From HAZOP to Risk Intelligence: Can AI Restore Confidence in Quantitative Risk Assessment?

The issue

For many years, safety and risk assessment has followed a process of progressive simplification. Detailed plant-wide quantitative risk assessments, cumulative societal risk calculations, extensive fault trees and event trees, and formal human-reliability studies have increasingly been replaced by risk registers, matrices, bow ties, barrier diagrams and broadly descriptive accounts of organisational resilience.

These simpler methods are not inherently wrong. They are accessible, economical and useful for communicating with managers. The problem arises when they cease to be summaries of deeper analyses and become substitutes for them. A red box on a risk matrix identifies a concern, but it does not explain the scenarios contributing to it, the dependencies between safeguards, the effectiveness of controls or the uncertainty surrounding the judgement. A bow tie can display barriers without demonstrating whether they are independent, available or effective under actual operating conditions.

The decline of extensive quantitative assessment was partly driven by cost, time and the diminishing availability of specialist expertise. But there was another reason: after a great deal of manual work, the result was often reduced to a single number and presented as ***the answer***.

Even a technically correct number may not tell decision-makers what they really need to know. An average annual loss can conceal whether a system experiences frequent, recoverable failures or retains the potential for one rare catastrophe. Two risks may have the same expected value while being completely different in their distribution, reversibility and social acceptability. An average may also obscure when the risk is concentrated, who bears it, how it changes during maintenance or degraded operation, and how uncertain the underlying model may be.

Quantitative risk assessment therefore lost influence not simply because it was difficult to perform, but because its outputs could become detached from the systems and decisions they were intended to inform. The number appeared precise while its meaning remained incomplete.

What AI changes

The ability of LLMs to assist with preliminary HAZOP studies using detailed P&IDs, PFDs and process descriptions provides an important demonstration of what is now possible. An LLM can identify process nodes, interpret intended functions, apply guidewords, propose deviations, trace possible causes and consequences, and identify existing safeguards. In a matter of hours, it can produce a preliminary analysis that provides a serious basis for expert discussion.

AIHAZOP should be regarded as the beginning of the opportunity, not its final destination.

The deeper capability is the construction of a structured, evidence-linked model of the system. The LLM can read engineering drawings, operating procedures, control narratives, equipment data, maintenance records and incident histories and convert them into a common representation of equipment, functions, flows, operating states, controls, dependencies, human activities, hazards and barriers.

The HAZOP then becomes one analytical view of this deeper model. Its findings can also improve the model by revealing previously unrecognised states, dependencies, failure mechanisms and safeguard requirements. Instead of producing a static worksheet, the HAZOP becomes part of an iterative process through which the model is progressively completed and challenged.

Once that common base exists, the same information can support other analyses. HAZOP causes can become initiating events and fault-tree branches. Consequences can be expanded into event-tree sequences and physical consequence models. Safeguards can become barriers whose availability, effectiveness and independence are examined quantitatively. Operator interventions can be developed into human-reliability analyses, while functional interactions and changing operating conditions can be explored through dynamic system models, resurrecting and reconnecting methods that have traditionally been conducted as separate studies, often by different teams using different assumptions.

From risk register to analytical question

This creates a different role for the Board risk register. A qualitative risk should no longer be treated as the completed assessment. It should become a question posed to the deeper system model.

A Board may identify “loss of critical cooling” as a major concern. The LLM could translate that statement into an investigation of the equipment and functions providing cooling, the operating modes in which it is required, the technical and human causes of its loss, the dependencies capable of defeating redundant systems, the subsequent process response and the barriers intended to prevent escalation.

The system could then determine which specialist studies were required. These might include HAZOP, fault-tree analysis, event-tree analysis, LOPA, human-reliability assessment and physical consequence modelling. The Board would not need to prescribe the methodology. It would identify what matters; the AI-supported system would determine how the concern

should be investigated and, as with AIHAZOP, provide a preliminary analysis for expert examination.

The result might be a quantified exposure, but it should not be forced into a single apparently precise answer. Depending on the available evidence, the risk could be quantified, expressed as a credible range, bounded by an upper estimate, screened as insignificant or declared data-deficient. An explicit statement that the evidence is insufficient is more useful than a number constructed from unsupported assumptions.

More than producing a better number

AI also offers the possibility of looking beyond the immediate technical calculation. A major accident does not end at the physical boundary of the plant. Its consequences can propagate through operational, commercial, environmental, regulatory, legal, social and political systems.

An incident may interrupt dependent services, expose weaknesses in previous assurances, damage relationships with local communities, trigger litigation or public inquiry, lead to regulatory intervention and influence policy across an entire industry. These consequences may ultimately exceed the immediate cost of the physical damage.

An LLM can examine comparable historical events and consider not only their technical causes but also the circumstances that shaped their wider consequences. Were earlier warnings ignored? Did affected communities already distrust the organisation? Was information disclosed promptly? Was the emergency response competent? Did the event become symbolic of a wider institutional failure?

This does not mean that political reactions can be predicted precisely or converted into arbitrary probability multipliers. They should instead be developed as evidence-informed scenarios. The purpose is to ensure that decision-makers see plausible chains of consequence that a narrowly technical calculation would omit. The output becomes a risk profile rather than merely a risk number. It can show the central estimate, credible range, extreme outcomes, conditions in which exposure increases, people and places affected, recovery requirements, systemic dependencies, historical analogues and possible institutional consequences.

Restoring confidence through evidence

None of this means allowing the LLM to become the source of quantitative truth. It should not invent equipment failure rates, human-error probabilities or barrier-effectiveness values. Its role is to find and organise evidence, construct candidate analytical models, identify missing information, call appropriate calculation tools and explain the resulting conclusions.

Every important element should retain its provenance. The system must distinguish what was directly evidenced, imported from an approved database, calculated, assumed, inferred or still missing. Numerical calculations should be performed through validated analytical engines, while qualified specialists continue to review the system boundary, assumptions, dependencies and conclusions.

AI can reduce the labour involved in quantitative analysis, but its greater contribution may be to make the analysis more transparent and open to challenge. A decision-maker should be able to move from a Board-level risk through the contributing scenarios and calculations to the engineering evidence on which they depend.

Risk registers, matrices and bow ties do not disappear. They remain useful management views, but they are generated from the deeper model. Their apparent simplicity no longer requires the underlying analysis to be discarded.

From safety case to executable assurance

The safety case provides perhaps the most telling example of the same problem. Safety cases were intended to move beyond prescriptive compliance by requiring the dutyholder to demonstrate that a particular system was acceptably safe in its actual operating context. Properly constructed, a safety case should establish a clear relationship between safety claims, supporting arguments and verifiable evidence.

In practice, however, it can become an enormous documentary construction, particularly where software, automation, human performance and organisational arrangements interact. The claim that a system is acceptably safe may be supported by layers of subsidiary claims, procedures, compliance statements, test reports and expert judgements that few people can examine as a whole.

The argument may be logically polished while the system itself remains poorly understood. Evidence may demonstrate that approved processes were followed without showing that the resulting system will behave safely. Quantitative claims may rest upon assumptions whose origins are difficult to trace. Software assurance may rely heavily on development procedures and testing while leaving questions about configuration, interfaces, emergent behaviour and operation outside the validated envelope unresolved.

This is where what some critics have called “safety theatre” can arise. The documents, diagrams and review processes display the performance of assurance without necessarily providing a reproducible demonstration that the safety claims are true.

LLMs could make this worse. Given a template, an LLM could generate fluent claims, plausible arguments and apparently comprehensive evidence tables, producing a more persuasive safety case without improving understanding. Prompting the LLM to write a convincing case would therefore be precisely the wrong use.

The LLM should become the cross-examiner rather than the advocate.

The safety case could be converted into a machine-readable network of claims, arguments, assumptions, contexts and evidence. Each claim could then be tested against the hazard analyses, functional model, barrier dependencies, calculations, requirements, software version, test results and operational history. Unsupported or circular claims, obsolete evidence, untested assumptions, contradictory documents and changes that invalidate previous conclusions could all be exposed.

For software, claims could be traced to the relevant requirements, architecture, interfaces, tests and executable version. If the software, configuration, operating procedure or system boundary changed, the affected safety claims could be identified. The safety case would become a living representation of the current assurance state rather than a static document assembled for periodic approval.

The result need not be reduced to pass or fail. Claims could be supported, conditionally supported, contradicted, untested, dependent upon assumptions or based on evidence that no

longer applies. The remaining weaknesses would form an explicit assurance-deficit register showing where further evidence, analysis or risk reduction was required.

Such auditing must remain independent and deliberately challenging. The same AI process should not be allowed to generate the argument, select its evidence and then declare it sound. AI can expose the structure and weaknesses of a safety case, but it cannot certify its own reasoning or decide what level of risk society should tolerate.

Implementation

Current AI HAZOP approaches demonstrate that an LLM can interpret P&IDs, process descriptions and operating information to identify nodes, deviations, causes, consequences and safeguards. The logical next step is to retain that analysis as a structured functional model rather than discard it after producing the HAZOP worksheet.

The AI would identify functional subsystems and develop each as an evidence-linked dynamic module containing its operating states, controls, failure behaviour and human, software and organisational interactions. A candidate Markov blanket would define the interface through which each subsystem exchanges material, energy, information or authority with the rest of the system. These boundaries would be tested for hidden dependencies and common causes rather than simply assumed.

The modules could then be connected through their interfaces, together with explicit environmental and background influences, to form a dynamic model of the complete system. HAZOP would become one view of that model—not its final product. The same underlying representation could generate fault trees, event trees, consequence assessments and quantified risk exposures, while also testing whether safety-case claims are genuinely supported by the available evidence.

This would move AI HAZOP from automating worksheets to constructing the foundation of an AI-supported risk compiler.

This is where we need a practical means of realising the proposition. FRAIXL[1] provides a common environment in which the structured system model, analytical metadata, quantitative parameters, safety claims, documentary evidence, operating data and execution results can be connected.

The LLM can populate, orchestrate and interrogate; FMV can visualise the functional model; specialist engines can perform reliability and consequence calculations; and DuckDB can retain the evidence, scenarios, model versions, runs and operational traces. As experience accumulates, calculated expectations can be compared with actual performance and the model progressively recalibrated.

A reviewer could move from a Board-level risk or top-level safety claim to the relevant hazards, functions, barriers, calculations, software versions and supporting evidence. Conversely, a newly identified hazard, anomalous operating trace or system change could be followed upwards to reveal every risk estimate and safety claim it might invalidate. The LLM assistance could therefore become more than a facility for conducting AI-supported HAZOPs or executing

[1] *FRAIXL is an integrated LLM-enabled systems-modelling facility combining AI reasoning, Excel-based model construction and analysis, FMV visualisation, and DuckDB data storage to create, execute and interrogate evidence-linked functional system models.*

functional models. It could provide both an AI-supported risk compiler and an executable assurance environment.

Conclusion

AI cannot restore confidence in quantitative risk assessment simply by producing more calculations, documents or persuasive numbers. It can do so by making detailed analysis affordable again, cost-effectively reviving and reconnecting different methodologies, exposing uncertainty and placing calculated risk within its operational, historical and societal context.

The real advance is the creation of a living, evidence-linked system model from which HAZOPs, fault trees, event trees, bow ties, human-reliability studies, consequence assessments, Board risk views and safety arguments can all be generated, tested and kept consistent. Risk and safety assessment should never require faith or blin trust. What AI may restore is warranted confidence: confidence that a number has an identifiable meaning, that a safety claim can be challenged, that its evidence can be examined, and that its limitations and wider consequences remain visible. The goal: -Q**ualitative concern in, evidence-linked risk intelligence out; safety claim in, transparent and executable assurance audit out.**

David Slater