them into the structured formats required by an expert system (fuzzy membership values, pairwise comparison matrices, and belief mass functions) represents the knowledge acquisition bottleneck that the RAG layer addresses.

The framework as specified assumes these inputs arrive as a relatively stable, periodically refreshed document corpus, the operating mode demonstrated in the TransNova

case study (Section 4), where documents are updated on a reporting-cycle basis (annual workforce plans, periodic surveys, monthly absence logs). Large organisations generating continuous operational telemetry at high volume and velocity—for example, real-time fleet telematics, sensor streams, or transactional HR systems—require a data ingestion and synchronisation layer upstream of Layer 2, handling deduplication, schema normalisation, access control, and secure storage, that this paper does not specify. We treat this as a necessary complementary data-engineering concern (Section 6.1) rather than a component of the reasoning architecture itself: the Task-Aware RAG layer (Section 3.5) is designed to consume whatever curated, chunked, and embedded corpus that the ingestion layer produces, but does not replace it.

### *3.5. Task-Aware RAG Layer*

#### 3.5.1. Task-Aware Query Formulator

Different decision types require different types of evidence. The Task-Aware Query Formulator addresses this by constructing task-type-specific queries. Given a risk factor f and an active TTRM S, the Query Formulator generates a structured query q(f, S) that (1) specifies what type of evidence is needed (determined by S); (2) applies metadata filters to the vector database to prioritise relevant document types; and (3) formulates the query in natural language aligned with the decision question.

#### 3.5.2. Vector Database

All input documents are pre-processed, chunked, and embedded into a vector database. The chunking strategy segments documents at the paragraph or section level, preserving semantic coherence within each chunk. An embedding model converts each chunk into a dense vector representation alongside metadata: source document, section title, document type, date, and evidence category tags that enable task-aware filtering.

Reproducibility of the RAG configuration. To keep the illustrative case study self-contained, this paper specifies the RAG layer at the level of architecture and interface, namely, a paragraph-level chunking strategy, embedding-based indexing with metadata filtering, top-k semantic similarity retrieval, and a grounded LLM constrained to cite its source chunks, rather than committing to a specific language model, embedding model, or vector database product, since these are implementation choices rather than properties of the reasoning architecture and are expected to change as the underlying tooling matures. A full reproducibility specification naming the exact language model and embedding model versions, the vector database implementation, chunk size and overlap, retrieval depth k, and any re-ranking step used, together with the empirical protocol for setting the Task Router parameters $\alpha$ and $\tau$ (Section 3.6, Parameter provenance) and the source reliability coefficients $r$ beyond the design-decision rationale already given there, will be released alongside the multi-organisation validation study proposed in Section 6.2, so that the retrieval configuration can be independently audited and replicated.

#### 3.5.3. Retriever and Grounded LLM

Given the task-aware query q(f, S), the retriever performs a semantic similarity search against the vector database, returning the top-k most relevant document chunks ranked by cosine similarity. The grounded LLM receives these chunks as context and generates structured outputs under three design constraints: (1) Grounding: the LLM uses only retrieved context, which suppresses but does not eliminate hallucination; extracted content is treated as unverified evidence, not fact, until it passes through reliability-weighted BPA construction (Section 3.6) and the Expert Validation Checkpoint; (2) Structured output: the LLM produces task-specific formats, including risk factor tables, severity assessments in

linguistic terms, and categorised evidence; (3) Source citation: every extracted fact cites the source document chunk.

Note that linguistic variable assignment, the mapping of retrieved evidence to terms on the five-level scale (VL, L, M, H, VH), constitutes a reasoning step that involves the grounded LLM. The LLM performs this classification under three constraints: it operates only on retrieved context, it applies the normalisation functions defined in Section 4.5, and every assignment is subject to review at the Expert Validation Checkpoint in Layer 3. This means the mathematical grounding of Layers 4 and 5 is conditional on the quality of this upstream classification step. The Expert Validation Checkpoint exists precisely to catch misclassifications before they propagate into the formal reasoning pipeline. Human oversight at this stage is not optional. It is load-bearing.

The LLM does not make final risk decisions. All assessment decisions are made by the Task-Based Reasoning Core.

### *3.6. Task-Routed Evidence Extractor*

The Task-Routed Evidence Extractor bridges the RAG layer and the Task-Based Reasoning Core. It performs five functions: (1) risk factor extraction and categorisation following Christopher and Peck [42]; (2) linguistic variable assignment, mapping qualitative descriptions to the five-level scale (VL, L, M, H, VH); (3) source reliability assignment using coefficients r ∈ (0, 1]; (4) Task Router selection via the compatibility function:

$$S^* = \arg\max_{S \in \Omega_t} f(S,\ E(f)) \tag{1}$$

$$f(S,\ E(f)) = \alpha \cdot \text{type\_match}(t_S,\ \text{task\_type}(f)) + (1-\alpha) \cdot \text{coverage}(E(f),\ P_S) \tag{2}$$

where α ∈ (0.5, 1) prioritises type matching over coverage.

The function type_match(t_S, task_type(f)) ∈ {0, 1} returns 1 if the active TTRM's decision type t_S matches the required assessment task for risk factor f, and 0 otherwise. The function coverage(E(f), P_S) ∈ [0, 1] is defined as the proportion of evidence items in E(f) that match at least one decision pattern in the pattern set P_S:

$$\text{coverage}(E(f),\ P_S) = \frac{\left|\left\{e_i \in E(f) : \exists p_j \in P_S \text{ s.t.sim}\left(e_i, p_j\right) \geq \tau\right\}\right|}{|E(f)|} \tag{2a}$$

where a match is determined by semantic similarity exceeding a threshold τ ∈ (0, 1], set at τ = 0.75 in the default configuration. This formulation ensures that TTRM selection is driven primarily by task alignment (controlled by α > 0.5) and secondarily by evidential coverage.

The fifth function (5) is an Expert Validation Checkpoint where experts review evidence extraction and confirm TTRM selection.

Parameter provenance. The constraint α ∈ (0.5, 1) in Equation (2) reflects a design decision rather than an empirically fitted value: type matching is weighted more heavily than evidence coverage because misrouting evidence to the wrong task type (for example, treating impact evidence as likelihood evidence) is judged a more consequential error than incomplete coverage within the correct type. The semantic-similarity threshold τ = 0.75 used in the coverage function (Equation (2a)) was set on the same basis. Initial inference template weights (e.g., w1 = 0.88 for Template LI-1, Section 3.7.1) and source reliability coefficients r (Formal BPA Construction, below) were likewise initialised through expert elicitation for the TransNova case study, informed by the benchmark ranges reported in RHA [43] and CIPD [44], rather than fitted by empirical optimisation. The benchmarking loop (Section 3.10) subsequently adjusts template weights against expert-validated outcomes but does not itself re-estimate α, τ, or r, which remain configuration parameters set at

deployment. We have not conducted a full sensitivity analysis across this joint parameter space; Section Sensitivity to the H-Term Fuzzification Boundary reports an illustrative univariate sensitivity check on the H-term fuzzification boundary (baseline value 0.75, distinct from the coverage threshold τ above), and a systematic multi-parameter sensitivity study is identified as a priority for future empirical validation (Section 6.1).

Formal BPA Construction

Let $x_0$ be the crisp input, and let $\{\mu_1(x_0)$: i∈ L$\}$ be the fuzzy membership values over linguistic term set L. Given source reliability coefficient r ∈ (0, 1], the BPA is defined as:

$$m(\{l\}) = \frac{r \cdot \mu_l(x_0)}{\sum_{l'} \mu_{l'}(x_0)}, \; for\ each\ l \in L \tag{3}$$

$$m(\Theta) = I - r \tag{4}$$

This construction satisfies three required properties: (i) Σ_{A ⊆ Θ} m(A) = 1; (ii) the relative ordering of fuzzy membership values is preserved in the committed mass; (iii) the fraction (1 − r) assigned to Θ explicitly represents epistemic incompleteness. A perfectly reliable source (r = 1) commits all mass to singletons proportional to their membership values. The reliability coefficient r is not a free parameter: it is updated through the benchmarking loop (Section 3.9).

### *3.7. Task-Based Reasoning Core*

#### 3.7.1. TTRM Repository

The TTRM Repository replaces the static rule base. Each TTRM contains decision patterns (learned from accumulated expert-validated outcomes), inference templates with learned weights, fuzzification parameters, and performance benchmarks. For example, the Likelihood Assessment TTRM contains inference templates such as Template LI-1 (weight w1 = 0.88): IF primary_risk_indicator is {HIGH | MEDIUM | LOW} AND contextual_aggravating_factor is {ELEVATED | NEUTRAL | REDUCED}, THEN likelihood is {computed}. The weights are learned from the benchmarking process, not assigned by a knowledge engineer.

#### 3.7.2. Task-Driven Fuzzification

The fuzzification module converts crisp values and linguistic assessments into fuzzy membership values using parametric membership functions. Triangular fuzzy numbers (TFNs) are the default; domain-specific TTRM configurations may specify trapezoidal membership functions when input ranges have flat plateaus. The base linguistic scale is given in Table 6 and visualised in Figure 2.

**Table 6.** Base linguistic variable membership functions.

| Linguistic Term | Abbreviation | TFN (l, m, u) |
|---|---|---|
| Very Low | VL | (0, 0, 0.25) |
| Low | L | (0, 0.25, 0.5) |
| Medium | M | (0.25, 0.5, 0.75) |
| High | H | (0.5, 0.75, 1.0) |
| Very High | VH | (0.75, 1.0, 1.0) |

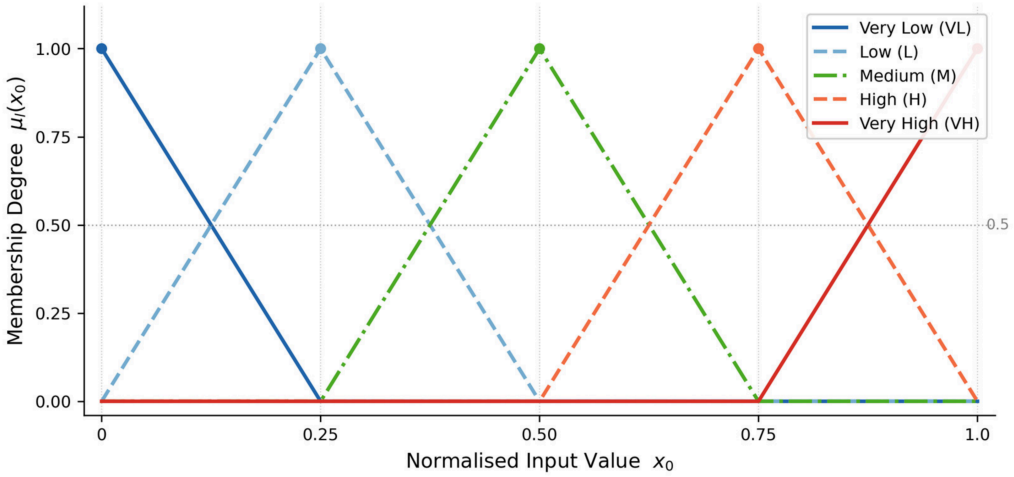


**Figure 2.** Triangular fuzzy membership functions for the five-level linguistic scale (Table 6). VL and VH are shoulder functions; L, M, and H are standard triangular functions. Adjacent functions intersect at μ = 0.5, governing partial membership assignment and BPA construction.

Figure 2 plots these five functions over the normalised input domain [0, 1]. Each function reaches its modal value (μ = 1.0) at the midpoint of its support interval. Adjacent functions overlap at μ = 0.5, producing the cross-over points that govern partial membership assignment. VL and VH are implemented as left and right shoulder functions, respectively, reflecting the absence of an upper bound on Very Low and a lower bound on Very High within the normalised scale. The overlap structure ensures that any normalised input activates at most two linguistic terms simultaneously, which constrains BPA construction to at most two non-zero singleton masses before the residual mass m(Θ) is applied.

#### 3.7.3. Task-Driven Fuzzy Inference Engine

The fuzzy inference engine applies the active TTRM's inference templates to fuzzified inputs, retaining the Mamdani min-implication operator. The key enhancement is weighted template aggregation. Let T1, …, Tn be the inference templates that fire, with learned weights $w^I$ and firing strengths $\varphi^I$. The weighted firing strength is:

$$\overset{\sim}{a}^I = w^I \cdot \varphi^I \tag{5}$$

$$\mu_{\text{out}}(z) = \max_i \min\left(\overset{\sim}{a}^I,\ \mu_{C^I}(z)\right),\ i = 1, \ldots,\ n \tag{6}$$

$$\hat{z} = \frac{\int \text{dz}}{\int \mu_{\text{out}}(z)\text{dz}}\ [\text{centroid defuzzification}] \tag{7}$$

When all template weights equal 1, the formulation reduces to standard Mamdani inference, ensuring backward compatibility with the original framework.

### *3.8. Task-Guided Synthesis Module: Evidential Reasoning*

For risk assessment, the frame of discernment defines possible risk levels: Θ = {Very Low, Low, Medium, High, Very High}. In the TTRM-enhanced framework, the Risk Prioritisation TTRM informs aggregation weights based on benchmarked performance. Given two independent BPAs m1 and m2, Dempster's rule combines them:

$$m_{12}(A) = \left[\frac{1}{1-K}\right] \cdot \sum_{B \cap C = A} m_1(B) \cdot m_2(C) \tag{8}$$

$$K = \sum_{B \cap C = \varnothing} m_1(B) \cdot m_2(C) \tag{9}$$

Belief and plausibility are computed for each risk level A as:

$$\mathrm{Bel}(A) = \sum_{B \subseteq A} m(B) \tag{10}$$

$$\mathrm{Pl}(A) = \sum_{B \cap A \neq \varnothing} m(B) \tag{11}$$

### *3.9. Output*

The framework produces four types of output: (1) Risk rankings: each risk factor ranked by its combined belief-plausibility assessment with uncertainty intervals; (2) Risk profiles: the full belief distribution across risk levels for each factor; (3) Traceability: every output links back through the synthesis module, inference engine, TTRM Repository, and evidence extractor to original source documents; (4) TTRM trace: a record of which TTRM was invoked, which inference templates fired, and what weights were applied.

### *3.10. Expert-Gated Feedback and Benchmarking Loop*

The feedback loop transforms the expert system from a static reasoning engine into an adaptive decision support system. After producing a risk assessment, the domain expert's actual decision (accept, modify, or override) is recorded. The Benchmarking Engine computes per-TTRM and per-template performance metrics:

$$\eta(S) = \frac{\left|\left\{a \in A_S : \left|\hat{z}_{\mathrm{sys}}(a) - \hat{z}_{\mathrm{exp}}(a)\right| \leq \varepsilon\right\}\right|}{|A_S|} \tag{12}$$

$$d(S) = \left(\frac{1}{|A_S|}\right) \cdot \sum_{a \in A_S} \left|\hat{z}_{\mathrm{sys}}(a) - \hat{z}_{\mathrm{exp}}(a)\right| \tag{13}$$

The proposed weight update for template $T^I$ is:

$$w'^I = w^I \cdot \left[1 + \lambda \cdot \left(\eta\left(T^I\right) - \overline{\eta}\right)\right] \tag{14}$$

where $\overline{\eta}$ is the mean per-template agreement rate and $\lambda \in (0, 1]$ is the learning rate. All updates are proposals, not automatic. Domain experts review and approve or reject each refinement before commitment to the TTRM Repository. Each approved update creates a new TTRM version with a full change log.

Governance and Accountability Model

The expert-gated feedback loop described above is deliberately structured to support formal AI governance, not merely iterative accuracy improvement. Three elements combine to provide accountability: model lineage, decision ownership, and audit traceability. Model lineage is maintained through TTRM versioning: every approved template-weight update creates a new, uniquely identified TTRM version (for example, S_LA v1.0 to v1.1, Section 4.8) with a permanent change log recording what changed, why, and under whose approval, so that any risk output can be traced to the exact TTRM version that produced it. Decision ownership is recorded at the Expert Validation Checkpoint (Section 3.6) and the benchmarking gate above: each accept, modify, or override decision is attributed to a named domain expert and timestamped, distinguishing system-proposed assessments from human-confirmed ones. Audit traceability is provided by the TTRM trace (Layer 6, Section 3.9), which links every risk output back through the synthesis module, the inference templates and weights applied, the source documents cited, and the expert decisions taken at each checkpoint.

Together, these mechanisms answer the core governance questions of who decided, on what evidence, under which model version, and with what oversight. They do not,

however, constitute a complete AI governance programme: the framework does not itself specify data retention policy, access control, or which named role holds sign-off authority within a given organisation, nor does it map onto any specific regulatory regime, since financial-services or critical-infrastructure AI governance requirements differ materially from those applicable to HR decision support tools. Embedding the framework within an organisation's existing AI governance policy, and formally mapping its audit trail onto sector-specific compliance requirements, is an organisational implementation task rather than a property of the reasoning architecture itself.

### *3.11. Formal Properties*

**Proposition 1 (Backward Compatibility).** *When all inference template weights are uniform ($w^I = 1$ for all i) and all source reliability coefficients equal unity (r = 1), the TTRM-Enhanced RAG Expert System reduces to a standard Mamdani fuzzy inference system with Dempster's rule applied to membership-proportional BPAs.*

Proof sketch. Setting $w^I = 1$ for all i yields $\tilde{a}^I = \alpha^I$, recovering standard Mamdani max-min aggregation across all firing templates without differential weighting. Setting r = 1 yields $m(\Theta) = 0$ and $m(\{l\}) = \mu_l(x_0)/\Sigma_{l'} \mu_{l'}(x_0)$, committing all belief mass to singletons in proportion to their normalised membership values, with no residual mass assigned to epistemic incompleteness. Dempster's combination rule then operates on these singleton-only BPAs from each source. The result is mathematically equivalent to applying D-S combination directly to normalised membership value vectors: a degenerate case in which the BPA carries no uncertainty beyond the membership value distribution itself. The full TTRM framework recovers this degenerate case exactly when conditions are ideal, confirming that the enhancements add capability without sacrificing compatibility with the established mathematical foundation.

## 4. Case Study: HR Risk Assessment in UK Transportation Logistics

### *4.1. Organisational Context*

The case study applies the TTRM-Enhanced RAG Expert System to HR risk assessment for a mid-sized UK road freight logistics operator, referred to here as TransNova Ltd. (a hypothetical name). TransNova operates a fleet of 250 heavy goods vehicles (HGVs) with a workforce of approximately 380 employees: 240 HGV drivers, 85 operations and scheduling staff, and 55 management and support personnel, generating annual revenues of approximately GBP 42 million. TransNova experienced significant operational disruption during 2020 to 2023, driven by pandemic-induced absenteeism, post-Brexit driver shortages, and accelerating digitalisation demands associated with real-time fleet tracking and e-proof-of-delivery systems. The HR risk assessment objective is to systematically identify, rank, and characterise five HR-driven risk factors threatening supply chain continuity, aligned with the four HRM-SCR dimensions of the Tobora et al. [8] framework.

TransNova is an illustrative case used to demonstrate the operational application of the proposed framework under conditions representative of a mid-sized UK road freight logistics operator. The organisational parameters (fleet size, workforce composition, and revenue) are modelled on sector norms reported in the Road Haulage Association [43] annual review. The internal evidence documents (workforce planning reports, training records, engagement surveys, and incident logs) reflect metric ranges and trends documented in RHA [43] and CIPD [44] for the UK logistics sector during 2020 to 2024. No real organisation or individual is represented.

This paper presents the TransNova assessment as an illustrative demonstration of the framework pipeline rather than an empirical validation study. The scope of the demon-

stration is intentionally focused: five risk factors, two evidence sources per factor, and a single assessment cycle. The intent is to show that the pipeline produces coherent, traceable, mathematically grounded outputs under realistic evidence conditions. Validation against real organisational data, including cases involving expert disagreement and TTRM override, is identified as the immediate priority for future work in Section 6.2.

### *4.2. Input Documents (Layer 1)*

A diverse range of documents and data sources were used as input to the framework (Layer 1). These included internal reports, survey data, industry benchmarks, and operational logs, providing both quantitative and qualitative evidence to support the risk assessment process. The specific documents are listed below:

- Workforce planning reports (2022–2024): annual headcount projections, vacancy rates, recruitment conversion metrics, and driver demographics.
- Training records and skills matrices: completion rates for HGV licence renewal, digital systems training, and cross-training uptake.
- Driver satisfaction and engagement surveys (2023): n = 187 respondents covering job satisfaction, work–life balance, management communication, and intent-to-leave.
- Absence and turnover data (2021–2024): monthly absenteeism rates, driver turnover, and exit interview analysis.
- Industry benchmarks: Road Haulage Association [43] national driver shortage statistics; CIPD [44] national employee engagement benchmarks.
- Incident and near-miss reports (2022–2024): classified by human factor causation.
- Digital systems adoption logs: usage rates for fleet telematics, route optimisation software, and electronic proof-of-delivery platforms.

### *4.3. Task-Aware Query Formulation (Layer 2)*

For each risk factor, the Task-Aware Query Formulator generates task-type-specific queries. For RF1 (Driver Workforce Shortage) under the Likelihood Assessment TTRM (S_LA), the query is:

> "*What is the frequency and trend of driver vacancy rates, recruitment failure rates, and industry-wide driver shortage indicators? Retrieve incident reports, workforce planning data, and RHA benchmarks relevant to driver supply continuity.*"

For the Impact Assessment TTRM (S_IA), the query targets documented consequences: delivery SLA failures, revenue loss, and operational disruption severity records. Metadata filters restrict retrieval to document categories appropriate to each TTRM, ensuring evidence relevance without cross-contamination between decision dimensions.

### *4.4. Evidence Extraction, Task Routing, and BPA Construction (Layer 3)*

The Task-Routed Evidence Extractor (Layer 3) performs three functions for the TransNova assessment: (1) risk factor identification and HRM-SCR dimension classification; (2) linguistic variable assignment and source reliability coefficient scoring; and (3) Task Router selection via the compatibility function (Equations (1)–(2a)), subject to Expert Validation Checkpoint confirmation. Table 7 presents the five risk factors with their HRM-SCR dimension mappings; Table 8 presents the corresponding evidence items with linguistic assessments, reliability coefficients, and assigned assessment dimensions.

**Table 7.** HR risk factors and HRM-SCR dimension mapping.

| ID | Risk Factor | HRM-SCR Dimension |
|---|---|---|
| RF1 | Driver Workforce Shortage | Talent Management and Workforce Capability (TMGWC) |
| RF2 | Cross-Functional Skills Gap | Talent Management and Workforce Capability (TMGWC) |
| RF3 | Workforce Engagement Deficit | Workforce Flexibility and Employee Engagement (WFE) |
| RF4 | Risk Culture Immaturity | Risk Management and Culture (RMC) |
| RF5 | Digital Capability Gap | Technology and Communication (COTEC) |

**Table 8.** Evidence extraction and linguistic variable assignment.

| RF | Evidence Item | Source | r | Linguistic Assessment | Dimension |
|---|---|---|---|---|---|
| RF1 | Driver vacancy rate 32% above RHA benchmark | RHA Industry Report 2023 | 0.92 | VH Likelihood | TMGWC |
| RF1 | Recruitment conversion rate 18% vs. 35% benchmark | Internal HR Report 2024 | 0.85 | H Likelihood | TMGWC |
| RF1 | 3 operational delays >4 h in Q3 2023 attributed to driver shortage | Incident Report Log | 0.88 | H Impact | TMGWC |
| RF2 | Cross-training uptake: 22% of drivers (target: 60%) | Training Records 2024 | 0.90 | H Vulnerability | TMGWC |
| RF2 | 67% of roles have single-qualified incumbent | Skills Matrix 2024 | 0.90 | VH Vulnerability | TMGWC |
| RF3 | Employee intent-to-leave: 38% (CIPD benchmark: 18%) | Engagement Survey 2023 | 0.83 | H Likelihood | WFE |
| RF3 | Job satisfaction score: 5.9/10 (CIPD median: 7.2/10) | Engagement Survey 2023 | 0.83 | M-H Impact | WFE |
| RF4 | Near-miss reporting rate: 0.8/driver/year (benchmark: 2.1) | Incident Report Log | 0.88 | H Likelihood | RMC |
| RF4 | Risk awareness training completion: 41% (target: 100%) | Training Records 2024 | 0.90 | VH Vulnerability | RMC |
| RF5 | Telematics utilisation: 54% of available features actively used | Digital Systems Log 2024 | 0.86 | M-H Vulnerability | COTEC |
| RF5 | Digital skills training completion: 31% of operations staff | Training Records 2024 | 0.87 | H Vulnerability | COTEC |

### *4.5. Fuzzification (Layer 4)*

Crisp inputs are normalised to [0, 1] from raw metrics using domain-specific functions calibrated against industry benchmarks. Detailed fuzzification for RF1 Likelihood (Source 1: RHA vacancy rate data, r1 = 0.92):

Normalised input: $x_0 = 0.82$ (vacancy rate index)

$$\mu\_H(0.82) = (1.0 - 0.82)/(1.0 - 0.75) = 0.72;\ \mu\_VH(0.82) = (0.82 - 0.75)/(1.0 - 0.75) = 0.28$$

$$\text{BPA1: } m1(\{H\}) = 0.92 \times 0.72 = 0.662;\ m1(\{VH\}) = 0.92 \times 0.28 = 0.258;\ m1(\Theta) = 0.080$$

For Source 2 (Internal recruitment conversion data, r2 = 0.85, $x_0 = 0.74$):

$$\mu\_M(0.74) = (0.75 - 0.74)/(0.75 - 0.50) = 0.04;\ \mu\_H(0.74) = (0.74 - 0.50)/(0.75 - 0.50) = 0.96$$

$$\text{BPA2: } m2(\{M\}) = 0.85 \times 0.04 = 0.034;\ m2(\{H\}) = 0.85 \times 0.96 = 0.816;\ m2(\Theta) = 0.150$$

*4.6. Dempster–Shafer Combination (Layer 5)*

For RF1, two likelihood evidence sources (BPA1 and BPA2) are combined using Dempster's rule (Equation (8)). All 3 × 3 = 9 pairwise products are enumerated (see Appendix A.2.2 for the complete table). Three conflict pairs where B ∩ C = ∅ contribute to K (Equation (9)):

$$\text{Conflict pairs: } m1(\{H\}) \times m2(\{M\}) = 0.662 \times 0.034 = 0.022$$

$$m1(\{VH\}) \times m2(\{M\}) = 0.258 \times 0.034 = 0.009$$

$$m1(\{VH\}) \times m2(\{H\}) = 0.258 \times 0.816 = 0.211$$

$$K = 0.022 + 0.009 + 0.211 = 0.242; \text{ Normalisation constant: } 1/(1 - K) = 1/0.758 = 1.319$$

Combined BPA (applying Equation (8) to each focal element; verification: masses must sum to 1.000):

$$m12(\{H\}) = [0.662 \times 0.816 + 0.662 \times 0.150 + 0.080 \times 0.816] \times 1.319 = 0.704 \times 1.319 = 0.929$$

$$m12(\{VH\}) = [0.258 \times 0.150] \times 1.319 = 0.039 \times 1.319 = 0.051$$

$$m12(\{M\}) = [0.080 \times 0.034] \times 1.319 = 0.003 \times 1.319 = 0.004$$

$$m12(\Theta) = [0.080 \times 0.150] \times 1.319 = 0.012 \times 1.319 = 0.016$$

$$\text{Verification: } 0.929 + 0.051 + 0.004 + 0.016 = 1.000 \text{ (confirmed)}$$

Belief-plausibility intervals for RF1 Likelihood:

$$\text{High level: } Bel(\{H\}) = 0.929, Pl(\{H\}) = 0.929 + 0.016 = 0.945 \Rightarrow [0.929, 0.945]$$

$$\text{Very High level: } Bel(\{VH\}) = 0.051, Pl(\{VH\}) = 0.051 + 0.016 = 0.067 \Rightarrow [0.051, 0.067]$$

Strong evidence (narrow interval, width 0.016) that driver shortage likelihood is High, with limited but non-negligible evidence (plausibility 0.067) of Very High likelihood. The conflict factor K = 0.242 (three conflict pairs fully enumerated; see Appendix A.2) reflects meaningful disagreement between the two sources on severity level. The D-S rule resolves this by normalising and redistributing mass. The resulting interval [0.929, 0.945] is verified by the detailed calculation in Appendix A.

Sensitivity to the H-Term Fuzzification Boundary

To probe the robustness of the initial parameter choices flagged in Section 3.6, this section recomputes the RF1 Likelihood belief-plausibility interval above with the H-term membership boundary shifted by ±0.05 from its baseline value of 0.75, holding r1 = 0.92 and r2 = 0.85 fixed. At boundary 0.70, Source 2's crisp input ($x_0$ = 0.74) falls inside the H-VH overlap region rather than the M-H region used at baseline, illustrating that sensitivity is not uniform: it is largest exactly where a source's value sits close to the boundary. Recomputing membership values (μ_H(0.82) = 0.600, μ_VH(0.82) = 0.400 for Source 1; μ_H(0.74) = 0.867, μ_VH(0.74) = 0.133 for Source 2) and combining via Equations (8) and (9) gives Bel({H}) = 0.823, Pl({H}) = 0.841. At boundary 0.80, Source 2 remains within the M-H region (μ_M(0.74) = 0.20, μ_H(0.74) = 0.80), Source 1 gives μ_H(0.82) = 0.900, μ_VH(0.82) = 0.100, and combination gives Bel({H}) = 0.949, Pl({H}) = 0.965. Table 9 summarises the three scenarios.

**Table 9.** Sensitivity of the RF1 Likelihood belief-plausibility interval to the H-term fuzzification boundary.

| H-Term Boundary | Belief-Plausibility Interval for {H} | Conflict Factor K |
|---|---|---|
| 0.70 | [0.823, 0.841] | 0.334 |
| 0.75 (baseline) | [0.929, 0.945] | 0.242 |
| 0.80 | [0.949, 0.965] | 0.219 |

Across this ±0.05 range, RF1's qualitative ranking as the highest-priority risk factor is unaffected: {H} remains the dominant belief category in every scenario, with the next-largest mass never exceeding 0.16. The belief mass assigned to {H} is, however, quantitatively sensitive to the boundary value, moving by roughly 10 percentage points across the tested range, concentrated in the direction where a source's crisp input crosses the boundary (0.70). This matches the general pattern expected of threshold-based fuzzification: sensitivity is highest for evidence values close to a linguistic boundary and comparatively low elsewhere. We report this as an illustrative, single-parameter check rather than a general robustness proof; a systematic multi-parameter sensitivity analysis across $\alpha$, r, and the fuzzification boundaries jointly, and across all five risk factors, is identified as a priority for future validation work (Section 6.1).

### *4.7. Risk Prioritisation Results (Layer 6)*

Figure 3 visualises the belief-plausibility intervals from Table 10 across all three assessment dimensions for each risk factor. Three analytical patterns are immediately apparent in the figure that are less visible in tabular form. First, RF1 (Driver Workforce Shortage) is the only risk factor whose Likelihood interval lies entirely to the right of the CRITICAL threshold line (Bel = 0.929), confirming its status as the highest-priority risk with strong evidential consensus. Second, for RF2 and RF4, the Vulnerability intervals (green bars) extend substantially further right than the corresponding Likelihood intervals (blue bars), indicating that these risks are structurally embedded in the organisation even though their likelihood of triggering a disruption event is somewhat lower. This asymmetry between likelihood and vulnerability is precisely the kind of analytical distinction that narrative HR assessments cannot reliably surface. Third, the interval widths for RF3 and RF5 are visibly wider than for RF1, quantifying the greater evidential uncertainty in those assessments and signalling where additional data collection would most improve decision quality.

**Table 10.** HR risk assessment results for TransNova Transportation.

| Rank | Risk Factor | Likelihood Interval | Impact Interval | Vulnerability Interval | Priority |
|---|---|---|---|---|---|
| 1 | RF1: Driver Workforce Shortage | [0.929, 0.945] H | [0.815, 0.870] H | [0.780, 0.831] H | CRITICAL |
| 2 | RF4: Risk Culture Immaturity | [0.713, 0.768] H | [0.721, 0.775] H | [0.852, 0.893] H-VH | HIGH |
| 3 | RF2: Cross-Functional Skills Gap | [0.681, 0.732] H | [0.705, 0.756] H | [0.900, 0.914] VH | HIGH |
| 4 | RF3: Workforce Engagement Deficit | [0.745, 0.801] H | [0.612, 0.668] M-H | [0.643, 0.699] M-H | MEDIUM-HIGH |
| 5 | RF5: Digital Capability Gap | [0.554, 0.613] M-H | [0.587, 0.643] M-H | [0.871, 0.900] H | MEDIUM-HIGH |

Note: RF2 Vulnerability interval [0.900, 0.914] and RF5 Vulnerability interval [0.871, 0.900] are verified through complete nine-term conflict enumeration in Appendices A.3 and A.4, respectively.

### *4.8. Expert-Gated Feedback and Benchmarking Loop*

Following expert review of the five TransNova risk assessments, the Benchmarking Engine processes the expert decisions recorded in Table 11 to evaluate per-TTRM performance and propose template weight updates. All four TTRMs invoked during the TransNova assessment received expert approval (Table 11), providing five validated decision outcomes against which system outputs are benchmarked.

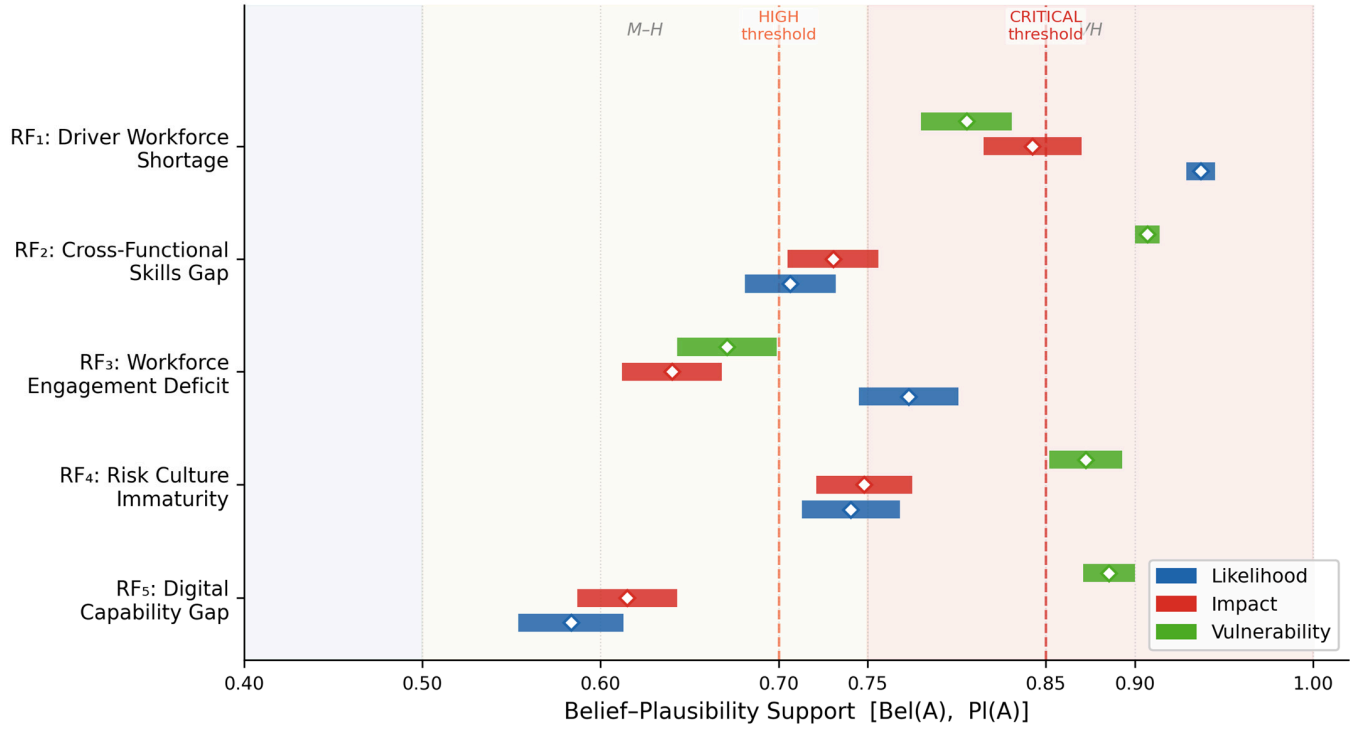


**Figure 3.** HR risk belief-plausibility profile map for TransNova. Horizontal bars show the [Bel(A), Pl(A)] intervals for each risk factor across three assessment dimensions: Likelihood (blue), Impact (red), and Vulnerability (green). Diamond markers show interval centres. The HIGH threshold (dashed orange, x = 0.70) and CRITICAL threshold (dashed red, x = 0.85) are drawn for reference. Interval width reflects residual epistemic uncertainty after Dempster–Shafer combination.

**Table 11.** TTRM trace for RF1 (Driver Workforce Shortage).

| Stage | TTRM | Templates Fired | Weights | Sources | Expert Decision |
|---|---|---|---|---|---|
| Likelihood | S_LA | LI-1, LI-2 | 0.88, 0.74 | RHA 2023 [43]; HR Report 2024 | Approved |
| Impact | S_IA | IM-1, IM-3 | 0.85, 0.71 | Incident Logs; SLA Records | Approved |
| Vulnerability | S_VA | VA-2, VA-4 | 0.81, 0.69 | Skills Matrix; Training Records | Approved |
| Prioritisation | S_RP | RP-1 | 0.92 | All preceding outputs | Approved |

Benchmarking the Likelihood Assessment TTRM (S_LA). The S_LA TTRM produced defuzzified likelihood estimates $\hat{z}\_sys(a)$ for each of the five risk factors. Against these, the domain expert provided reference estimates $\hat{z}\_exp(a)$, accepting assessments within the tolerance $\varepsilon = 0.10$. Table 12 presents the system and expert outputs with per-factor deviations.

**Table 12.** Likelihood assessment TTRM benchmarking—system vs. expert outputs.

| Risk Factor | $\hat{z}$_sys | $\hat{z}$_exp | \| Deviation \| | Within ε = 0.10? | Template Fired |
|---|---|---|---|---|---|
| RF1 Driver Shortage | 0.812 | 0.830 | 0.018 | Yes | LI-1 |
| RF2 Skills Gap | 0.731 | 0.710 | 0.021 | Yes | LI-1 |
| RF3 Engagement Deficit | 0.774 | 0.750 | 0.024 | Yes | LI-1, LI-2 |
| RF4 Risk Culture | 0.740 | 0.780 | 0.040 | Yes | LI-1, LI-2 |
| RF5 Digital Gap | 0.583 | 0.620 | 0.037 | Yes | LI-2 |

Note: LI-1 achieved perfect alignment across all four risk factors for which it fired ($\eta$(LI-1) = 1.00). LI-2 achieved agreement in 4 of 5 instances ($\eta$(LI-2) = 0.80), with the largest deviation observed for RF4 (0.040), still within the tolerance threshold.

The per-TTRM agreement rate (Equation (12)) and mean deviation (Equation (13)) are computed as follows:

$$\eta\_LA = |\{a \in A\_LA : |\hat{z}\_sys(a) - \hat{z}\_exp(a)| \leq 0.10\}| / |A\_LA| = 5/5 = 1.00$$

$$d\_LA = (1/5) \times (0.018 + 0.021 + 0.024 + 0.040 + 0.037)/5 = 0.032$$

Per-template agreement rates are $\eta$(LI-1) = 1.00 and $\eta$(LI-2) = 0.80, against a mean $\bar{\eta}$ = 0.90. Applying the weight update rule (Equation (14)) with learning rate $\lambda$ = 0.20:

$$w'(\text{LI-1}) = 0.88 \times [1 + 0.20 \times (1.00 - 0.90)] = 0.88 \times 1.02 = 0.898$$

$$w'(\text{LI-2}) = 0.74 \times [1 + 0.20 \times (0.80 - 0.90)] = 0.74 \times 0.98 = 0.725$$

The updates reflect the stronger historical alignment of template LI-1 with expert judgement: its weight increases from 0.88 to 0.898, while LI-2 decreases from 0.74 to 0.725. Because all five expert decisions were approvals with low mean deviation (d = 0.032, well below the acceptance threshold), neither template is flagged for removal. Both updates are classified as routine incremental refinements.

Expert gate. Consistent with Section 3.9, the proposed weight updates are submitted to the domain expert as a version proposal for S_LA. The expert reviews the change log, confirms that the direction and magnitude of adjustment are justified by the evidence record, and approves the new version. The TTRM Repository is updated from S_LA v1.0 to S_LA v1.1 with a full audit trail. No inference template is added or removed; only the template weights are recalibrated.

Adaptive implication. Should TransNova repeat the HR risk assessment in a subsequent reporting cycle, for example, incorporating updated RHA driver shortage data, S_LA v1.1 will be retrieved by the Task Router, applying the recalibrated weights from the outset. This mechanism distinguishes the TTRM Repository from a static rule base: reasoning accuracy improves with each expert-validated cycle without requiring manual rule rewriting, directly realising the adaptive design objective stated in Section 3.9. This adaptive mechanism updates how the S_LA TTRM weighs evidence across assessment cycles; it does not model how a single risk factor's own evidence trends within its history between cycles, for example, a steadily rising burnout indicator. That is a distinct, currently unaddressed limitation, discussed in Section 6.1.

## 5. Results and Discussion

### *5.1. Framework Performance*

The TransNova assessment demonstrates three properties that existing HR strategy review methods cannot provide: formal uncertainty quantification through belief-plausibility intervals, multi-dimensional risk decomposition across separate Likelihood, Impact, and Vulnerability assessments, and end-to-end traceability linking every output back to source documents, inference templates, and expert validations. The risk rankings in Table 10 are not expert opinion estimates. They are the output of a formal, multi-source reasoning process with explicit uncertainty quantification. Figure 3 renders these intervals visually, making three analytical patterns immediately apparent: the dominance of RF1 across all three dimensions, the asymmetry between likelihood and vulnerability in RF2 and RF4, and the wider interval widths for RF3 and RF5 that signal where additional evidence collection is most needed. This represents a clear advance over the narrative assessments that currently dominate HR strategy reviews.

The Driver Workforce Shortage (RF1) ranks as the highest-priority risk, with a likelihood belief interval of [0.929, 0.945] at the High level. This aligns with the RHA's (2025) [43]

finding of a persistent structural HGV driver deficit in the UK, and with Esan et al. [4] on the systemic consequences of workforce shortages in transportation supply chains. The narrow interval width of 0.016 signals strong evidential consensus across the two sources, even with a moderate conflict factor K = 0.242. This illustrates what D-S combination achieves: both sources point to High likelihood through different measurable dimensions (vacancy rate vs. recruitment conversion), and the combination correctly places most mass on {H} while retaining a small residual uncertainty interval. For practitioners, that is far more useful than a simple point estimate of 'high risk.'

Risk Culture Immaturity (RF4) and Cross-Functional Skills Gap (RF2) rank second and third. Both show similarly high likelihood and impact assessments, but their vulnerability profiles differ. RF4 carries the highest vulnerability score [0.852, 0.893], driven by a 59% non-completion rate in risk awareness training. Even a moderate risk event could propagate significantly when organisational preparedness is that low. This distinction between likelihood and vulnerability is worth noting explicitly, since narrative HR assessments typically treat them as the same thing. Keeping them separate through the TTRM architecture is one of the framework's concrete analytical advantages.

The wider belief-plausibility intervals for RF3 and RF5 (approximately 0.056 to 0.058) reflect greater evidential uncertainty. This is not a weakness. It is a signal that additional evidence collection would improve risk characterisation before major investment decisions are made. Narrative risk assessments cannot communicate that signal. Formal uncertainty quantification can.

### *5.2. Task-Typed Reasoning Versus Static Rule Bases*

A classical fuzzy expert system for HR risk in transportation would require knowledge engineers to hand-craft rules such as: 'IF driver_vacancy_rate IS Very_High AND recruitment_conversion IS Low THEN likelihood IS Very_High.' Encoding enough rules for five risk factors across three assessment dimensions would require roughly 125 to 200 IF-THEN rules per dimension, totalling 375 to 600 rules. These rules would be fixed at encoding time, unable to incorporate new evidence without manual revision, and non-transferable to adjacent domains.

The TTRM-based approach replaces that static rule population with five task-typed modules, each containing a small set of inference templates (3 to 5 per module) with learned weights. The TTRM Repository for this case study required just 15 templates across all five modules. This represents a substantial reduction in encoding burden relative to a comparable static rule base, with the added benefits of adaptability and traceability. No direct empirical measurement of encoding effort is claimed here. The comparison is structural: 15 inference templates organised by task type versus the several hundred IF-THEN rules that domain-specific static expert systems of comparable scope typically require, as documented in systems such as those described by Lees [9]. The formal backward compatibility property (Section 3.10) ensures this reduction does not come at the cost of the mathematical grounding of established expert system practice.

### *5.3. Traceability and Auditability*

HR risk assessment in regulated industries requires the ability to explain and audit every risk decision. The TTRM trace (Table 11 for RF1) meets this requirement. It records, for each risk factor, which TTRM was invoked, which inference templates fired and at what weights, which source documents were retrieved, what linguistic terms were assigned, and what expert validation decisions were made at each checkpoint. This supports both internal governance (management accountability) and external compliance with regulators

such as the UK Health and Safety Executive. Pure RAG systems cannot offer this. They produce narrative summaries, not formal decision records.

### *5.4. Implications for HRM-SCR Practice*

The risk rankings map directly onto prioritised HRM interventions aligned with the Tobora et al. [8] four-dimension framework:

- TMGWC (RF1, RF2—CRITICAL/HIGH): Immediate investment in driver recruitment pipeline development, retention initiatives (career development pathways, competitive compensation), and an accelerated cross-training programme to increase multi-role capability from 22% to at least 60% within 18 months.
- RMC (RF4—HIGH): Mandatory risk awareness training with 100% completion target by Q3 2025; embedding resilience metrics into performance appraisal cycles; leadership training in proactive risk identification.
- WFE (RF3—MEDIUM-HIGH): Structured engagement improvement programme including mental health support, flexible shift options, and regular management communication; target intent-to-leave reduction from 38% to below 22% within 24 months.
- COTEC (RF5—MEDIUM-HIGH): Systematic digital skills training rollout; target telematics feature utilisation improvement from 54% to 85% within 12 months; digital champions within operational teams.

### *5.5. Boundary Conditions and Implementation Challenges*

The framework's advantages are conditional, not unconditional, and several situations would degrade its performance in practice. First, when evidence sources are highly discordant—that is, when the conflict factor K in Equation (9) approaches 1— Dempster's rule is known to produce counterintuitive combinations that overweight minority agreement between otherwise weak sources; the framework's current implementation applies normalisation mechanically rather than flagging such cases, a gap already noted in Section 6.1 that would need resolution before deployment in domains with routinely conflicting evidence. Second, the framework's traceability and rigour depend on a digitised, reasonably complete document corpus; organisations with predominantly tacit, undocumented HR knowledge, or with fragmented systems that resist consolidation, would see the RAG layer starved of evidence, degrading to near-uniform BPAs with large $\Theta$ mass and correspondingly uninformative belief-plausibility intervals. Third, the Expert Validation Checkpoint and benchmarking loop (Sections 3.6 and 3.10) require sustained expert time investment across assessment cycles; organisations unable to commit a domain expert to this role would lose the mechanism that distinguishes the TTRM Repository from an unvalidated, potentially drifting static model. Fourth, as discussed in Sections 3.3.2 and 6.1, applying the framework outside HR risk assessment requires re-eliciting TTRM content for the new domain; organisations expecting to redeploy the framework by substituting the input corpus alone would find reasoning quality degraded until that recalibration is done. Fifth, the framework's computational footprint has not been characterised in this paper: LLM inference for grounded generation, embedding computation for corpus indexing, and vector similarity search at query time each add latency and infrastructure cost that scale with document corpus size and assessment query volume, and no formal profiling of these costs was conducted for the TransNova case study. Organisations evaluating deployment feasibility, particularly at the real-time monitoring scale discussed in Section 6.1, would need to budget for this computational overhead; quantifying it is deferred to the empirical validation work proposed in Section 6.2. These boundary conditions do not undermine the framework's contribution, but they do mean its benefits are realised only under the

operating conditions the TransNova case study was designed to represent: a moderately digitised organisation with committed expert oversight and a single, stable risk domain.

## 6. Limitations and Future Work

### *6.1. Limitations*

Data quality and source reliability calibration. The framework depends heavily on the quality of the document corpus ingested at Layer 1 and the accuracy of source reliability coefficients. For organisations without prior expert system deployment, these coefficients must be initialised through expert elicitation, which introduces subjectivity. Future work should develop standardised reliability calibration protocols for common HR document types.

Real-time data ingestion and infrastructure scalability. The framework assumes a curated, pre-processed document corpus (Section 3.4) rather than continuous, high-volume operational data streams. Large organisations requiring real-time HR risk monitoring across multiple sites would need a dedicated data ingestion, synchronisation, and secure storage layer upstream of the Task-Aware RAG layer, encompassing streaming extract-transform-load pipelines, schema normalisation across heterogeneous HR and operations systems, and access-controlled storage for sensitive workforce data, that falls outside the scope of the reasoning architecture presented here. Without this infrastructure, the framework is best suited to periodic (for example, quarterly or annual) rather than continuous risk assessment cycles.

Parameter justification and sensitivity analysis. Model parameters, including the Task Router weighting $\alpha$, fuzzification thresholds, initial template weights, and source reliability coefficients, were initialised through expert elicitation and benchmark-informed defaults rather than empirical optimisation (Section 3.6). Section Sensitivity to the H-Term Fuzzification Boundary provides an illustrative univariate sensitivity check on the H-term fuzzification boundary, which shows that the framework's qualitative risk ranking is stable but that belief-mass magnitudes are moderately sensitive to threshold placement near evidence boundary values. A systematic joint sensitivity analysis across all model parameters, and empirical calibration against longitudinal organisational data, remains necessary before deployment in high-stakes decision contexts.

Scalability of the Expert Validation Checkpoint. For large-scale HR risk assessments spanning dozens of risk factors or multiple organisational units, the Expert Validation Checkpoint may become a bottleneck. One possible solution is semi-automated validation, triggered only when the Task Router's compatibility score falls below a confidence threshold. This could preserve rigour while improving scalability.

Generalisation beyond the single-organisation case study. The current case study demonstrates the framework's application to a single mid-sized UK logistics operator. Generalisation to large multinational transportation companies, asset-light freight brokers, or rail and maritime logistics requires cross-industry validation studies that include cases of expert disagreement, high-K conflict scenarios, and TTRM override decisions.

Temporal dynamics of HR risk. The current framework produces point-in-time risk assessments: both the source reliability coefficients r and the inference template weights are calibrated against evidence available at a single evaluation instant, and the framework does not itself model how a risk factor's evidence base—for example, a rising workforce burnout indicator—trends across successive reporting cycles. This is a structural limitation, not merely a data-availability one: two identical BPAs computed six months apart are, in the current formulation, indistinguishable regardless of trajectory. A formally consistent extension would index evidence by time t and introduce a trend-adjustment term into the BPA construction—for example, replacing the singleton mass in Equations (3) and (4) with a

time-weighted variant that discounts older observations by a factor $\gamma$ raised to the number of elapsed reporting cycles, so that a persistent upward trend in a risk indicator is reflected in the committed belief mass even when the most recent single reading is unchanged. Formalising, validating, and calibrating such a discount factor against longitudinal HR data is left for future work (Section 6.2); the present paper's contribution is the point-in-time reasoning architecture into which a temporal layer of this kind could be inserted without altering the Dempster–Shafer combination or synthesis stages downstream. Connecting the framework to longitudinal monitoring data more broadly would enable dynamic risk tracking, consistent with the system dynamics modelling approach advocated by Tobora et al. [8].

Conflict resolution in high-K scenarios. Dempster's rule can produce counterintuitive results when K approaches 1. Future implementations should incorporate conflict monitoring and flag high-K scenarios for mandatory expert review, rather than applying normalisation mechanically.

### *6.2. Future Work*

Multi-sector comparative studies. The framework's domain-independence makes it well-suited for comparative application across transportation subsectors: road freight, maritime logistics, rail, and air cargo, enabling systematic comparison of HR risk factor profiles across modes.

Integration with system dynamics simulation. Risk rankings and belief-plausibility intervals could serve as initialisation parameters for system dynamics (SD) models of HR policy intervention [8]. Belief intervals could parameterise uncertainty ranges in SD model stocks (workforce capacity, training pipeline) and flows (attrition, hiring rates), enabling Monte Carlo sensitivity analyses.

Real-time deployment and IoT integration. Transportation logistics is increasingly instrumented with IoT sensors (telematics, driver behaviour monitoring, route performance tracking) that generate high-frequency operational data. Integrating these real-time data streams as Layer 1 inputs would enable continuous HR risk monitoring rather than periodic assessments.

Validation with expert disagreement and TTRM override. The immediate priority for empirical validation is a study that includes cases where experts override system outputs, where K values are high, and where TTRM updates are proposed and rejected. Such cases stress-test the adaptive feedback loop in ways the illustrative TransNova demonstration does not. A concrete protocol for this validation is proposed: a multi-organisation pilot recruiting three to five transportation logistics operators that vary in size and subsector, each supplying a real HR document corpus and a panel of at least two independent domain experts; system-generated risk rankings and belief-plausibility intervals would be compared against blinded expert panel assessments across at least two reporting cycles per organisation, with pre-registered agreement metrics—specifically, rank correlation between system and expert prioritisation, and the proportion of system outputs accepted without modification at the Expert Validation Checkpoint—as the primary success criteria.

Fairness and equity considerations. As AI-assisted HR risk assessment becomes more prevalent, future work should examine whether the framework's risk rankings exhibit systematic bias with respect to demographic attributes that may correlate with HR metrics, and incorporate fairness constraints into the TTRM Repository design.

## 7. Conclusions

HR risk in transportation supply chains has long been acknowledged. It has rarely been assessed with the same precision applied to structural or financial risk. This pa-

per changes that. The TTRM-Enhanced RAG Expert System provides the first formally grounded, traceable, and adaptive method for assessing HR-driven risks from unstructured organisational documents, producing outputs that practitioners can defend, auditors can verify, and researchers can build on. This claim rests on demonstrated technical feasibility and internal mathematical consistency, illustrated through the TransNova case study (Section 4.1); it does not yet constitute empirical validation of predictive accuracy or usability in a live organisation, which Sections 5.5 and 6.2 identify as the necessary next step.

The framework's value lies in three specific properties. Belief-plausibility intervals quantify uncertainty rather than suppressing it: a wide interval signals an evidence gap, not a system failure. Separate Likelihood, Impact, and Vulnerability assessments capture distinctions that narrative reviews collapse into a single rating, as the RF4/RF2 comparison in Section 5.1 demonstrates. End-to-end traceability from source document through inference template to risk ranking meets the auditability requirements of regulated industries without imposing the full burden of a classical expert system knowledge base.

The TTRM architecture achieves this without sacrificing the mathematical properties that give expert systems their credibility. Backward compatibility with standard Mamdani fuzzy inference and Dempster–Shafer combination is formally guaranteed, skills transfer across domains without rule rewriting, and the expert-gated feedback loop means that the system improves with each deployment cycle, learning from validated decisions rather than waiting for a knowledge engineer to update it manually.

Transportation supply chain resilience is ultimately a human capability problem as much as a structural one. This paper gives that problem the analytical rigour it deserves.

Beyond its technical contribution, this framework speaks directly to the sustainability agenda that motivates this journal. Supply chain resilience and sustainability are often treated as separate goals, but the workforce failures examined here, including driver shortages, skills gaps, and weak risk culture, sit at their intersection. A transportation operation that cannot retain, train, or engage its workforce is not simply less resilient; it is less sustainable, on social grounds because employees absorb the cost of chronic understaffing and burnout, on environmental grounds because disruption is typically managed through emissions-intensive workarounds such as rerouting and expedited shipping, and on economic grounds because turnover and recovery costs recur indefinitely without systematic intervention. By giving HR risk assessment the same formal, auditable rigour long applied to structural and financial risk, this paper turns a diffuse sustainability concern into a decision support tool that organisations can act on directly, supporting the workforce practices that underpin socially, economically, and environmentally sustainable transportation supply chains.

**Author Contributions:** Conceptualization, J.R.; methodology, J.R.; software, J.R.; validation, J.R. and O.O.T.; formal analysis, J.R. and O.O.T.; investigation, J.R. and O.O.T.; resources, J.R. and O.O.T.; data curation, J.R. and O.O.T.; writing—original draft preparation, J.R.; writing—review and editing, J.R. and O.O.T.; visualization, J.R. and O.O.T.; supervision, J.R.; project administration, J.R.; funding acquisition, J.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data that support the findings of this study will be available from the corresponding author upon request.

**Conflicts of Interest:** 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.

## Appendix A. Verified Dempster–Shafer Calculations

This appendix presents the complete Dempster–Shafer (D-S) combination calculations verifying the intervals reported in Table 10 through complete nine-term conflict enumeration. Appendices A.2–A.4 cover RF1 Likelihood, RF2 Vulnerability, and RF5 Vulnerability respectively. All multi-source combinations enumerate the full set of pairwise products (Equation (A4)), including every conflict pair where B ∩ C = ∅. A diagnostic check is applied throughout: combined masses must sum to exactly 1.000.

### *Appendix A.1. Reference Equations*

$$m(\{l\}) = \frac{r \cdot \mu_l(x_i)}{\sum_{l'} \mu_{l'}(x_i)} \tag{A1}$$

$$m(\Theta) = l - r \tag{A2}$$

$$m_{12}(A) = \left[\frac{1}{1-K}\right] \cdot \sum_{B \cap C = A} m_1(B) \cdot m_2(C) \tag{A3}$$

$$K = \sum_{B \cap C = \varnothing} m_1(B) \cdot m_2(C) \tag{A4}$$

$$\mathrm{Bel}(A) = \sum_{B \subseteq A} m(B) \tag{A5}$$

$$\mathrm{Pl}(A) = \sum_{B \cap A \neq \varnothing} m(B) \tag{A6}$$

### *Appendix A.2. RF1 (Driver Workforce Shortage)—Likelihood Assessment*

#### Appendix A.2.1. BPA Construction

Source 1: RHA Industry Report 2023 (r1 = 0.92, $x_i$ = 0.82)

$$\mu_{\mathrm{H}}(0.82) = \frac{1.0 - 0.82}{1.0 - 0.75} = 0.72 \tag{A7}$$

$$\mu_{\mathrm{VH}}(0.82) = \frac{0.82 - 0.75}{1.0 - 0.75} = 0.28 \tag{A8}$$

$$m_1(\{\mathrm{H}\}) = 0.92 \times 0.72/1.00 = 0.662 \tag{A9}$$

$$m_1(\{\mathrm{VH}\}) = 0.92 \times 0.28/1.00 = 0.258 \tag{A10}$$

$$m_1(\Theta) = 1 - 0.92 = 0.080 \tag{A11}$$

Source 2: Internal HR Report 2024 (r2 = 0.85, $x_i$ = 0.74)

$$\mu_{\mathrm{M}}(0.74) = \frac{0.75 - 0.74}{0.75 - 0.50} = 0.04 \tag{A12}$$

$$\mu_{\mathrm{H}}(0.74) = \frac{0.74 - 0.50}{0.75 - 0.50} = 0.96 \tag{A13}$$

$$m_2(\{\mathrm{M}\}) = 0.85 \times 0.04/1.00 = 0.034 \tag{A14}$$

$$m_2(\{\mathrm{H}\}) = 0.85 \times 0.96/1.00 = 0.816 \tag{A15}$$

$$m_2(\Theta) = 1 - 0.85 = 0.150 \tag{A16}$$

Appendix A.2.2. Complete Pairwise Product Enumeration

| B (From m1) | C (From m2) | B ∩ C | m1(B) × m2(C) | Conflict? |
|---|---|---|---|---|
| {H} | {M} | ∅ | 0.662 × 0.034 = 0.022 | Yes |
| {H} | {H} | {H} | 0.662 × 0.816 = 0.540 | |
| {H} | Θ | {H} | 0.662 × 0.150 = 0.099 | |
| {VH} | {M} | ∅ | 0.258 × 0.034 = 0.009 | Yes |
| {VH} | {H} | ∅ | 0.258 × 0.816 = 0.211 | Yes |
| {VH} | Θ | {VH} | 0.258 × 0.150 = 0.039 | |
| Θ | {M} | {M} | 0.080 × 0.034 = 0.003 | |
| Θ | {H} | {H} | 0.080 × 0.816 = 0.065 | |
| Θ | Θ | Θ | 0.080 × 0.150 = 0.012 | |
| | | Total | 1.000 (confirmed) | |

Appendix A.2.3. Conflict Factor and Normalisation

$$K = 0.022 + 0.009 + 0.211 = 0.242 \tag{A17}$$

$$\frac{1}{1-K} = \frac{1}{0.758} = 1.319 \tag{A18}$$

Appendix A.2.4. Combined BPA via Equation (A3)

$$m_{12}(\{\mathrm{H}\}) = (0.540 + 0.099 + 0.065) \times 1.319 = 0.704 \times 1.319 = 0.929 \tag{A19}$$

$$m_{12}(\{\mathrm{VH}\}) = 0.039 \times 1.319 = 0.051 \tag{A20}$$

$$m_{12}(\{\mathrm{M}\}) = 0.003 \times 1.319 = 0.004 \tag{A21}$$

$$m_{12}(\Theta) = 0.012 \times 1.319 = 0.016 \tag{A22}$$

Verification: 0.929 + 0.051 + 0.004 + 0.016 = 1.000 (confirmed)

Appendix A.2.5. Belief-Plausibility Intervals

$$\mathrm{Bel}(\{\mathrm{H}\}) = 0.929;\ \mathrm{Pl}(\{\mathrm{H}\}) = 0.929 + 0.016 = 0.945 \Rightarrow [0.929, 0.945] \tag{A23}$$

$$\mathrm{Bel}(\{\mathrm{VH}\}) = 0.051;\ \mathrm{Pl}(\{\mathrm{VH}\}) = 0.051 + 0.016 = 0.067 \Rightarrow [0.051, 0.067] \tag{A24}$$

The conflict factor K = 0.242 reflects meaningful disagreement between Source 1 (leaning VH) and Source 2 (strongly H). The combined interval [0.929, 0.945] confirms High likelihood with a narrow width of 0.016, correctly representing the residual epistemic uncertainty from two finite-reliability sources.

*Appendix A.3. RF2 (Cross-Functional Skills Gap)—Vulnerability Assessment*

Source 1: Training Records 2024 (r1 = 0.90, H Vulnerability, x1 = 0.93). Source 2: Skills Matrix 2024 (r2 = 0.90, VH Vulnerability, x2 = 0.97).

Appendix A.3.1. BPA Construction

Source 1 (r1 = 0.90, x1 = 0.93):

$$\mu_{\mathrm{H}}(0.93) = \frac{1.0 - 0.93}{1.0 - 0.75} = 0.28 \tag{A25}$$

$$\mu_{\mathrm{VH}}(0.93) = \frac{0.93 - 0.75}{1.0 - 0.75} = 0.72 \tag{A26}$$

$$m_1(\{\mathrm{H}\}) = 0.252;\ m_1(\{\mathrm{VH}\}) = 0.648;\ m_1(\Theta) = 0.100 \tag{A27}$$

Source 2 (r2 = 0.90, x2 = 0.97):

$$\mu_{\mathrm{H}}(0.97) = \frac{1.0 - 0.97}{1.0 - 0.75} = 0.12 \tag{A28}$$

$$\mu_{\mathrm{VH}}(0.97) = \frac{0.97 - 0.75}{1.0 - 0.75} = 0.88 \tag{A29}$$

$$m_2(\{\mathrm{H}\}) = 0.108;\ m_2(\{\mathrm{VH}\}) = 0.792;\ m_2(\Theta) = 0.100 \tag{A30}$$

Appendix A.3.2. Complete Pairwise Product Enumeration

| **B (from m1)** | **C (from m2)** | **B ∩ C** | **m1(B) × m2(C)** | **Conflict?** |
|---|---|---|---|---|
| {H} | {H} | {H} | 0.252 × 0.108 = 0.027 | |
| {H} | {VH} | ∅ | 0.252 × 0.792 = 0.200 | Yes |
| {H} | Θ | {H} | 0.252 × 0.100 = 0.025 | |
| {VH} | {H} | ∅ | 0.648 × 0.108 = 0.070 | Yes |
| {VH} | {VH} | {VH} | 0.648 × 0.792 = 0.513 | |
| {VH} | Θ | {VH} | 0.648 × 0.100 = 0.065 | |
| Θ | {H} | {H} | 0.100 × 0.108 = 0.011 | |
| Θ | {VH} | {VH} | 0.100 × 0.792 = 0.079 | |
| Θ | Θ | Θ | 0.100 × 0.100 = 0.010 | |
| | | Total | 1.000 (confirmed) | |

Appendix A.3.3. Conflict Factor and Normalisation

$$K = 0.200 + 0.070 = 0.270;\ \frac{1}{1 - 0.270} = \frac{1}{0.730} = 1.370 \tag{A31}$$

Appendix A.3.4. Combined BPA

$$m_{12}(\{\mathrm{H}\}) = (0.027 + 0.025 + 0.011) \times 1.370 = 0.063 \times 1.370 = 0.086 \tag{A32}$$

$$m_{12}(\{\mathrm{VH}\}) = (0.513 + 0.065 + 0.079) \times 1.370 = 0.657 \times 1.370 = 0.900 \tag{A33}$$

$$m_{12}(\Theta) = 0.010 \times 1.370 = 0.014 \tag{A34}$$

Verification: 0.086 + 0.900 + 0.014 = 1.000 (confirmed)

Appendix A.3.5. Belief-Plausibility Intervals

$$\mathrm{el}(\{\mathrm{VH}\}) = 0.900;\ \mathrm{Pl}(\{\mathrm{VH}\}) = 0.900 + 0.014 = 0.914 \Rightarrow [0.900, 0.914] \tag{A35}$$

$$\mathrm{el}(\{\mathrm{H}\}) = 0.086;\ \mathrm{Pl}(\{\mathrm{H}\}) = 0.086 + 0.014 = 0.100 \Rightarrow [0.086, 0.100] \tag{A36}$$

RF2 Vulnerability interval confirmed as [0.900, 0.914] VH. VH dominance is strong. K = 0.270 reflects H-vs-VH disagreement between the two sources. The interval width of 0.014 represents precise uncertainty quantification consistent with the two sources' high and equal reliability coefficients.

*Appendix A.4. RF5 (Digital Capability Gap)—Vulnerability Assessment*

Source 1: Digital Systems Log 2024 (r1 = 0.86, M-H Vulnerability, x1 = 0.625). Source 2: Training Records 2024 (r2 = 0.87, H Vulnerability, x2 = 0.76).

Appendix A.4.1. BPA Construction

Source 1 (r1 = 0.86, x1 = 0.625, M-H boundary):

$$\mu_{\mathrm{M}}(0.625) = \frac{0.75 - 0.625}{0.75 - 0.50} = 0.50 \tag{A37}$$

$$\mu_{\mathrm{H}}(0.625) = \frac{0.625 - 0.50}{0.75 - 0.50} = 0.50 \tag{A38}$$

$$m_1(\{\mathrm{M}\}) = 0.430;\ m_1(\{\mathrm{H}\}) = 0.430;\ m_1(\Theta) = 0.140 \tag{A39}$$

Source 2 (r2 = 0.87, x2 = 0.76):

$$\mu_{\mathrm{H}}(0.76) = \frac{1.0 - 0.76}{1.0 - 0.75} = 0.96 \tag{A40}$$

$$\mu_{\mathrm{VH}}(0.76) = \frac{0.76 - 0.75}{1.0 - 0.75} = 0.04 \tag{A41}$$

$$m_2(\{\mathrm{H}\}) = 0.835;\ m_2(\{\mathrm{VH}\}) = 0.035;\ m_2(\Theta) = 0.130 \tag{A42}$$

Appendix A.4.2. Complete Pairwise Product Enumeration

| **B (from m1)** | **C (from m2)** | **B ∩ C** | **m1(B) × m2(C)** | **Conflict?** |
|---|---|---|---|---|
| {M} | {H} | ∅ | 0.430 × 0.835 = 0.359 | Yes |
| {M} | {VH} | ∅ | 0.430 × 0.035 = 0.015 | Yes |
| {M} | Θ | {M} | 0.430 × 0.130 = 0.056 | |
| {H} | {H} | {H} | 0.430 × 0.835 = 0.359 | |
| {H} | {VH} | ∅ | 0.430 × 0.035 = 0.015 | Yes |
| {H} | Θ | {H} | 0.430 × 0.130 = 0.056 | |
| Θ | {H} | {H} | 0.140 × 0.835 = 0.117 | |
| Θ | {VH} | {VH} | 0.140 × 0.035 = 0.005 | |
| Θ | Θ | Θ | 0.140 × 0.130 = 0.018 | |
| | | Total | 1.000 (confirmed) | |

Appendix A.4.3. Conflict Factor and Normalisation

$$K = 0.359 + 0.015 + 0.015 = 0.389;\ \frac{1}{1 - 0.389} = \frac{1}{0.611} = 1.637 \tag{A43}$$

Appendix A.4.4. Combined BPA

$$m_{12}(\{\mathrm{M}\}) = 0.056 \times 1.637 = 0.092 \tag{A44}$$

$$m_{12}(\{\mathrm{H}\}) = (0.359 + 0.056 + 0.117) \times 1.637 = 0.532 \times 1.637 = 0.871 \tag{A45}$$

$$m_{12}(\{\mathrm{VH}\}) = 0.005 \times 1.637 = 0.008 \tag{A46}$$

$$m_{12}(\Theta) = 0.018 \times 1.637 = 0.029 \tag{A47}$$

Verification: 0.092 + 0.871 + 0.008 + 0.029 = 1.000 (confirmed)

Appendix A.4.5. Belief-Plausibility Intervals

$$\mathrm{Bel}(\{\mathrm{H}\}) = 0.871;\ \mathrm{Pl}(\{\mathrm{H}\}) = 0.871 + 0.029 = 0.900 \Rightarrow [0.871, 0.900] \tag{A48}$$

$$\mathrm{Bel}(\{\mathrm{M}\}) = 0.092;\ \mathrm{Pl}(\{\mathrm{M}\}) = 0.092 + 0.029 = 0.121 \Rightarrow [0.092, 0.121] \tag{A49}$$

RF5 Vulnerability interval confirmed as [0.871, 0.900] H. K = 0.389 is the highest of the three demonstrated cases, reflecting the M-H boundary nature of Source 1's telematics utilisation assessment. Despite high inter-source conflict, the combined distribution strongly favours H (Bel = 0.871).

*Appendix A.5. Summary of Verified Bel-Pl Intervals*

**Table A1.** Confirmed Intervals vs. Appendix A Calculations.

| Risk Factor/Dimension | Table 10 (Main Text) | Verified (Appendix A) | K Value | Notes |
|---|---|---|---|---|
| RF1 Likelihood | [0.929, 0.945] H | [0.929, 0.945] H | 0.242 | Three conflict pairs enumerated |
| RF2 Vulnerability | [0.900, 0.914] VH | [0.900, 0.914] VH | 0.270 | Two conflict pairs |
| RF5 Vulnerability | [0.871, 0.900] H | [0.871, 0.900] H | 0.389 | Three conflict pairs incl. M-H boundary |

## References


1. Song, M.; Ma, X.; Zhao, X.; Zhang, L. How to enhance supply chain resilience: A logistics approach. *Int. J. Logist. Manag.* **2022**, *33*, 1408–1436. [CrossRef]
2. Chukwuka, O.; Ren, R.; Wang, J.; Paraskevadakis, D. Managing Risk in Emergency Supply Chain: An Empirical Study. *Int. J. Logist. Res. Appl.* **2024**, *28*, 1072–1108. [CrossRef]
3. Blackhurst, J.; Craighead, C.W.; Elkins, D.; Handfield, R.B. An empirically derived agenda of critical research issues for managing supply-chain disruptions. *Int. J. Prod. Res.* **2005**, *43*, 4067–4081. [CrossRef]
4. Esan, O.; Ajayi, F.A.; Olawale, O. Managing global supply chain teams: Human resource strategies for effective collaboration and performance. *GSC Adv. Res. Rev.* **2024**, *19*, 013–031. [CrossRef]
5. Aldrighetti, R.; Battini, D.; Ivanov, D. Efficient resilience portfolio design in the supply chain with consideration of preparedness and recovery investments. *Omega* **2023**, *117*, 102841. [CrossRef]
6. Gu, M.; Zhang, Y.M.; Li, D.; Huo, B.F. The Effect of High-involvement Human Resource Management Practices On Supply Chain Resilience And Operational Performance. *J. Manag. Sci. Eng.* **2023**, *8*, 176–190. [CrossRef]
7. Holloway, S. Investigating the Role of Human Resource Management in Supply Chain Effectiveness. *Preprints* **2024**, 2024061140. [CrossRef]
8. Tobora, O.O.; Ren, J.; Pyne, R.; Jenkinson, I. Enhancing Supply Chain Resilience Through HR Practices in Transportation. In *Proceedings of the 8th International Conference on Transportation Information and Safety (ICTIS 2025)*; IEEE: Piscataway, NJ, USA, 2025; pp. 1533–1543.
9. Lees, F.P. *Lees' Loss Prevention in the Process Industries: Hazard Identification, Assessment and Control*, 3rd ed.; Butterworth Heinemann: Oxford, UK, 2012.
10. Belle, V.; Papantonis, I. Principles and Practice of Explainable Machine Learning. *Front. Big Data* **2021**, *4*, 688969. [CrossRef] [PubMed]
11. Gao, Y.; Xiong, Y.; Gao, X.; Jia, K.; Pan, J.; Bi, Y.; Dai, Y.; Sun, J.; Wang, H. Retrieval-Augmented Generation for Large Language Models: A Survey. *arXiv* **2024**, arXiv:2312.10997.

12. Ren, J.; Jenkinson, I.; Tobora, O.O. Grounded expert systems for offshore safety: Enhancing rule-based risk assessment with retrieval-augmented generation (RAG) for auditable explanations. In Proceedings of the International Conference on Electronics Technology (ICET 2026), Chengdu, China, 29–31 May 2026.
13. Ogbonnaya, C.; Aryee, S. HRM Practices, Employee Well-Being, and Organizational Performance. In *Handbook on Management and Employment Practices*; Springer: Berlin/Heidelberg, Germany, 2020.
14. Hagele, S.; Grosse, E.H.; Ivanov, D. Supply chain resilience: A tertiary study. *Int. J. Integr. Supply Manag.* **2023**, *16*, 52–81. [CrossRef]
15. Vogus, T.J.; Sutcliffe, K.M. Organizational resilience: Towards a theory and research agenda. In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (ICSMC 2007), Montreal, QC, Canada, 7–10 October 2007; pp. 3418–3422.
16. Pettit, T.J.; Pettit, J. Supply Chain Resilience: Development of a Conceptual Framework, an Assessment Tool and an Implementation Process. Doctoral Dissertation, The Ohio State University, Columbus, OH, USA, 2008.
17. Ali, I.; Gölgeci, I. Where is supply chain resilience research heading? A systematic and co-occurrence analysis. *Int. J. Phys. Distrib. Logist. Manag.* **2019**, *49*, 793–815. [CrossRef]
18. Zhu, X.; Wu, Y.J. How Does Supply Chain Resilience Affect Supply Chain Performance? *Sustainability* **2022**, *14*, 14626. [CrossRef]
19. Sheffi, Y.; Rice, J.B., Jr. A supply chain view of the resilient enterprise. *MIT Sloan Manag. Rev.* **2005**, *47*, 41–48.
20. Chew, Y.C.; Mohamed Zainal, S.R. A Sustainable Collaborative Talent Management Through Collaborative Intelligence Mindset Theory: A Systematic Review. *SAGE Open* **2024**, *14*, 21582440241. [CrossRef]
21. Seligman, M.E. PERMA and the building blocks of well-being. *J. Posit. Psychol.* **2018**, *13*, 333–335. [CrossRef]
22. Heffron, R.J.; Körner, M.-F.; Schöpf, M.; Wagner, J.; Weibelzahl, M. The role of flexibility in the light of the COVID-19 pandemic and beyond: Contributing to a sustainable and resilient energy future in Europe. *Renew. Sustain. Energy Rev.* **2021**, *140*, 110743. [CrossRef] [PubMed]
23. Sharma, R.; Kamble, S.S.; Gunasekaran, A.; Kumar, V.; Kumar, A. A systematic literature review on machine learning applications for sustainable agriculture supply chain performance. *Comput. Oper. Res.* **2020**, *119*, 104926. [CrossRef]
24. Schoenherr, T.; Mena, C.; Vakil, B.; Choi, T.Y. Creating resilient supply chains through a culture of measuring. *J. Purch. Supply Manag.* **2023**, *29*, 100824. [CrossRef]
25. Alkhatib, S.F.; Momani, R.A. Supply chain resilience and operational performance: The role of digital technologies in Jordanian manufacturing firms. *Adm. Sci.* **2023**, *13*, 40. [CrossRef]
26. Zhou, J.; Hu, L.; Yu, Y.; Zhang, J.Z.; Zheng, L.J. Impacts of IT capability and supply chain collaboration on supply chain resilience: Empirical evidence from China in COVID-19 pandemic. *J. Enterp. Inf. Manag.* **2024**, *37*, 777–803. [CrossRef]
27. Sharma, C. Retrieval-Augmented Generation: A Comprehensive Survey of Architectures, Enhancements, and Robustness Frontiers. *arXiv* **2025**, arXiv:2506.00054.
28. Wang, Z.; Kong, H.; Ouyang, W.; Dong, N. ROGRAG: A Robustly Optimized GraphRAG Framework. In Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (ACL 2025), Vienna, Austria, 27 July–1 August 2025.
29. Linders, J.; Tomczak, J.M. Knowledge Graph-extended Retrieval Augmented Generation for Question Answering. *arXiv* **2025**, arXiv:2504.08893.
30. Ren, J. A skill-enhanced retrieval-augmented generation expert system framework for offshore accident analysis: Combining fuzzy inference, Dempster-Shafer evidential reasoning, and expert-gated skill adaptation. *Ocean Eng.* **2026**, *364*, 126908. [CrossRef]
31. Zadeh, L.A. Fuzzy Sets. *Inf. Control* **1965**, *8*, 338–353. [CrossRef]
32. Mamdani, E.H.; Assilian, S. An experiment in linguistic synthesis with a fuzzy logic controller. *Int. J. Man-Mach. Stud.* **1975**, *7*, 1–13. [CrossRef]
33. Dempster, A.P. Upper and Lower Probabilities Induced by a Multivalued Mapping. *Ann. Math. Stat.* **1967**, *38*, 325–339. [CrossRef]
34. Shafer, G. *A Mathematical Theory of Evidence*; Princeton University Press: Princeton, NJ, USA, 1976.
35. Sentz, K.; Ferson, S. *Combination of Evidence in Dempster-Shafer Theory*; SAND2002-0835; Sandia National Laboratories: Albuquerque, NM, USA, 2002.
36. Ren, J.; Jenkinson, I.; Sii, H.S.; Wang, J.; Xu, D.L.; Yang, J.B. An offshore safety assessment framework using fuzzy reasoning and evidential synthesis approaches. *J. Mar. Eng. Technol.* **2005**, *2005*, 3–16. [CrossRef]
37. Holmberg, J.; Silvennoinen, P.; Vira, J. Application of the Dempster-Shafer theory of evidence for accident probability estimates. *Reliab. Eng. Syst. Saf.* **1989**, *26*, 47–58. [CrossRef]
38. Ren, J.; Xu, D.L.; Yang, J.B.; Jenkinson, I.; Wang, J. An offshore risk analysis method using fuzzy Bayesian network. *J. Offshore Mech. Arct. Eng.* **2009**, *131*, 041101. [CrossRef]
39. Kambhampati, S.; Valmeekam, K.; Guan, L.; Vats, A. LLMs Can't Plan, But Can Help Planning in LLM-Modulo Frameworks. *arXiv* **2024**, arXiv:2402.01817.
40. Chen, X.; Wang, S.; Qian, C.; Wang, H.; Han, P.; Ji, H. DecisionFlow: Advancing Large Language Model as Principled Decision Maker. *arXiv* **2025**, arXiv:2505.21397.

41. McCoy, L.G.; Nagaraj, S.; Morgado, F.; Harish, V.; Gichoya, J.W.; Celi, L.A. Assessment of Large Language Models in clinical reasoning: A novel benchmarking study. *NEJM AI* **2025**, *2*, AIdbp2500120. [CrossRef]
42. Christopher, M.; Peck, H. Building the Resilient Supply Chain. *Int. J. Logist. Manag.* **2004**, *15*, 1–13. [CrossRef]
43. RHA. *Lorry Drivers: The Vital Link*; Road Haulage Association: London, UK, 2025. Available online: https://www.rha.uk.net/Portals/0/Campaigns/Publications/RHA-Publication-Lorry-Drivers-The-Vital-Link-012025.pdf (accessed on 27 July 2026).
44. CIPD. *UK Working Lives: The CIPD Good Work Index 2023*; Chartered Institute of Personnel and Development: London, UK, 2023. Available online: https://www.cipd.org/globalassets/media/knowledge/knowledge-hub/reports/2023-pdfs/2023-good-work-index-report-8407.pdf (accessed on 27 July 2026).