

Comprehensive AI-Driven Climate Risk Assessment: A Theoretical Framework for Transcending Human Cognitive Limitations in Asset-Intensive Industries

Shan Vahora

ESGAgent.ai, Brisbane, Queensland, Australia

Abstract Climate risk assessment in asset-intensive industries is constrained by the cognitive limits of expert-led practice, which can evaluate only a small, non-representative fraction of the asset-hazard-scenario space. This chapter presents a multi-agent, generative-AI framework that reframes asset-level climate risk identification as a systematic search problem, and reports a single anonymized case study comparing the framework against an independent expert assessment of the same facility. The framework is organised as a sequential pipeline of specialised agents spanning climate-hazard interpretation, risk enumeration, verification, financial quantification, and disclosure mapping, drawing on CMIP6 and NARClIM 2.0 regional climate projections. On the case facility, the framework reproduced all 28 pathways identified by the expert assessment and extended coverage to 93 validated pathways following automated screening and independent expert review, a 73 per cent retention rate from 127 enumerated candidates, at substantially lower marginal cost. The chapter sets out the framework's data flows and loss model in reproducible detail, examines the implications for engineering asset-management decisions including useful-life revision, replacement scheduling, and impairment testing under AASB S2 and IFRS S2, and treats the framework's assumptions and limitations critically, including the single-facility scope, the absence of ground truth for risk validity, and the resolution constraints of regional climate projections. The contribution is a formalised framework and an illustrative case: the result demonstrates enhanced risk enumeration on that facility and is not offered as a statistically validated claim of superior risk assessment performance.

Keywords Climate risk assessment; Generative AI; Multi-agent systems; Engineering asset management; Physical climate risk; Risk enumeration; AASB S2; IFRS S2; NARClIM; CMIP6; Asset impairment; Useful-life estimation

1 Introduction

Asset-intensive industries such as mining, energy, manufacturing, and transport infrastructure manage portfolios whose useful lives span decades, over which climate conditions will shift materially. For engineering asset management, climate risk is therefore not a peripheral disclosure obligation but a direct input to the core asset-management decisions: lifecycle planning, maintenance and replacement scheduling, capital allocation, and asset performance optimisation under changing operating envelopes. Mispriced physical climate exposure propagates into impairment, useful-life estimation, and resilience investment decisions, with consequences that compound over an asset's life.

Incumbent practice assesses these risks through expert-led consultation. While valuable for contextual judgement, this approach is constrained by well-documented bounds on human information processing. Experts typically evaluate a limited set of primary risks before cognitive load degrades performance; interdependencies across asset systems are frequently missed due to siloed specialisation; and the combinatorial space of asset-hazard-scenario-horizon interactions vastly exceeds what any team can exhaustively traverse. The result is an assessment that samples, rather than enumerates, the risk space.

Existing computational approaches address parts of this gap but not the whole. Physical climate risk tools and catastrophe models quantify specific hazards well but operate hazard-by-hazard and require expert framing of which assets and scenarios to test. Machine-learning applications in climate risk have focused largely on hazard prediction rather than on systematic, asset-level risk enumeration and its translation into asset-management decisions. The use of generative, multi-agent systems to perform exhaustive asset-level climate risk identification, and to connect that identification to engineering asset-management and financial decision outputs, remains underexplored.

This chapter makes four contributions. First, it formalises a multi-agent architecture for climate risk assessment, specifying agent roles, data flows, and the enumeration logic in sufficient detail to support reproducibility. Second, it defines a set of measures for comparing assessment coverage against an expert baseline. Third, it reports a structured, anonymized case study applying the framework to a representative asset-intensive facility and comparing its output against a conventional independent expert assessment of the same facility. Fourth, it examines the implications for engineering asset management and for climate-related financial reporting under AASB S2 and IFRS S2, and sets out the framework's assumptions and limitations critically.

The remainder of the chapter is organised as follows. Section 2 positions the work against existing climate-risk and machine-learning literature. Section 3 establishes the cognitive and combinatorial basis for the approach. Section 4

specifies the framework and methodology. Section 5 sets out the case study design and protocol. Section 6 presents the case study results. Section 7 discusses implications for engineering asset management and financial reporting. Section 8 addresses limitations and assumptions, and Section 9 concludes.

2 Related Work

2.1 Physical climate risk assessment for assets

Established practice decomposes physical climate risk into hazard, exposure, and vulnerability, translating asset-level shocks into financial loss (World Bank, 2022; Bressan et al., 2024). Industry methodologies such as the Physical Climate Risk Appraisal Methodology (PCRAM 2.0; IIGCC, 2025) and commercial datasets (S&P Global Sustainable1, 2025) combine climate science, engineering, and finance to support investment and asset-management decisions. These approaches are rigorous within scope but share a structural property: the analyst frames which assets, hazards, scenarios, and horizons enter the assessment, and the evaluation proceeds largely hazard-by-hazard. The risk space is therefore sampled according to expert judgement rather than exhaustively enumerated.

The cost of this sampling is measurable. Asset-level studies show that aggregate or score-based approaches materially understate exposure: in a study of 1,820 physical assets, financial losses were underestimated by up to 70 per cent when asset-level information was neglected, and by up to 82 per cent when tail acute risks were excluded (Bressan et al., 2024). This establishes the central problem the present work addresses, namely that incomplete traversal of the asset-hazard-scenario space systematically biases risk estimates and, by extension, the lifecycle and capital decisions that depend on them.

2.2 Machine learning in climate risk assessment

A substantial body of work applies machine learning to climate risk. A systematic review of more than 1,200 studies (Zennaro et al., 2021) found that the field is dominated by decision trees, random forests, and neural networks applied to a narrow set of perils, principally floods and landslides, and concluded that compound and cascading hazards, together with forward-looking scenario analysis, remain underexplored in combination with machine learning. These methods are predominantly predictive: they estimate the magnitude or likelihood of a given hazard. They do not perform systematic, asset-level enumeration of the full set of climate-asset interactions, nor do they translate identified risks into the decision outputs, such as useful-life revision, impairment, and resilience prioritisation, required for engineering asset management.

2.3 Multi-agent large language model systems

Recent work has moved from single language-model agents to multi-agent systems that specialise individual models into distinct roles and coordinate their interaction to solve complex tasks (Guo et al., 2024; Tran et al., 2025). Frameworks including AutoGen (Wu et al., 2023), MetaGPT (Hong et al., 2024), and ChatDev (Qian et al., 2024) demonstrate role specialisation and orchestrated collaboration, with domain applications emerging in software engineering, medical decision-making, and financial analysis. This paradigm is well suited to exhaustive, structured tasks because specialised agents can partition a large problem space and process partitions under a consistent protocol. To date, however, this architecture has not been applied to physical climate risk assessment in asset-intensive industries, and the framework presented here is an original design rather than a derivative of any existing multi-agent framework.

2.4 Research gap

The three strands above define the gap this chapter addresses. Incumbent physical-risk methodologies are defensible but expert-scoped and sampling-based. Machine-learning approaches are systematic but predictive rather than enumerative, and are not connected to asset-management decisions. Multi-agent language-model systems offer the enumeration and specialisation capability but have not been brought to bear on climate risk for assets. The contribution of this work is to position a multi-agent generative architecture at this intersection: to treat asset-level climate risk identification as a systematic search problem, and to connect the resulting enumeration to the lifecycle, impairment, and resilience decisions that constitute engineering asset management.

3 Theoretical Foundation

3.1 Cognitive constraints on expert assessment

Expert climate risk assessment is bounded by well-established limits on human information processing. Miller's (1956) work on working-memory capacity and Sweller's (1988) cognitive load theory together imply that an analyst can hold and relate only a small number of interacting variables at once, while the intrinsic complexity of climate science, asset engineering, and regulatory requirements imposes a high baseline load. Faced with this, analysts satisfice in Simon's (1955) sense, seeking a satisfactory rather than a complete account of the risk space.

The omissions that result are not random. Tversky and Kahneman's (1974) work on the availability heuristic and confirmation bias indicates that the risks an expert is most likely to overlook are the unfamiliar, the novel, and the compound, precisely because these are less easily recalled and do not fit established categories. For engineering asset management this is consequential. Assets in heavy industry have lifecycles measured in decades, over which the material risks are disproportionately

the long-horizon, compounding, and cross-system interactions that expert sampling is structurally most likely to miss. The decision quality of lifecycle planning, replacement scheduling, and resilience investment (ISO 55000) is therefore limited not by the competence of experts but by the cognitive form of the assessment itself.

3.2 The combinatorial structure of asset-level risk

The scale of the problem can be made concrete. Consider a facility characterised by 50 distinct physical assets, 8 climate hazard categories, 5 time horizons, 3 warming scenarios, and 10 potential cascade effects. The space of candidate risk pathways is the product of these dimensions, $50 \times 8 \times 5 \times 3 \times 10 = 60,000$. Not every pathway corresponds to a material risk, but the analyst cannot know in advance which do. Constrained by time and cognitive capacity, expert teams typically evaluate on the order of 100 pathways, under 0.2 per cent of the candidate space, and the unexamined remainder includes an unknown number of material interactions. The difficulty is not that experts evaluate risks poorly, but that they can evaluate only a small, non-representative fraction of them.

3.3 Reframing assessment as systematic search

The theoretical move proposed here is to treat asset-level climate risk identification not as a recall task but as a search problem over the candidate space. A computational system is not subject to the working-memory and sequential-processing limits that force human sampling, and can therefore traverse a far larger fraction of the space under a consistent protocol, applying the same analytical rigour to the ten-thousandth pathway as to the first. Where an expert's coverage is bounded by cognition, a computational approach shifts the binding constraint elsewhere.

This relocation is the precise claim of this work, and it is important not to overstate it. A computational approach does not eliminate limitation; it changes its character. It relaxes the cognitive constraint on coverage, but it introduces distinct constraints of its own: the assessment is only as good as the asset, hazard, and climate data available to it; generative models carry their own error modes, including fabrication and miscalibration; and a larger set of candidate risks demands a correspondingly larger verification burden to separate material findings from spurious ones. These constraints are not incidental. They define the conditions under which the theoretical advantage holds, and they are treated directly in the framework's verification design (Section 4) and limitations (Section 8). The contribution of the approach is to move climate risk assessment from a regime bounded by what experts can hold in mind to one bounded by data quality and verification, which are addressable through engineering rather than fixed by cognition.

4 Framework and Methodology

4.1 Design overview

The framework is an original multi-agent design (see Figure 1 for what a sequential multi-agent pipeline architecture looks like). It adopts the now-established paradigm in which specialised agents are coordinated over a shared task (Guo et al., 2024; Wu et al., 2023; Hong et al., 2024), but the agent decomposition, the enumeration protocol, and the financial quantification model described below are specific to this work and are not derived from an existing framework. The system is organised as three sequential layers: a climate data layer, an interpretive layer that maps climate fields to asset-specific hazard probabilities and risk types, and a financial quantification layer that converts those probabilities into a monetised asset impact. The layers are realised as a pipeline of specialised agents in which each agent consumes the structured output of the preceding agent as its sole input, so that the output of one stage forms the well-defined input contract of the next.

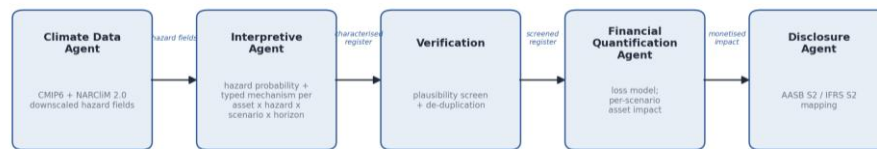


Figure 1 Sequential multi-agent pipeline architecture

4.2 Climate data layer

Hazard characterisation draws on the Coupled Model Intercomparison Project Phase 6 (CMIP6) global ensemble for scenario forcing (Eyring et al., 2016) and on NARCLiM 2.0 for regional detail (NSW Government, 2024). NARCLiM 2.0 dynamically downscales five CMIP6 global models using the WRF regional model at 4 km resolution for south-eastern Australia and 20 km for the wider Australasian domain, under the SSP1-2.6, SSP2-4.5, and SSP3-7.0 scenarios across the period to 2100. Because the case facility lies in south-east Queensland, outside the 4 km south-eastern Australia domain, the assessment uses NARCLiM 2.0 at the 20 km Australasian (CORDEX) resolution under the SSP1-2.6 (low) and SSP3-7.0 (high) scenarios. The coarser 20 km resolution relative to the 4 km domain is a known constraint on hazard localisation for this facility and is treated as a limitation in Section 8. Because the Shared Socioeconomic Pathways carry no assigned likelihoods, results are reported separately for each scenario rather than aggregated into a single probability-weighted figure, consistent with the data provider's guidance.

4.3 Interpretive layer

The interpretive layer is the core analytical contribution. For each asset in the facility register and each hazard category, it maps the relevant downscaled climate variables, for example extreme-heat days, rainfall intensity, flood depth, and wind

exceedance, onto two quantities: the annual probability that the hazard affects the asset, and the physical mechanism through which the impact occurs. Rather than the analyst pre-selecting which asset-hazard pairs to evaluate, the layer traverses the full cross-product of assets, hazard categories, scenarios, and time horizons, producing a probability and a characterised risk type for each cell. This systematic traversal, rather than expert sampling, is what distinguishes the approach from the methodologies reviewed in Section 2.

4.4 Financial quantification layer

Each enumerated and characterised risk is converted into a monetary impact by combining the hazard probability with asset and company-specific parameters. For an asset a , hazard h , scenario s , and year t , the gross expected loss in a single enumeration cell is:

$$E[L]_{a,h,s,t} = P(a, h, s, t) \times [d(a, h) \cdot r(a) + \Phi(a, h, s, t)]$$

where P is the annual probability of the hazard affecting the asset, d is the number of days of operation impacted, and r is revenue lost per day of disruption. The physical-damage term Φ is hazard-dependent: for hazards parameterised against asset value it takes the form $\phi(a, h, s, t) \cdot V(a)$, a damage fraction of the exposed asset value; for hazards with a directly estimable repair cost it takes an absolute value $c(a, h, s, t)$. Insurance is applied as a per-event deductible $k(a, h)$, with the insurer assumed to cover the excess, so the owner-retained loss is the gross loss less recovery. The monetised impact for the asset under a given scenario is the discounted sum of retained losses across hazards and horizons:

$$Impact_{a,s} = \sum_t (1 + \delta)^{-t} \sum_h E[L]_{a,h,s,t} - \iota$$

with δ the discount rate and ι the insurance recovery net of the deductible. Reporting the impact separately for SSP1-2.6 and SSP3-7.0 yields the figure used downstream for useful-life revision, impairment testing, and resilience prioritisation (Section 7). Stating this derivation explicitly is what converts the financial outputs from asserted numbers into reproducible ones.

4.5 Pipeline and verification

The agents execute as an ordered pipeline. The climate data agent produces, for each location in the asset register, the relevant downscaled hazard fields. The interpretive agent consumes these and emits a characterised risk register: for every asset, hazard, scenario, and horizon, an annual probability and a typed risk mechanism. A verification stage then screens this register for physical plausibility and de-duplication before it is passed onward, with low-confidence items flagged for human review rather than admitted automatically. Only the screened register proceeds to the financial quantification agent, which combines it with the company-specific parameters to produce the monetised impact, which in turn is passed to the disclosure agent for mapping to AASB S2 and IFRS S2.

Because each stage exposes a fixed output schema to the next, the pipeline is auditable at every boundary: the hazard fields, the characterised register, the screened register, and the monetised output are each inspectable artefacts. This staged transparency supports the explainability and audit-trail requirements discussed in Section 7. Because exhaustive traversal at the interpretive stage produces a larger candidate set than expert sampling, the proportion of candidate risks retained after the verification stage is reported as a quantity in its own right (Section 6), since it bears directly on the credibility of the comprehensiveness claim.

5 Case Study: Design and Protocol

5.1 Purpose and scope

This section reports a single case study applying the framework to one facility and comparing its output against a conventional expert assessment of the same facility. The case is illustrative rather than confirmatory. It demonstrates how the framework behaves on a real asset base and provides a like-for-like comparison against expert practice on one instance. It is not a statistical evaluation, and the figures reported are point observations for this facility, not estimates drawn from a sample with a known distribution. The conditions under which these observations would or would not generalise are set out in Section 8.

The facility is anonymised at the asset owner's request and is referred to as Facility A: a food production facility located on the Brisbane River in south-east Queensland, Australia, with a replacement value of the order of AUD 45 million, treated for the purposes of this assessment as a single high-value production asset. Location and asset profile are retained because they are methodologically necessary, the location determines the applicable downscaled climate projections and the asset profile determines the hazard pathways, while the owner's identity and the commercially sensitive financial parameters are withheld or expressed as ratios where their disclosure would be identifying.

5.2 Comparison protocol and unit of analysis

The expert assessment was conducted independently by an engineering consultancy as a full, paid engagement on the same facility. Both assessments enumerate risk at a single shared unit: the typed asset-hazard-scenario-horizon pathway, defined as a distinct combination of the asset, one characterised hazard mechanism, one SSP scenario, and one projection horizon. A risk is counted once, on each side, when such a pathway is identified. Both assessments were given the same facility scope, the same asset register, and the same climate scenarios, so that any difference in the count reflects coverage of the candidate space and not a difference in counting convention or scope.

5.3 Measures reported

Three measures are reported for the case. Coverage is the number of distinct typed pathways each method identified, together with their overlap, reported as pathways found by both methods, by the expert only, and by the framework only. Validity is the proportion of the framework's pathways retained after the automated verification stage (Section 4.5) and after subsequent independent expert review, which quantifies the false-positive burden of exhaustive enumeration. Efficiency is the time and direct cost of each assessment; because the framework's cost is a marginal compute cost while the expert figure is a fully loaded professional fee inclusive of overhead, liability, and reporting, the two are presented as what each engagement cost in practice rather than as a normalised unit-cost comparison.

5.4 What the case study can and cannot establish

The case study is positioned against five constraints that bound its inference, each treated more fully in Section 8. It is a single facility, so no variance or generalisation can be claimed. The expert baseline is one consultancy's output, and expert assessments vary between teams, so the comparison is against one instance of expert practice rather than expert practice in general. The chosen counting unit, the typed pathway, is one the framework's enumeration architecture is structurally suited to, so coverage measured this way is a measure of enumeration completeness specifically and not of overall assessment quality. Validity is assessed by expert review, which is itself fallible and provides no independent ground truth, and the design can measure false positives but not false negatives, since the true set of material risks for the facility is unknown. Accordingly, coverage is reported as relative coverage between two methods, not as coverage of the true risk space. Reported together and within these bounds, the measures support a defensible account of how the framework performed on this facility, which is the claim the case study is intended to make.

6 Case Study: Results

6.1 Coverage

The framework enumerated 127 candidate pathways for Facility A across two scenarios (SSP1-2.6 and SSP3-7.0), five horizons to 2100, and the hazard mechanisms relevant to a riverside food-production asset. Automated verification (Section 4.5) and independent expert review retained 93 of these, a retention rate of 73 per cent, with 34 candidates removed as physically implausible or duplicative. The conventional expert assessment, conducted independently on the same facility and scope, identified 28 pathways. All 28 expert-identified pathways fell within the framework's 93 retained pathways, so the validated framework output is a strict superset of the expert output: there were no pathways the expert identified that the framework missed. The framework additionally identified 65 validated pathways

not present in the expert assessment. The comparison is summarised in Table 1 and shown graphically in Fig. 2.

Table 1 Coverage of the framework and the independent expert assessment on Facility A

Measure (Facility A)	Count
Pathways identified by both methods	28
Expert only (framework misses)	0
Framework only, validated	65
Total framework retained (validated)	93
Framework candidates removed in verification	34
Total framework candidates enumerated	127
Total expert pathways	28
Validation retention rate	73%

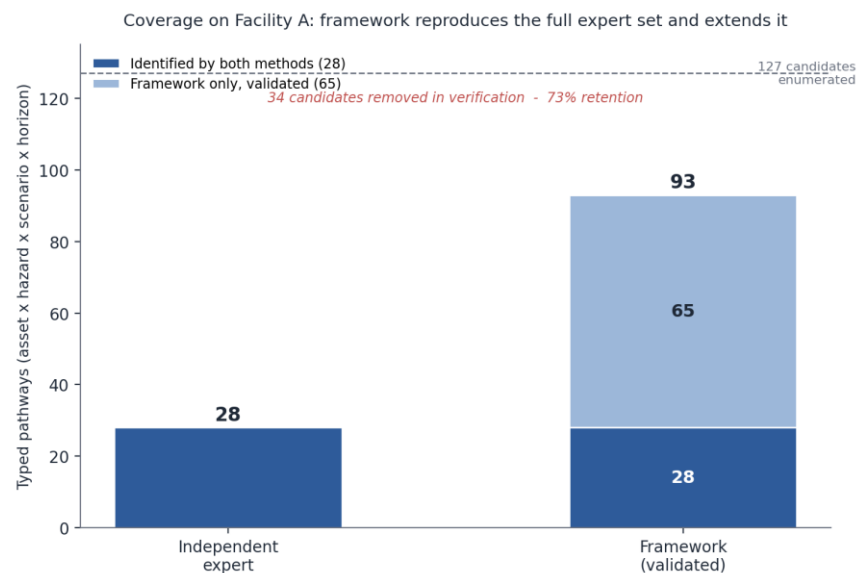


Figure 2 Coverage on Facility A: validated framework output contains and extends the expert set

For this facility, the framework enumerated 3.3 times as many validated pathways as the expert baseline (93 versus 28), with the containment property that it reproduced the entire expert assessment and extended it. This figure measures risk-identification coverage, that is, the breadth of pathways enumerated, and not the accuracy or quality of the resulting risk assessment, for which no ground truth is available (Section 8.2). It is reported as a point observation for one facility, not as an expected multiple across facilities.

6.2 Validity

The 73 per cent retention rate is reported as the central qualifier on the coverage figure. It states directly that exhaustive enumeration carries a false-positive burden, here roughly one candidate in four, and that the comprehensiveness claim rests on the 93 validated pathways rather than the 127 raw candidates. The 65 framework-only pathways were predominantly compound and long-horizon interactions, the categories that Section 3 predicts expert sampling is most likely to omit and that the machine-learning literature (Zennaro et al., 2021) identifies as least addressed by existing methods. This profile is consistent with the framework's design intent, although it should be read as a description of this single case rather than as evidence of a general pattern.

6.3 Financial quantification

Applying the loss model of Section 4.4 to the retained pathways produces a monetised impact for Facility A under each scenario, computed as the discounted sum of owner-retained losses across pathways and horizons. Fig. 3 illustrates the per-pathway computation for the dominant pathway, riverine flooding, showing the decomposition into direct damage, business interruption, emergency response, supply-chain disruption, and insurance effects, and the resulting risk-adjusted net present value over the assessment period. The facility-level impact under each scenario is the aggregate of these per-pathway figures across the retained register; the worked pathway is shown in full to make the model's inputs and discounting transparent rather than to stand in for the aggregate.



Figure 3 Worked per-pathway financial model for the riverine flooding pathway under the SSP3-7.0 scenario, Facility A. The worked example applies the facility's actual AUD 5M single-event policy limit; the general loss model in Section 4.4 is deductible-only.

The downstream use of these per-scenario impact figures for useful-life revision and impairment testing is treated in Section 7.

6.4 Efficiency

The framework completed the assessment of Facility A in approximately 18 hours of computation at a direct cost of the order of AUD 3,200. The independent expert assessment was a full engagement involving three senior experts over six weeks, approximately 240 person-hours, at a fee of the order of AUD 180,000. As noted in Section 5.3, these are not normalised unit costs: the framework figure is marginal compute, while the expert figure is a fully loaded professional fee inclusive of overhead, liability, and reporting. They are reported as the actual cost of each engagement, not as a like-for-like cost ratio.

6.5 Behaviour across additional facilities

While the head-to-head comparison is scoped to Facility A, the framework was subsequently applied to further facilities spanning manufacturing, energy infrastructure, transport networks, and a commercial-property portfolio. No paired independent expert assessment was conducted for these, so they do not support a coverage comparison. The enumeration behaviour was nonetheless consistent with the case facility: candidate registers were dominated by compound and long-horizon interactions, and the proportion of candidates retained after verification was of a similar order. These observations are offered as supporting context on the framework's behaviour, not as evaluation results, and the realised financial

outcomes attributed to those engagements are deliberately omitted here because they were not subject to the controlled comparison applied to Facility A.

7 Discussion

7.1 Implications for engineering asset management

The case study's relevance to engineering asset management lies in the decision outputs the enumeration feeds rather than in the enumeration itself. A per-asset, per-scenario monetised climate impact, decomposed into typed pathways, is a direct input to the central asset-management decisions. For useful-life and depreciation, the long-horizon physical-degradation and inundation pathways revise the expected economic life of the asset and therefore its depreciation profile. For maintenance and replacement scheduling, the pathways with the highest expected retained loss identify where condition monitoring and intervention should be concentrated. For capital allocation, the per-scenario impact allows resilience investments to be ranked by risk reduction per dollar and by payback. For asset performance optimisation, the typed pathways indicate the operating conditions, for example extreme-heat days affecting refrigeration, under which performance is most exposed. In each case the contribution of comprehensive enumeration is to widen the set of pathways considered in these decisions beyond those an expert sample would surface, particularly the compound and long-horizon interactions that a decades-long asset life makes material (ISO 55000).

7.2 Implications for climate-related financial reporting

Comprehensive physical-risk quantification has a direct, and analytically defensible, consequence for climate-related financial reporting. The per-scenario monetised impact, taken net of insurance recovery as the owner-retained loss, provides an input to value-in-use and recoverable-amount calculations for asset impairment testing under IAS 36, and the long-horizon degradation pathways inform the review of useful lives that determines depreciation. Under AASB S2 and IFRS S2, which require disclosure of the anticipated financial effects of climate-related physical risks, the typed and quantified register provides the underlying analysis from which such disclosures can be constructed. The evidentiary standard these treatments require is supported by the pipeline's staged transparency: because each stage exposes an inspectable artefact (Section 4.5), the path from a climate projection to a financial adjustment is auditable. This is framed here as a methodological consequence of comprehensive quantification, and the determination of any specific accounting treatment remains a matter for the preparer and its auditors.

7.3 Explainability, governance, and deployment

Two features of the design bear on its practical use. First, explainability: the staged pipeline produces, for each retained pathway, the data sources, the characterised mechanism, the probability, and the loss computation, which together constitute a reasoning chain suitable for technical and audit review. Second, governance: because exhaustive enumeration produces findings that may inform consequential decisions, the design routes low-confidence items and adjustments above a materiality threshold to human review and professional sign-off rather than admitting them automatically. Deployment is subject to practical requirements summarised in Table 2, principally a complete and current asset register, access to high-resolution climate projection data, sufficient computational capacity for parallel enumeration, and secure integration with the systems that consume the outputs. These requirements, and the constraints they imply, are discussed further in Section 8.

Table 2 Data and computational requirements for deployment

Data requirements	Computational infrastructure
• Comprehensive asset registers with component-level detail • High-resolution climate projection data • Historical incident databases • Regulatory requirement mappings • Industry-specific vulnerability functions • Financial reporting systems and historical accounting data	• Cloud-based parallel processing capability • Acceleration for parallel scenario processing • Distributed storage for climate data • API integrations with climate data providers • Monitoring interfaces for updated projections • Secure connections to financial reporting systems

8 Limitations and Assumptions

The framework and the evidence presented for it are subject to several material limitations, which are set out here directly because they bound the claims the chapter can support.

8.1 Scope and inference

The empirical evidence is a single facility. No variance, confidence interval, or generalisation can be drawn from one case, and the reported coverage multiple and retention rate are point observations specific to Facility A. The expert baseline is the output of one consultancy; because expert assessments vary substantially between teams, the comparison measures the framework against one instance of expert practice rather than against expert practice in general, and a different team might have produced a materially different baseline. Establishing generalisable performance would require assessment of multiple facilities and, ideally, comparison against more than one independent expert baseline per facility.

8.2 Measurement and ground truth

The chosen counting unit, the typed pathway, is one the framework's enumeration architecture is structurally suited to maximise, so the coverage measure should be understood as a measure of enumeration completeness rather than of assessment quality in the round. Validity is assessed by expert review, which is itself fallible and, where the reviewing expert is connected to the framework, not fully independent. More fundamentally, the design can measure false positives, through the retention rate, but it cannot measure false negatives, because the true set of material risks for the facility is unknown. The candidate space is therefore not a meaningful denominator for materiality, and coverage is reported only as relative coverage between two methods, not as a fraction of the true risk space.

8.3 Climate data and uncertainty

For the case facility in south-east Queensland, NARCLiM 2.0 provides regional projections at 20 km resolution rather than the 4 km available for south-eastern Australia, which limits the spatial localisation of hazards such as flooding for this specific site. Structural uncertainty in the underlying global models and in the downscaling propagates into the hazard probabilities produced by the interpretive layer. Because the Shared Socioeconomic Pathways carry no assigned likelihoods, results are reported per scenario and cannot be collapsed into a single expected value without an external and subjective weighting. These uncertainties condition every downstream financial figure.

8.4 Model validity

The generative components of the interpretive layer carry their own error modes, including fabrication and probability miscalibration. The verification stage screens for physical implausibility and duplication and is reflected in the 73 per cent retention rate, but it does not guarantee that retained probabilities are well calibrated, and calibration against historical incident data remains necessary. The

hazard-to-mechanism mapping encodes assumptions about asset vulnerability that require domain validation for asset classes beyond those tested.

8.5 Loss-model assumptions

The loss model applies several simplifications. In its general form insurance is modelled as a per-event deductible with the insurer covering the excess; policy limits and the prospect of cover withdrawal or repricing under rising risk are not represented in the general model, although the worked example in Fig. 3 applies the facility's actual single-event limit. Where such a limit applies and is omitted, the retained loss would be understated. Physical damage is modelled either as a fraction of asset value or as an absolute repair cost depending on hazard type, and the netting of insurance recovery against impact is a reporting choice that affects the figure carried into impairment analysis. Discounting assumptions, and the treatment of cost inflation over long horizons, materially affect the resulting net present values.

8.6 Deployment constraints

Practical deployment depends on data and integration conditions that will not always hold. A complete, current, component-level asset register is required, and its absence degrades coverage; access to high-resolution climate projections is uneven across jurisdictions; and the tacit, contextual knowledge that experienced assessors bring is not fully captured by the data the framework consumes. Professional sign-off remains necessary for consequential adjustments, so the framework is best understood as augmenting rather than replacing expert judgement.

9 Conclusion

Expert-led climate risk assessment is constrained not by the competence of experts but by the cognitive form of the assessment, which forces a small and biased sample of a vast asset-hazard-scenario space. The primary contribution of this chapter is the use of a multi-agent, generative-AI framework to systematically explore the space of asset-level climate-risk pathways and to support engineering asset-management decision-making under uncertainty. To that end the chapter formalises asset-level climate risk identification as a systematic search problem, specifies its data flows and loss model in reproducible detail, and reports a single anonymized case study comparing it against an independent expert assessment of the same facility.

On that facility, the framework reproduced all 28 pathways identified by the expert assessment and extended validated coverage to 93 pathways after screening and independent review, a 73 per cent retention rate from 127 candidates, and did so at substantially lower marginal cost for this engagement. These observations evidence enhanced risk identification and enumeration on the case facility, rather than a generally validated improvement in risk assessment, efficiency, or decision

quality. The result is reported as a point observation, with its false-positive burden stated and its inferential limits made explicit. The broader proposition the framework supports is structural: it relocates the binding constraint on enumeration from human cognition to data quality and verification, which are addressable through engineering. For engineering asset management, this widens the set of pathways feeding useful-life, replacement, capital-allocation, and resilience decisions; for financial reporting, it provides auditable inputs to impairment and disclosure under AASB S2 and IFRS S2.

The evidence offered here is a formalised framework and an illustrative case, not a statistically validated evaluation. The priority for further work is multi-facility assessment against multiple independent expert baselines, calibration of the interpretive layer's probabilities against incident data, and validation of the hazard-to-mechanism mappings across asset classes. These steps would establish whether the enhanced enumeration coverage observed on one facility generalises, and whether broader enumeration translates into measurably better risk assessment.

Acknowledgments

The author thanks the organisations that provided facility data and access for the assessments described here, and the engineering and accounting professionals who provided independent review of the framework's outputs.

Competing Interests

Shan Vahora is the founder of ESGAgent.ai, which is developing commercial applications of the framework described in this chapter. The company has commercial partnerships related to the delivery of climate risk-informed services. This interest is disclosed in the interest of transparency, and the case study has been reported with its methodological limitations made explicit to mitigate the risk of favourable bias.

References

- AASB (2024) AASB S2 Climate-related Disclosures. Australian Accounting Standards Board, Melbourne.
- Bressan G, Duranovic A, Monasterolo I, Battiston S (2024) Asset-level assessment of climate physical risk matters for adaptation finance. *Nature Communications* 15. <https://doi.org/10.1038/s41467-024-48820-1>
- Eyring V et al. (2016) Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development* 9(5):1937-1958.
- Guo T et al. (2024) Large language model based multi-agents: a survey of progress and challenges. *Proceedings of IJCAI 2024*:8048-8057.
- Hong S et al. (2024) MetaGPT: meta programming for a multi-agent collaborative framework. *International Conference on Learning Representations (ICLR)*.
- IFRS Foundation (2023) IFRS S2 Climate-related Disclosures. International Financial Reporting Standards Foundation, London.
- IIGCC (2025) The Physical Climate Risk Appraisal Methodology (PCRAM) 2.0. Institutional Investors Group on Climate Change, London.
- IAS 36 (2023) Impairment of Assets. International Accounting Standards Board, London.
- IPCC (2023) Climate Change 2023: Synthesis Report. Intergovernmental Panel on Climate Change, Geneva.
- ISO (2024) ISO 55000:2024 Asset management - Vocabulary, overview and principles, 2nd edn. International Organization for Standardization, Geneva.
- Miller GA (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review* 63(2):81-97.
- NSW Government (2024) NARCLiM 2.0 climate projections. NSW Department of Climate Change, Energy, the Environment and Water (AdaptNSW), Sydney.
- Qian C et al. (2024) ChatDev: communicative agents for software development. *Proceedings of the Association for Computational Linguistics (ACL)*.
- S&P Global Sustainable1 (2025) Climate Change Physical Risk Exposure Scores and Financial Impact Data Methodology. S&P Global.
- Simon HA (1955) A behavioral model of rational choice. *Quarterly Journal of Economics* 69(1):99-118.
- Sweller J (1988) Cognitive load during problem solving: effects on learning. *Cognitive Science* 12(2):257-285.
- TCFD (2017) Recommendations of the Task Force on Climate-related Financial Disclosures. Task Force on Climate-related Financial Disclosures, Basel.
- Tran K-T et al. (2025) Multi-agent collaboration mechanisms: a survey of LLMs. *arXiv preprint arXiv:2501.06322*.
- Tversky A, Kahneman D (1974) Judgment under uncertainty: heuristics and biases. *Science* 185(4157):1124-1131.
- World Bank (2022) Physical Climate Risk Assessment: Practical Lessons for the Development of Climate Risk Assessment Methodologies. World Bank, Washington DC.
- Wu Q et al. (2023) AutoGen: enabling next-generation LLM applications via multi-agent conversation. *arXiv preprint arXiv:2308.08155*.
- Zennaro F et al. (2021) Exploring machine learning potential for climate change risk assessment. *Earth-Science Reviews* 220:103752.