

Article

AI Decision Support for Urban Fire Risk Management: A Framework for Validation, Governance, and Bounded Deployment

Eric Scheepbouwer 

Department of Civil and Environmental Engineering, University of Canterbury, Christchurch 8140, New Zealand; eric.scheepbouwer@canterbury.ac.nz

Abstract

AI-based decision support is moving into fire practice and governance, where it is used to prioritise inspections, analyse building and community risk, examine station coverage, support evacuation planning, interpret warnings, and explore fire scenarios. These tools can extend analytical capacity, but they also create a decision role migration problem: an output developed for prediction, prioritisation, warning, simulation, or planning may later be treated as clearance, justification, or authority. Existing fire model evaluation guidance recognises that validation is use-specific; AI systems add a further challenge because outputs can migrate across dashboards, reusable software components, interfaces, and institutional procedures. This article develops a role-sensitive framework for bounded deployment of AI decision support in urban fire risk management. The framework classifies AI outputs by epistemic role, decision proximity, validation basis, temporal coupling, consequence asymmetry, and governance explicitness. Its central synthesis is that evidence sufficient to warn may be insufficient to clear. Probabilistic outputs can support screening, investigation, prioritisation, and scenario analysis; permissive or safety-proximate claims require stronger assurance, uncertainty communication, fallback rules, and explicit authority allocation. The contribution is a governance logic for keeping exploratory, advisory, policy-shaping, and safety-proximate AI roles separate in urban fire management and policy formulation.

Keywords: artificial intelligence; decision support; urban fire risk management; fire safety engineering; safety assurance; municipal fire services; bounded deployment; human–automation interaction



Academic Editors: Dingli Liu,
Zhisheng Xu, Jinghong Wang and
Zongjia Zhang

Received: 3 June 2026

Revised: 24 July 2026

Accepted: 27 July 2026

Published: 4 August 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

AI-based decision support is no longer confined to laboratory fire modelling or post-event analysis. In urban fire management, data-driven systems increasingly present fire professionals with ranked buildings, risk scores, coverage maps, predicted hazards, route estimates, and warnings. These outputs can help services and authorities see patterns that would otherwise remain difficult to detect. Their significance, however, depends on what they are allowed to do once they enter inspection, planning, emergency response, regulatory, or policy workflows [1–8].

The practical question is therefore not whether AI can support urban fire management, but what role AI output is allowed to play once it enters a decision process. The central problem is role migration. A fire risk score can begin as an analytical screen and later

become a reason to defer or justify inspection. A detection output can begin as a prompt for human investigation and later become part of automatic escalation. A rapid surrogate can be developed for scenario exploration and later be read as operational guidance because it produces a result while a decision is unfolding. The model may be unchanged, but the decision burden attached to its output has changed.

Three application types illustrate the practical setting. First, Firebird was developed in collaboration with the Atlanta Fire Rescue Department to combine municipal datasets, estimate commercial property fire risk, and support inspection prioritisation and inspector allocation [1]. Second, FSLens uses fire records, spatial forecasting, and interactive multi-criteria analysis to identify potentially under-resourced areas and compare candidate fire station layouts [2]. Third, data-driven building risk analysis can combine predicted occurrence and potential consequence to support prevention and resource planning decisions [3]. These examples represent different stages of implementation, including documented municipal collaboration, case study evaluation, and research-oriented analytics, and are not all presented as routine operational deployments. They nevertheless expose the same governance risk: an advisory score or analytical comparison may acquire stronger authority when embedded in inspection schedules, resource plans, or management dashboards.

Fire engineering already recognises that model validity is use-specific. ASTM E1355-23 treats fire model evaluation as a process for assessing predictive capability for a specified use, with explicit attention to model scope, scenarios, uncertainty, limitations, and misuse prevention [9]. Fire model verification and validation literature similarly stresses that agreement, numerical robustness, assumptions, and applicability must be judged against the intended modelling context [10,11]. This article accepts that foundation and extends it to AI decision support: the relevant use is not only the modelling task, but the decision role that the output acquires in a workflow.

Fire safety decisions are often asymmetric because false-positive and false-negative outputs lead to different actions and harms. A false warning may cause unnecessary evacuation or resource use, whereas a false non-warning may delay intervention or be mistaken for clearance. Both can cause harm, but they fail in different directions and may therefore require different evidentiary burdens [9,12–17].

The article develops a role-sensitive evaluative framework for bounded deployment of AI decision support in urban fire risk management. The framework integrates six dimensions: epistemic role (the claim the output makes), decision proximity (how directly it can affect action or authority), validation basis (the evidence supporting its intended use), temporal coupling (how much time users have to examine, contest, or override it), consequence asymmetry (whether different error directions produce different harms), and governance explicitness (whether limits, responsibilities, fallback rules, and authority are stated). These dimensions are not presented as a universal taxonomy of AI or as individually novel concepts. They are practical tests for determining whether AI output is functioning as information, prioritisation, warning, clearance, policy justification, or safety-proximate authority. The article first defines the study design and synthesis approach, then presents the framework, applies it illustratively to selected studies, and discusses the implications for validation, assurance, and governance.

2. Materials and Methods

2.1. Study Design, Scope, and Unit of Analysis

This article is a conceptual framework study supported by governance analysis and illustrative mapping. It does not develop or benchmark a new AI model, nor does it estimate the prevalence of AI governance gaps across the fire safety literature. Its purpose is to construct a role-sensitive framework for assessing when AI decision support outputs

in urban fire risk management require additional validation, assurance, uncertainty communication, fallback rules, or authority controls. This article type is appropriate because the relevant problem is not only technical performance, but the way AI outputs acquire decision roles in socio-technical systems [12,18,19].

The scope is urban and built-environment fire decision support. This includes municipal fire risk management, inspection prioritisation, building and community risk analysis, fire station and resource planning, evacuation and route support tools, fire detection and warning, rapid simulation, and policy formulation. Community-scale emergency corridor evaluation also illustrates the wider urban fire governance context because it can identify corridor deficiencies and prioritise municipal interventions, thereby influencing planning, investment, and policy decisions [20]. Wildfire and wildland–urban interface studies are used only where they illuminate spatial risk, peri-urban exposure, rapid prediction, resource allocation, or decision support transfer relevant to urban fire governance [7,21]. Here, peri-urban exposure means the exposure of people, buildings, and infrastructure on the urban fringe, where developed areas meet rural or vegetated land. These studies are not treated as a separate wildfire AI review stream.

The unit of analysis is the AI output as it is situated in a decision process. The same technical model may produce a score, prediction, warning, ranking, route estimate, or simulation output, while the governance burden depends on what that output is allowed to influence. A risk score used to screen cases, a warning used to trigger caution, a route suggestion used during evacuation, and a model output used to justify regulatory acceptance do not have the same evidential implications. The framework therefore evaluates AI decision support by output role and decision context rather than by algorithm type alone [12–14].

The source material comprises four kinds of evidence: fire-domain AI and decision support studies [1–8,21]; fire model verification, validation, and evaluation guidance [9–11]; AI governance and risk management standards [12,18,19]; and safety-critical assurance and human–automation literature [13–17,22–25].

2.2. Conceptual Synthesis and Framework Construction

The framework was developed by comparing the four bodies of literature and drawing together the concepts relevant to how AI outputs are used in decisions. It does not pool numerical findings or estimate an overall effect. Fire-domain AI studies commonly report model architecture, data source, performance metrics, or application area. Governance concerns arise later, when the output enters a decision process. The same output may remain exploratory in one setting and become decision-shaping or safety-proximate in another. The synthesis therefore asked how adjacent safety-critical fields handle the movement from model output to decision authority. Four bodies of literature were used. AI governance literature treats AI systems as socio-technical systems whose outputs may operate as predictions, recommendations, or decisions in context [12,18,19]. Fire model verification and validation literature treats model adequacy as bounded by assumptions, scenarios, validation basis, and intended use [9–11]. Safety-critical ML and assurance-case literature treats model or component performance as evidence within a broader system-level argument about hazards, controls, operating context, and acceptable use [13,14,23,24]. Human–automation research explains how decision aids can reshape reliance, oversight, and practical authority under workload, salience, and time pressure [15–17,22,25].

The questions were developed by comparing what the four bodies of literature require before a model output can influence action. Similar concerns were grouped into six questions covering the output's claim, decision role, supporting evidence, time for review, consequences of error, and responsibility. This was an interpretive comparison rather than

a formal frequency count. These questions were translated into the fire-domain framework presented in Section 3.

2.3. Illustrative Mapping Approach

The illustrative mapping assesses whether the framework differentiates decision roles across selected fire-domain and fire-adjacent examples. It is illustrative rather than systematic and is used to compare how similar AI or data-driven outputs acquire different governance burdens across decision contexts. Cases were purposively selected to maximise contrast across decision support roles rather than to represent the prevalence of applications in the literature. Selection required each case to expose a distinct combination of output role, decision proximity, temporal coupling, validation basis, or consequence asymmetry. The FIREBIRD inspection-prioritisation framework and the FSLens visual-analytics system serve as urban anchor cases, together with the data-driven building-fire-risk study From Occurrence to Consequence, because they connect data-driven analytics with municipal inspection, fire-station planning, and building or community risk assessment [1–3]. Flashover prediction, evacuation support, surrogate modelling, and WUI/spatial risk examples are used as boundary cases because they show how AI outputs can move closer to warning, time-compressed guidance, or policy-relevant prioritisation [4–8,21]. These examples sit near the limits of the study's urban planning focus: flashover and evacuation tools approach time-critical life safety decisions; surrogate models can move from scenario analysis towards operational guidance; and WUI/spatial risk tools extend into peri-urban planning and resource allocation. The cases therefore span planning, policy-shaping, warning, and near-operational roles; they are not intended to represent the prevalence of fire-domain AI applications. Each example was read against the same analytical questions: what role does the output perform; what validation basis supports it; how close is it to action or authority; what are the asymmetric consequences of false-positive and false-negative outputs; and what governance risk would arise if the output were used beyond the role supported by its evidence base. The final column in the mapping table is interpretive. It identifies possible role migration risks, not documented failures or unsafe conduct by the cited studies.

3. Results

3.1. Framework Overview

The conceptual synthesis produced a role-sensitive framework for bounded deployment of AI decision support in urban fire risk management. The framework starts from the premise that the governance burden of an AI system depends not only on the algorithm or performance metric, but on the role its output is allowed to play in a decision process. This premise follows from AI governance and safety assurance literature, which treats AI outputs as predictions, recommendations, or decisions situated in socio-technical contexts rather than as isolated computational artefacts [12–14,18,19].

The unit of analysis is the AI output in use. A risk score, warning, route estimate, optimisation result, or simulation output may be low-proximity when used to structure analysis, but more consequential when used to guide action, justify policy, defer inspection, or support clearance. Fire model evaluation guidance already recognises that validity is bounded by model assumptions, scenarios, validation basis, and intended use [9–11]. The framework extends this logic to AI decision support by asking what decision role the output acquires once it enters dashboards, operating procedures, public warning systems, inspection regimes, evacuation tools, or regulatory decisions.

The framework organises this assessment through six dimensions: epistemic role, decision proximity, validation basis, temporal coupling, consequence asymmetry, and governance explicitness. Together, they ask what the output claims, what it influences,

what evidence supports it, how quickly it may be acted on, which error direction matters, and whether operating limits and responsibilities are explicit. These questions draw on fire model V&V, AI risk management, safety-case reasoning, and human–automation research on reliance and authority under workload and time pressure [9–19,22–25].

The operating logic proposed here is that evidence sufficient to warn, rank, investigate, or explore may be insufficient to clear, authorise, defer action, or justify safety acceptance. This warning–clearance distinction is proposed here as an analytical synthesis grounded in intended use validation, consequence asymmetry, safety-case reasoning, and human–automation concerns [9,13–17,22–25]. As an output moves closer to action, clearance, policy justification, or safety-proximate use, the required validation, assurance, uncertainty communication, fallback provision, and governance control increase.

Relationship with Existing Frameworks

NIST AI RMF, ISO/IEC 23894, and ISO/IEC 42001 provide domain-general processes for identifying, assessing, treating, monitoring, and governing AI risk across the system life cycle [12,18,19]. They do not, by themselves, determine what evidential burden attaches when a particular fire-domain output moves from information or prioritisation toward warning, clearance, regulatory acceptance, or operational control. Fire model evaluation and V&V guidance provide a complementary principle: predictive capability is bounded by the model, scenario, implementation, assumptions, and specified use [9–11]. The present framework does not replace either body of guidance. Its incremental contribution is to organise them around the decision role of the AI output in use, extend intended use boundedness to the interfaces and institutional workflows through which authority can migrate, and make warning–clearance asymmetry an explicit fire safety governance test. The six dimensions are therefore a role-sensitive bridge between domain-general AI risk management and fire-specific use validation, rather than six independently novel risk categories.

3.2. Six Dimensions of Bounded Deployment

The six dimensions were selected to capture recurring concerns across the four bodies of literature: what the output claims, what decision it affects, what evidence supports it, how much time is available for review, which error matters most, and whether responsibilities and limits are explicit. Each dimension addresses a different way in which an output may acquire more authority than its evidence supports [9–19,22–25].

Epistemic role asks what kind of claim the AI output makes. In urban fire decision support, an output may detect a possible condition, estimate a risk score, rank cases, forecast development, recommend action, optimise resources, warn of danger, or support an assurance-like claim. These roles carry different implications. A detection output that prompts investigation is not equivalent to a recommendation that guides action, and a prediction is not equivalent to a safety assurance claim [12–14,18,19].

Decision proximity asks how close the output sits to action, policy authority, or operational control. Some outputs support awareness, analysis, or later deliberation. Others influence inspection priorities, investment, resource allocation, or regulatory attention. Safety-proximate outputs may affect evacuation, route selection, entry, withdrawal, emergency escalation, or acceptance of a fire safety claim. Decision proximity concerns what the output is allowed to influence, not merely what the system developer intended it to do [12,15–19,22,25].

Validation basis asks what evidence supports the output for the role in which it is used. Fire-domain AI systems may be evaluated against labelled images, historical incident records, property data, experiments, expert judgement, CFD or FDS outputs,

agent-based simulations, operational trials, or standards-based criteria. These evidence bases do not support identical claims. Fire model V&V guidance is directly relevant because it treats model validity as bounded by assumptions, scenario, implementation, and intended use [9–11].

Temporal coupling asks how much time exists to interpret, contest, or override the output before action is taken. Offline planning tools allow review, sensitivity analysis, and deliberation. Near-real-time and real-time outputs may be acted on before users can inspect assumptions, compare alternatives, or seek independent verification. A strategic resource allocation tool may be close to policy authority but weakly time-coupled; a route support output during an incident may be both time-coupled and safety-proximate [15–17,22,25].

Consequence asymmetry asks whether false-positive and false-negative outputs have materially different implications. In fire decision support, error direction often matters as much as aggregate accuracy. A false warning may cause unnecessary caution, investigation, evacuation, or resource use; a false non-warning may delay response or create false reassurance. This dimension supports the paper’s warning–clearance synthesis: evidence sufficient to warn, investigate, or prioritise may be insufficient to clear, authorise, defer action, or justify safety acceptance [9,12–17,22–25].

Governance explicitness asks whether the system, study, or deployment plan states the conditions under which the output may be used, challenged, overridden, updated, or ignored. Relevant controls include operating limits, uncertainty representation, out-of-domain warnings, fallback behaviour, responsibility allocation, auditability, model version control, data quality checks, and conditions for non-use. AI risk management and management system standards emphasise that AI risk is not controlled by model performance alone; it also requires organisational processes, life cycle management, monitoring, and accountability structures [12,18,19]. Applied together, the six dimensions assess whether the authority granted to an output is supported by its evidence base and deployment controls.

Practical Application in a Fire and Rescue Service

Using Firebird as the example, the output is first classified as a risk ranking. The service would then check whether, in local use, the ranking supports inspector judgement or becomes a basis for deferring low-ranked properties; whether it is validated for local buildings and data; whether there is time for review; what harm may follow from false low scores; and who can override or audit the ranking. If the ranking becomes a reason not to inspect, its decision role has migrated, and stronger validation or controls are required [1]. The assessment can be applied sequentially during procurement, implementation, or periodic review.

- Step 1—Epistemic role. The tool produces a relative fire risk score and a ranked inspection list. Its supported claim is prioritisation, not a declaration that low-ranked buildings are safe.
- Step 2—Decision proximity. If inspectors use the ranking to sequence additional inspections, the role remains advisory or policy-shaping. If a low score removes a building from a mandatory or otherwise justified inspection programme, the output has migrated toward a permissive non-inspection decision.
- Step 3—Validation basis. The auditor examines the incident, property, occupancy, violation, and geospatial data; temporal and geographic validation; calibration across building classes; missing data behaviour; and whether the current building stock and inspection purpose match the evaluated domain.
- Step 4—Temporal coupling. A weekly planning cycle permits review and challenge, so coupling is low. Coupling increases if the score automatically cancels, defers, or suppresses a visit before an inspector can assess other evidence.

- Step 5—Consequence asymmetry. A false high-risk score consumes inspection capacity, whereas a false low-risk score may defer attention to a hazardous building. The latter does not automatically dominate, but its potential life safety consequence requires an explicit rule against treating a low rank as clearance.
- Step 6—Governance explicitness and decision. Deploy the model as a supplemental prioritisation aid only if statutory or risk-based inspection floors remain, uncertainty and reasons are visible, inspectors can override the ranking, overrides and outcomes are logged, drift is monitored, and a named authority owns changes. Using the model to eliminate inspections would require separate validation and authorisation for that role.

3.3. Decision Role Migration and Warning–Clearance Asymmetry

Decision role migration occurs when an output developed, validated, or described for one purpose is reused or interpreted in a different or more consequential role without corresponding reassessment of evidence, uncertainty, fallback, and authority. In urban fire decision support, this can occur when a risk score becomes a reason to defer inspection, when a detection output becomes an automated escalation trigger, when a simulation surrogate becomes operational guidance, or when an absence of warning is treated as evidence of safety.

This migration is not necessarily caused by technical model change. It can arise through deployment conditions: dashboards, thresholds, procurement specifications, institutional routines, user interfaces, public warning systems, or compressed decision time. Fire model V&V guidance already supports the principle that model adequacy is tied to intended use and should not be transferred automatically across contexts [9–11]. Safety-critical ML and safety-case literature extend that logic by treating model or component performance as part of a wider system-level argument about hazards, controls, operating context, and change management [13,14,23,24]. Human–automation research provides a plausible cross-domain mechanism: under workload and time pressure, a concise automated output may influence action more strongly than its nominally advisory status suggests [15–17,22,25]. This mechanism has not been demonstrated here in fire service operations and is treated as a hypothesis for empirical testing, not as a fireground finding. The warning–clearance distinction is the clearest expression of this problem. A warning-oriented output supports protective action under uncertainty: investigation, caution, escalation, evacuation, withdrawal, or further analysis. A clearance-oriented output supports permissive action: remaining in place, entering a space, deferring inspection, accepting a design, or treating a condition as sufficiently safe. These roles do not carry the same error burden. A false warning and a false clearance can both cause harm, but they do not fail in the same direction.

The warning–clearance distinction is used here as an analytical synthesis: warning supports protective action under uncertainty, whereas clearance supports permissive action. A negative warning output should not be treated as positive clearance unless the system has been designed, validated, and governed for that role [9,12–17,22–25].

3.4. Illustrative Mapping of Selected AI Fire Decision Support Studies

The framework was applied to selected fire-domain and fire-adjacent studies to test whether it distinguishes decision roles that would otherwise be grouped under the broad label of AI decision support. The mapping examines whether the framework distinguishes decision roles that would otherwise be grouped under the broad label of AI decision support; it does not rank the studies, assess their technical quality, or estimate the prevalence of governance risks.

The selected examples cover several decision support roles relevant to urban fire governance. Firebird illustrates machine learning risk scoring for municipal inspection prioritisation [1]. FSLens illustrates visual analytics and optimisation for fire station layout and resource planning [2]. From Occurrence to Consequence illustrates building and community fire risk analytics using large-scale fire and building data [3]. Flashover prediction illustrates a safety-proximate warning function where absence of warning should not be treated as clearance [4]. Wildfire and WUI examples are included only as bounded comparison cases where rapid prediction, spatial exposure, or resource prioritisation may inform urban interface decision support [5–7,21]. A-Evac/Aamks illustrates evacuation and route support modelling, which may remain planning-oriented or become safety-proximate if interpreted as live guidance [8].

The examples differ in validation basis. Firebird relies on historical incident, property, and geospatial data; FSLens combines fire records, spatial layout modelling, optimisation, and expert feedback; From Occurrence to Consequence uses large-scale incident, building, and social data; image-based detection studies rely on labelled image, video, sensor, or remote observation data; flashover cGANs are evaluated against limited flashover video data and model tests; WUI and wildfire examples draw on simulation, remote sensing, fire weather, or historical fire data; and A-Evac/Aamks uses a CFAST-coupled simulation environment and stochastic evacuation modelling [1–8,21]. Because these evidence bases support different claims, the mapping focuses on what decision role each output could acquire and what governance risk would follow if that role expanded beyond the evidence base.

Table 1 maps these examples against the framework. The governance explicitness column should be read as an analytical inference about bounded use and possible role migration, not as evidence that the cited studies caused unsafe deployment. The point is to identify the governance risk that would arise if an output validated or presented for one role were reused, embedded, or interpreted in a more consequential role without corresponding reassessment.

Illustrative role migration scenario 1: inspection prioritisation. Suppose a municipal dashboard orders properties through a Firebird-style risk score. Low-scoring properties are placed below the visible inspection queue; staffing constraints mean that the lower queue is rarely reached; a threshold automatically marks some properties for deferred review; and managers are assessed on completion of the recommended high-risk list. Although the score remains nominally advisory, the combined interface, threshold, resource constraint, and performance incentive make it a *de facto* reason for non-inspection. The required controls are visibility of excluded cases, periodic random or rule-based inspection outside the ranking, review of false low outcomes, and explicit authority for deferral.

Illustrative role migration scenario 2: flashover non-warning. Consider a hypothetical incident command display that presents a cGAN-based flashover forecast as a prominent alert but does not display uncertainty or the limits of the video conditions on which the model was evaluated. Under time pressure, the absence of an alert could be read as permission to enter or remain, particularly if the workflow lacks independent temperature, gas, smoke, or structural checks. The cited study does not establish that this occurs in actual operations; the scenario identifies the conditions under which non-detection could migrate into clearance. Appropriate controls include persistent “not a clearance” wording, uncertainty display, independent confirmation, fallback doctrine, and a defined human authority for entry and withdrawal.

Table 1. Role-sensitive mapping of selected AI and adjacent data-driven fire decision support studies across the six dimensions.

Panel A: Output Role, Decision Proximity, and Validation Basis.			
Study/Source	Epistemic Role	Decision Proximity	Validation Basis
Firebird [1]	Commercial property fire risk score and inspection ranking.	Policy-shaping support for inspection prioritisation; migration occurs if a low rank becomes a reason for non-inspection.	Historical municipal incident, property, occupancy, violation, and geospatial data; local transfer and calibration remain relevant.
FSLens [2]	Spatial forecasting, multi-criteria comparison, and candidate fire station layouts.	Policy-shaping support for station and resource planning.	Fire records, spatial modelling, optimisation, multi-criteria analysis, and expert feedback.
From Occurrence to Consequence [3]	Building and community fire risk analysis combining occurrence and potential consequence.	Policy-shaping support for prevention and resource planning; not building-level clearance.	Large-scale incident, building, and social data used for data-driven risk analysis.
Flashover cGANs [4]	Video-based prediction of rapid fire growth and potential flashover.	Warning support; safety-proximate if incorporated into incident command.	Limited flashover video data and model tests reported in the study.
Cheng et al. [5]	Generative surrogate for spatial–temporal wildfire nowcasting.	Scenario awareness or planning; may become decision-shaping if embedded in live procedures.	Study-specific training and test evidence for spatial–temporal wildfire nowcasting.
Burge et al. [6]	Time-resolved wildfire-spread modelling using recurrent convolutional networks.	Scenario analysis or planning; may approach operational support if used in real time.	Study-specific time-resolved wildfire-spread modelling and evaluation evidence.
Jain et al. [7]	Review and synthesis of machine learning applications in wildfire science and management; not a single operational output.	Indirect knowledge and policy support.	Review evidence covering heterogeneous applications, datasets, and validation approaches.
A-Evac [8]	Probabilistic evacuation simulation and exit route planning.	Planning or advisory; safety-proximate if interpreted as live route direction.	CFAST-coupled simulation environment and stochastic evacuation modelling.
Cao et al. [20] adjacent non-AI comparison	Composite evaluation and ranking of community emergency corridor accessibility across physical, social, and information spaces.	Policy-shaping support for community planning, infrastructure investment, and intervention prioritisation; not a safety clearance output.	77 Chinese communities; three dimensions, seven elements, and 17 measures; questionnaires, semi-structured interviews, expert consultation, and entropy-weighted TOPSIS.
Lim and Chae [21] adjacent non-AI comparison	Statistical characterisation of human-caused wildfire frequency, Fire Weather Index thresholds, ignition patterns, and spatial clusters; not a real-time warning or clearance output.	Planning and prevention prioritisation; potential migration toward operational readiness or dispatch if thresholds are embedded in live procedures.	2004–2023 South Korean wildfire records, percentile-based Fire Weather Index classification using negative binomial regression, and bivariate LISA analysis.

Table 1. Cont.

Panel B: Temporal Coupling, Consequence Asymmetry, and Governance Explicitness.			
Study/Source	Temporal Coupling	Consequence Asymmetry	Governance Explicitness/Bounded Use
Firebird [1]	Primarily periodic planning, with time for inspector review.	A false high score consumes inspection capacity; a false low score may defer scrutiny of a hazardous building.	Retain inspection floors, local validation, visible reasons and uncertainty, override authority, audit logs, and drift monitoring.
FSLens [2]	Offline strategic assessment with time for sensitivity analysis and deliberation.	Underestimation may leave an area under-resourced; overestimation may divert scarce resources from other areas.	Expose assumptions, criteria weights, uncertainty, sensitivity, and distributional effects; do not make the output an automatic siting decision.
From Occurrence to Consequence [3]	Offline or periodic planning with time for review.	A false low estimate may obscure a vulnerable building or community; a false high estimate may misdirect prevention resources.	State aggregation and transfer limits, validate locally, and require additional evidence before building-level or life safety use.
Flashover cGANs [4]	Potentially high if the forecast is presented during an incident.	A false warning may prompt unnecessary withdrawal; a false non-warning may create dangerous reassurance.	Treat non-warning as non-detection, not clearance; display uncertainty and domain limits, require independent checks, and retain human entry/withdrawal authority.
Cheng et al. [5]	Potentially high where rapid outputs are used as live triggers.	Overprediction may divert resources; underprediction may understate spread or exposure.	State transfer and out-of-domain limits, retain uncertainty and human review, and avoid autonomous operational or clearance use without further validation.
Burge et al. [6]	Low in design or retrospective analysis; higher if used during an evolving event.	A false high spread estimate may overcommit resources; a false low estimate may understate developing exposure.	Validate the operating domain and update rate, display uncertainty, preserve fallbacks, and prevent automatic escalation or clearance.
Jain et al. [7]	Low for the review itself; application-specific for the systems discussed.	Overgeneralising reviewed findings may misdirect policy or transfer a result beyond its original domain.	Retain application-specific validation and context; do not treat review-level findings as operational authority.
A-Evac [8]	Low for planning; high if outputs are used during evacuation.	A route error may expose occupants to congestion, smoke, or untenable conditions; overly conservative routing may also delay egress.	Do not convert planning output into live instruction without role-specific validation, uncertainty display, fallback routes, and explicit authority.
Cao et al. [20] adjacent non-AI comparison	Offline strategic assessment with time for review and deliberation.	A falsely high accessibility score may defer investment in a vulnerable community; a falsely low score may redirect scarce resources from other communities.	Retain dimension-level results, document weights and sensitivity, permit local review, and prohibit an automatic funding or non-investment rule.
Lim and Chae [21] adjacent non-AI comparison	Low for retrospective or seasonal planning; higher if thresholds are used as live triggers.	A false high-risk classification can divert preparedness resources; a false low-risk classification can understate ignition or WUI exposure.	State geographic and temporal transfer limits, retain threshold uncertainty, validate locally, and prevent autonomous dispatch or clearance rules.

The governance explicitness column identifies bounded use controls and the risk that arises when an output exceeds the decision role supported by its validation basis. The mapping shows three patterns. First, some AI outputs are primarily policy-shaping: they rank buildings, compare spatial coverage, or identify community-level risk. Their governance burden concerns transparency, data quality, sensitivity, equity, auditability, and whether a ranking becomes a justification for non-action. Second, some outputs are warning-oriented or safety-proximate: they indicate possible fire, smoke, flashover, route constraint, or evacuation risk. Their governance burden concerns false reassurance, fallback, authority allocation, and whether absence of warning is treated as clearance. Third, some outputs occupy a boundary zone between exploration and operational use: rapid simulation, surrogate modelling, and near-real-time prediction may be valuable for scenario awareness, but speed can make outputs appear more authoritative than their validation basis supports.

The mapping indicates that governance depends not only on model performance but also on the authority assigned to an output. An output suitable for awareness, prioritisation, or planning may be insufficient for clearance, regulatory acceptance, or safety-proximate action unless its validation basis and controls support that role.

4. Discussion

4.1. Prediction, Assurance, and Decision Role Migration

The framework indicates that model performance alone does not determine an output's acceptable use. Accuracy, agreement with validation data, and simulation fidelity remain relevant, but the governance burden also depends on whether the output remains informational, becomes policy-shaping, or approaches warning, clearance, or other safety-proximate authority. This distinction follows from both fire model validation and safety-critical assurance logic. Fire model evaluation guidance treats validity as bounded by model assumptions, scenario, implementation, and intended use [9–11]. Safety-critical ML and assurance-case literature similarly treats model or component performance as evidence within a wider system-level argument about hazards, operating context, controls, and acceptable use [13,14,23,24]. The contribution of the present framework is to apply that logic to AI outputs whose decision role may change through dashboards, reusable components, institutional routines, or compressed decision time.

Prediction estimates what may happen under stated data, modelling assumptions, and operating conditions. Assurance evaluates whether the evidence and controls are sufficient to manage the associated risks for a specified use. This distinction becomes practical at the decision support layer, where technical outputs become institutional prompts, rankings, warnings, recommendations, justifications, or clearance signals.

4.2. Error Asymmetry: Warning Is Not Clearance

The mapping also shows why false-positive and false-negative errors cannot be treated as mirror image failures. Their significance depends on the action the output enables, delays, suppresses, or authorises. In an inspection prioritisation tool, a false high-risk classification may over-allocate attention, while a false low-risk classification may defer scrutiny of a vulnerable building or community. In a detection or flashover warning tool, a false warning may trigger unnecessary caution, while a false non-warning may create dangerous false reassurance. In an evacuation or route support tool, a false route assumption may expose occupants to congestion, smoke, or untenable conditions.

The warning–clearance distinction follows from this directional difference in error consequences. A warning-oriented output supports protective action under uncertainty: further investigation, caution, evacuation, withdrawal, escalation, or review. A clearance-

oriented output supports permissive action: remaining in place, entering a space, deferring inspection, accepting a design, or treating a condition as sufficiently safe. These are not equivalent decision roles. Evidence sufficient to warn may be insufficient to clear. The distinction is proposed here as an analytical synthesis, not as a settled fire safety doctrine. It is grounded in intended use validation, consequence asymmetry, safety-case reasoning, and human–automation concerns [9,13–17,22–25]. Its practical implication is that a negative warning output should not ordinarily be treated as a positive clearance output unless the system has been designed, validated, and governed for that clearance role. The absence of an alert is not, by itself, evidence of safety; it is only non-detection under the system’s assumptions, data quality, thresholds, and operating limits.

4.3. Implications for Bounded Deployment

Bounded deployment permits AI use within a decision role supported by the available evidence and controls. AI decision support can expand awareness, reveal spatial patterns, support inspection targeting, improve scenario exploration, or assist resource planning. The governance burden increases when outputs become policy-shaping, time-coupled, safety-proximate, or clearance-like.

For municipal fire services, AI risk scores and prioritisation tools can support inspection targeting and prevention planning, provided their assumptions, data quality, uncertainty, and equity implications are visible. They should not become automatic reasons to defer inspection or downgrade concern without review. For regulators and approving authorities, AI-supported submissions should be assessed by decision role. Exploratory or screening use may require a lower evidential burden than use for compliance, evacuation adequacy, tenability, design acceptance, or clearance. For researchers and developers, reporting model accuracy is insufficient. Studies should state the intended decision role, validation domain, uncertainty limits, temporal coupling, failure modes, and conditions for non-use.

Bounded deployment therefore requires alignment between output authority and evidential support. The more an AI output influences action, policy justification, clearance, or safety-proximate decision-making, the more explicit its validation basis, uncertainty communication, fallback rules, auditability, and authority allocation must become [9–14,18,19,23,24].

Table 2 converts the six dimensions into observable audit questions. It follows the output from claimed role to actual workflow effect, tests whether validation supports that effect, and then examines meaningful override, consequential error direction, and explicit accountability.

The dimensions are neither numerically weighted nor compensatory. Epistemic role and decision proximity establish actual use; validation must support that use; temporal coupling and consequence asymmetry affect the assurance burden; and governance explicitness is treated as a condition of deployment.

Bounded deployment can impose costs if every safety-related output is held to maximal assurance. Poorly differentiated requirements may increase compliance costs, delay beneficial adoption, favour large vendors, discourage monitored experimentation, or lead agencies to avoid useful tools. Boundaries may also displace responsibility where a nominal human reviewer lacks the time, information, competence, or authority to challenge the system. A proportionate approach permits low-proximity uses through sandboxes, pilots, advisory modes, monitoring, and staged evaluation, while requiring stronger evidence and controls as outputs approach inspection deferral, clearance, regulatory acceptance, evacuation, or operational control. Authorised roles should be reviewed as evidence accumulates and may expand, contract, or be suspended.

Table 2. Six-step checklist for assessing decision role migration and governance burden in AI-supported fire risk management.

Step	Audit Question	Observable Indicators	Required Response
1	What role is claimed?	Output label and specification: detect, estimate, rank, warn, recommend, clear, or authorise.	Record the intended role, user, decision, operating context, and prohibited uses.
2	What does the output actually change?	It reorders queues, hides cases, triggers or withholds action, pre-populates a decision, or becomes a performance target.	Classify the actual role as informational, advisory, decision-shaping, or decision-controlling.
3	Does validation match the actual role?	Evidence covers a prediction task but not the operational population, threshold, interface, workflow, or clearance use.	If actual use exceeds the validated role, restrict use, add review, or revalidate before deployment.
4	Can a person meaningfully contest or override it?	Time pressure, automation, default settings, poor uncertainty display, or no practicable fallback.	Strengthen uncertainty display, independent checks, override authority, fallback, and training.
5	Which error direction causes the greater harm?	A false low score, non-warning, route recommendation, inspection deferral, or apparent clearance suppresses protection.	Set thresholds and controls for the consequential error direction; do not treat non-warning as clearance.
6	Are authority and accountability explicit?	No named owner, audit log, review trigger, appeal route, expiry condition, or change control process.	Assign ownership, logging, monitoring, review, incident response, and authority for suspension.

4.4. Limitations and Future Research

Several limitations qualify the contribution. First, the illustrative mapping is not a systematic review. It shows how the framework differentiates selected applications, but it does not estimate how frequently governance gaps occur across the AI fire safety literature. A future scoping review could test how often fire AI studies specify decision role, validation domain, uncertainty communication, fallback, and authority allocation.

Second, the role migration risks identified in the mapping are analytical inferences rather than claims of unsafe deployment or invalid technical work. The analysis concerns possible deployment interpretations rather than the technical validity of the cited studies. Third, the boundary zone claim for surrogate modelling and rapid prediction is moderate and uneven across subdomains. It is strongest where studies claim rapid, near-real-time, real-time, or operational relevance, and weaker where tools are clearly framed as design-stage or exploratory aids [4–7]. Accordingly, the operational risk argument applies principally to studies claiming operational relevance, not to surrogate modelling generally.

Fourth, human–automation research provides a plausible cross-domain mechanism: under workload and time pressure, a concise automated output may influence action more strongly than its nominally advisory status suggests [15–17,22,25]. This mechanism has not been demonstrated here in fire service operations and is treated as a hypothesis for empirical testing, not as a fireground finding.

Fire-specific research should test this mechanism through observable behaviour. Interviews and workflow observations with incident commanders, inspectors, regulators, control room staff, and developers could compare stated advisory roles with actual effects on action and non-action. Scenario-based experiments could vary interface salience, uncertainty display, time pressure, default recommendations, and override requirements, measuring reliance, response time, escalation, deferral, entry or withdrawal, and error detection. Operating procedures, override records, and audit logs could then be analysed for evidence of role migration, including whether controls remain effective as systems and institutional incentives change. Such work requires appropriate ethics review, sampling, consent, and jurisdiction-specific interpretation.

The framework uses AI governance standards and assurance literature as conceptual anchors rather than ready-made fire-specific deployment rules [12–14,18,19,23,24]. Application to particular jurisdictions, building codes, operational doctrines, and municipal governance structures therefore requires local validation and legal interpretation. The next empirical task is to test role-sensitive governance in specific fire management settings.

5. Conclusions

This paper developed a six-dimensional framework for determining when AI outputs in urban fire risk management require stronger validation and governance. Its central concern is decision role migration: a score, warning, simulation, or ranking may acquire greater authority when embedded in dashboards, workflows, procurement specifications, emergency procedures, or regulatory decisions. The six dimensions—epistemic role, decision proximity, validation basis, temporal coupling, consequence asymmetry, and governance explicitness—connect the evidence supporting an output to what it is permitted to influence.

The framework distinguishes prediction from assurance and highlights the directional asymmetry of error: evidence sufficient to warn, investigate, prioritise, or explore may be insufficient to clear, defer action, or justify safety acceptance. The illustrative cases show that inspection prioritisation, resource planning, building risk analytics, detection, flashover prediction, evacuation support, and spatial risk tools impose different burdens because their decision proximity, time pressure, validation basis, and failure consequences differ.

The practical implication is bounded deployment. AI can defensibly support awareness, prioritisation, planning, and scenario exploration when its assumptions, uncertainty, limits, fallback arrangements, audit trail, and responsible authority are explicit. Stronger assurance is required when outputs affect evacuation, tenability, emergency response, inspection deferral, regulatory acceptance, or clearance. Bounded deployment requires the authority assigned to an output to remain within the role supported by its evidence and controls.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article. No new data were created or analysed. Further inquiries can be directed to the author.

Acknowledgments: During the preparation of this manuscript, the author used ChatGPT with the GPT-5.5 model (OpenAI, San Francisco, CA, USA) for editorial review and language refinement. The author reviewed and edited all outputs and takes full responsibility for the content of this publication.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Madaio, M.; Chen, S.-T.; Haimson, O.L.; Zhang, W.; Cheng, X.; Hinds-Aldrich, M.; Chau, D.H.; Dilkina, B. Firebird: Predicting Fire Risk and Prioritizing Fire Inspections in Atlanta. *arXiv* **2016**, arXiv:1602.09067.
2. Chen, L.; Wang, H.; Ouyang, Y.; Zhou, Y.; Wang, N.; Li, Q. FSLens: A Visual Analytics Approach to Evaluating and Optimizing the Spatial Layout of Fire Stations. *arXiv* **2023**, arXiv:2307.12227.
3. Ma, C.; Du, H.; Luan, S.; Dong, E.; Gardner, L.M.; Gernay, T. From Occurrence to Consequence: A Comprehensive Data-driven Analysis of Building Fire Risk. *arXiv* **2025**, arXiv:2503.22689.
4. Yun, K.; Bustos, J.; Lu, T. Predicting Rapid Fire Growth (Flashover) Using Conditional Generative Adversarial Networks. *arXiv* **2018**, arXiv:1801.09804.
5. Cheng, S.; Guo, Y.; Arcucci, R. A Generative Model for Surrogates of Spatial-Temporal Wildfire Nowcasting. *arXiv* **2023**, arXiv:2308.02810.

6. Burge, J.; Bonanni, M.R.; Hu, R.L.; Ihme, M. Recurrent Convolutional Deep Neural Networks for Modeling Time-Resolved Wildfire Spread Behavior. *arXiv* **2022**, arXiv:2210.16411.
7. Jain, P.; Coogan, S.C.P.; Subramanian, S.G.; Crowley, M.; Taylor, S.; Flannigan, M.D. A Review of Machine Learning Applications in Wildfire Science and Management. *arXiv* **2020**, arXiv:2003.00646.
8. Krasuski, A.; Krenski, K. A-Evac: The Evacuation Simulator for Stochastic Environment. *arXiv* **2017**, arXiv:1711.09229.
9. ASTM E1355-23; Standard Guide for Evaluating the Predictive Capability of Deterministic Fire Models. ASTM International: West Conshohocken, PA, USA, 2023. [[CrossRef](#)]
10. Van Hees, P. Validation and verification of fire models for fire safety engineering. *Procedia Eng.* **2013**, *62*, 154–168. [[CrossRef](#)]
11. National Institute of Standards and Technology. *FDS-SMV Manuals: Fire Dynamics Simulator User's Guide, Technical Reference Guide, Verification Guide, and Validation Guide*. NIST Fire Research Division/Firemodels Documentation Portal; National Institute of Standards and Technology: Gaithersburg, MD, USA. Available online: <https://pages.nist.gov/fds-smv/manuals.htm> (accessed on 1 June 2026).
12. National Institute of Standards and Technology. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST AI 100-1; NIST: Gaithersburg, MD, USA, 2023. [[CrossRef](#)]
13. Porter, Z.; Habli, I.; McDermid, J.; Kaas, M. A Principles-Based Ethics Assurance Argument Pattern for AI and Autonomous Systems. *arXiv* **2022**, arXiv:2203.15370.
14. Hawkins, R.; Paterson, C.; Picardi, C.; Jia, Y.; Calinescu, R.; Habli, I. Guidance on the Assurance of Machine Learning in Autonomous Systems (AMLAS). *arXiv* **2021**, arXiv:2102.01564.
15. Endsley, M.R.; Kiris, E.O. The out-of-the-loop performance problem and level of control in automation. *Hum. Factors* **1995**, *37*, 381–394. [[CrossRef](#)]
16. Parasuraman, R.; Riley, V. Humans and automation: Use, misuse, disuse, abuse. *Hum. Factors* **1997**, *39*, 230–253. [[CrossRef](#)]
17. Mosier, K.L.; Manzey, D. Humans and automated decision aids: A match made in heaven? In *Human Performance in Automated and Autonomous Systems*; CRC Press: Boca Raton, FL, USA, 2019; pp. 19–42. [[CrossRef](#)]
18. ISO/IEC 23894:2023; Information Technology—Artificial Intelligence—Guidance on Risk Management. International Organization for Standardization: Geneva, Switzerland, 2023.
19. ISO/IEC 42001:2023; Information Technology—Artificial Intelligence—Management System. International Organization for Standardization: Geneva, Switzerland, 2023.
20. Cao, J.; Wang, Y.; Li, C. Research on Community Emergency Corridor Systems in Urban Fire Risk Governance: An Empirical Study of 77 Chinese Communities. *Fire* **2026**, *9*, 186. [[CrossRef](#)]
21. Lim, C.J.; Chae, H. Characterizing Human-Caused Wildfire Based on the Fire Weather Index in South Korea. *Fire* **2026**, *9*, 147. [[CrossRef](#)]
22. Zhang, Y.; Liao, Q.V.; Bellamy, R.K.E. Effect of Confidence and Explanation on Accuracy and Trust Calibration in AI-Assisted Decision Making. *arXiv* **2020**, arXiv:2001.02114.
23. Borg, M.; Henriksson, J.; Socha, K.; Lennartsson, O.; Sonnsjo Lonegren, E.; Bui, T.; Tomaszewski, P.; Sathyamoorthy, S.R.; Brink, S.; Helali Moghadam, M. Ergo, SMIRK is Safe: A Safety Case for a Machine Learning Component in a Pedestrian Automatic Emergency Brake System. *arXiv* **2022**, arXiv:2204.07874.
24. Jaradat, O.; Graydon, P.; Bate, I. An Approach to Maintaining Safety Case Evidence After a System Change. *arXiv* **2014**, arXiv:1404.6846.
25. Rosbach, E.; Ganz, J.; Ammeling, J.; Riener, A.; Aubreville, M. Automation Bias in AI-Assisted Medical Decision-Making under Time Pressure in Computational Pathology. *arXiv* **2024**, arXiv:2411.00998.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.