Using Eliminative Argumentation to Enhance Trust in ILI Results

Adam Casey, Critical Systems Labs Inc., Vancouver, BC, Canada,
Laure Millet, Critical Systems Labs Inc., Vancouver, BC, Canada,
Jeff Joyce, Critical Systems Labs Inc., Vancouver, BC, Canada,
Vijay Nachiappan, Enbridge, Edmonton, Canada,
Sean Keane, Enbridge, Edmonton, Canada

Proceedings of the ASME 2024 14th International Pipeline Conference
IPC2024
September 23-27, 2024, Calgary, Alberta, Canada

IPC2024 – 134003

USING ELIMINATIVE ARGUMENTATION TO ENHANCE TRUST IN ILI RESULTS

Adam Casey
Critical Systems Labs
Calgary, Canada

Laure Millet
Critical Systems Labs
Vancouver, Canada

Jeff Joyce
Critical Systems Labs
Vancouver, Canada

Vijay Nachiappan
Enbridge
Edmonton, Canada

Sean Keane
Enbridge
Edmonton, Canada

ABSTRACT

Pipeline operators have traditionally relied on unity plots from integrity digs and their confidence in the in-line inspection (ILI) tool vendor as a basis for trust in the results of ILI. However, past digs provide a narrow view of ILI success, and operators have limited visibility into the vendor's equipment and processes.

In this paper, we describe an analytical approach for the pipeline operator and the tool vendor to collaboratively enhance trust in ILI results. Borrowing methods from safety assurance decision-making in the automotive, rail and nuclear power industries, we present a live and reuseable assurance case framework in eliminative argumentation (EA) produced following this approach. This approach covers all factors impacting inspection results, from identifying required inspection performance to equipment and processes used by the vendor. Safety performance indicators derived from the assurance case can be used as warning signs that adverse events might have occurred during the inspection and that an ILI run might require further examination to confirm the trustworthiness of the results.

We also describe our experience applying this methodology to create an assurance case for an actual ILI program. Our experience demonstrated the benefits of involving the vendor directly in constructing the assurance case. The structure of the assurance case clearly defines the causal connection or "golden thread" between the evidence (including indicators) and trust in the inspection. This traceability allows the operator to differentiate between minor deviations from the norm that do not impact the trustworthiness of the ILI results, and anomalies that are of greater concern. Overall, this approach yields a comprehensive, robust, and examinable basis for trust in ILI results while reducing reliance on integrity digs.

Keywords: Assurance Case, Eliminative Argumentation, In-line Inspection, Axial Crack, Safety Performance Indicators, Pipeline Integrity

NOMENCLATURE

AC	Assurance Case
DQA	Data Quality Assessment
EA	Eliminative Argumentation
ILI	In-Line Inspection
SPI	Safety Performance Indicator

1. INTRODUCTION

The trust an operator has in a pipeline's ability to operate at a given pressure is related to the confidence in the integrity condition. The integrity condition can be monitored using multiple inspection methods, whose results inform risk treatment decision-making. One such method is In-Line Inspection (ILI), which uses tools that travel through the pipeline, detecting and identifying defects [1]. Several types of ILI tool are employed in the pipeline industry, each designed to detect and size specific types of defects.

Ultrasonic axial crack tools are of particular interest, as crack inspection technology is rapidly evolving and increasingly complex, with different technologies having varying capabilities [2, 3]. The data analysis used to detect, identify, and size a crack is a multistep process involving both automated algorithms and human decision-making [4]. Because operators rely on the correctness and completeness of the results provided by ILI, it is crucial that operators understand the trustworthiness of these results and the extent to which error or omission could occur.

Throughout this paper, "trust" and "trustworthiness" are defined to mean "*the degree to which decision-makers can be confident that the ILI results accurately and completely describe any and all axial cracks that pose a risk of material harm in the inspected pipeline segment*".

Operators have traditionally relied on two primary sources of trust for ILI results: their confidence in the tool vendor, and unity plots from integrity digs [5]. However, past digs provide a narrow view of ILI success, and operators have limited visibility into the vendor's equipment and processes and the impact these have on the trustworthiness of the ILI results [6].

The vendor's confidence in their ILI system is embodied in the tool performance specification, which defines the tool's expected performance within a specific operational envelope of essential variables including temperature, pressure, and velocity limits. This performance specification is derived from multiple sources including historical data, large scale tests on real or artificial anomalies, and advanced simulations and modelling [7]. The performance specification is a useful starting point for trust in the ILI system, but it can be undermined by operating conditions outside the tool's performance envelope, unique characteristics of the pipeline or defect, human error, and other complexities. ILI tool vendors have established practices and safeguards to ensure that the data they provide to pipeline operators is correct and complete, including compliance with industry standards like API 1163 [8]. Some techniques are common to most vendors, such as qualifying a tool with validation data and providing a data quality assessment (DQA), but each vendor has their own implementation of these techniques, in addition to practices unique to that vendor and their technology. Understanding the specific techniques used by a vendor and any abnormalities of the specific inspection that was performed gives the operator insight into the trustworthiness of the results of that inspection.

The other traditional source of confidence in ILI results is from integrity digs, which involve excavating and manually inspecting sections of the pipeline where anomalies were reported. Besides allowing the operator to repair any defects found, integrity digs also provide data to validate the features reported by ILI. However, integrity digs are expensive, disruptive to the environment, and do not always increase the confidence associated with an ILI run [9, 10], as only a subset of features can be excavated and inspected. Furthermore, field measurements rely on the accuracy of the instruments used and the skills of the technician, introducing potential for error. Finally, digs only provide a comparison point for the end-result of inspection, and offer no insight into the processes and controls used by the vendor. This limits the usefulness of integrity digs as a basis for trust in ILI results.

Building a comprehensive understanding of the trustworthiness of ILI results requires understanding both the factors that could undermine the results, and the controls in place to mitigate those factors. Furthermore, it requires understanding the extent to which these undermining factors and controls have manifested during the inspection and analysis that produced the ILI results, and the effect they had on those results. Existing assurance activity occurs throughout the lifecycle of an ILI system, including when qualifying a vendor and tool, during the execution of a run, and during data analysis and reporting.

While these individual assurance activities conducted by the ILI vendor or the operator provide useful information individually, a more robust trust argument can be established when they are integrated together. By associating the impact of any abnormality from a specific inspection on the trustworthiness of ILI results, the trust argument can also facilitate communication between the ILI vendor and the operator.

Contribution – A methodology to develop a live, reusable assurance case for the trustworthiness of specific ILI results.

This paper describes a methodology for a pipeline operator to construct a reuseable, live assurance case for the trustworthiness of ILI results. This methodology uses close collaboration with the tool vendor and a technique called eliminative argumentation (EA) to construct an assurance case that assesses the entire ILI lifecycle, including tool development, tool qualification, tool selection, inspection, and data analysis. This paper describes a technique to derive safety performance indicators (SPIs) and critical documents from this assurance case, and to use these documents and indicators to evaluate the trustworthiness of the results of specific ILI runs. Altogether, this methodology allows the pipeline operator to build an evidence-based understanding of the trustworthiness of ILI results on a run-by-run basis.

This paper also describes an assurance case for ultrasonic axial crack ILI that was developed using this approach. The high-level structure of the assurance case is shown, along with examples of the critical documents and SPIs that were identified.

2. BACKGROUND

Assurance cases (ACs) are required by internationally recognized safety standards across many safety-critical industries [11, 12, 13, 14, 15, 16, 17] to argue that a system has satisfied the necessary goals and objectives. One commonly used definition of an assurance case is: "A reasoned and compelling argument, supported by a body of evidence, that a system, service or organisation will operate as intended for a defined application in a defined environment." [18] Although often used to argue the safety (a "safety case") or security (a "security case") of a system, they can be used to argue that any given system property holds and are thus generically referred to as "assurance cases". For this paper, we are considering the "trustworthiness" of an ILI system.

Assurance cases can be constructed in a variety of notations [18, 19, 20], including simple written language, i.e., a report. The proposed methodology uses a notation called Eliminative Argumentation (EA) to construct the assurance case and uses safety performance indicators (SPIs) for per-run evaluation of ILI results trustworthiness.

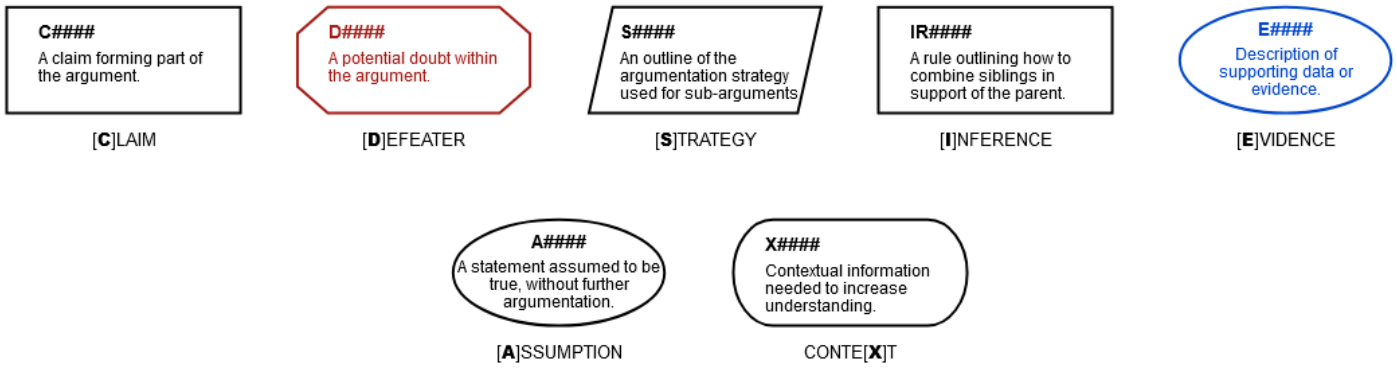

Figure 1 - EA nodes representation in the Socrates assurance case editor.

2.1 Eliminative Argumentation

The methodology proposed in this paper uses a technique called Eliminative Argumentation (EA) [19], which is a graphical notation for assurance cases that builds upon the widely-used Goal Structuring Notation (GSN) [18, 21] by including a notation for doubt called "defeaters". EA encourages the practitioner to think of and record doubts, counterarguments, questions, concerns and potential problems. This helps to combat confirmation bias, which is the tendency to search for or recall information that supports a desired outcome, and to ignore information that disagrees with one's preconceptions [22]. EA allows engineers to reason about system properties such as safety or security and build confidence while also noting objections, concerns, and open questions.

Assurance cases in this notation use a tree-like structure of nodes beginning with a top-level claim or goal about the safety, reliability, or behaviour of a system. The top-level claim is then decomposed into subclaims, which are themselves either decomposed into further subclaims, supported by evidence, or challenged with defeaters. Eventually, each branch is either supported with evidence that is not refuted by defeaters (facts that can be verified outside the argument) or terminated by defeaters that cannot be resolved. This explicit link between a safety goal and evidence is sometimes referred to as "the golden thread" and ensures that decisions are made based on the relevance of the evidence available.

Eliminative Argumentation includes seven main types of nodes (shown in Figure 1) and three types of terminal nodes, described below:

- **Claims** are assertions about the system that can either be supported with evidence or refuted by defeaters.
- **Defeaters** describe a doubt that can be raised in response to a claim, evidence, or inference.
- **Strategy** nodes describe how a claim is decomposed into subclaims.
- **Inference rules** suggest how child nodes can be combined to support a parent claim.
- **Evidence** nodes point to artifacts that support a claim, such as test results.
- **Assumptions** are statements that are assumed to be true for the structure of the argument.
- **Context** nodes provide additional information that is not part of the argument itself but might assist a reader in understanding the argument.

Evidence nodes and defeaters can be followed by a terminal node:

- The **Complete** terminator indicates that a line of reasoning is complete and adequately supported by evidence.
- The **Residual** terminator identifies that there is a residual doubt that is not resolved by the argument and evidence presented and that contributes to the overall doubt in the top-level claim.
- The **Undeveloped** terminator indicates that a line of reasoning has been deliberately left incomplete.

2.2 Safety Performance Indicators

Safety Performance Indicators (SPIs) are numeric values tracked and trended over time and used to evaluate the safety performance of a system [13, 23, 24]. They can be thought of as the safety equivalent of Key Performance Indicators (KPIs), which are often used to measure business performance. SPIs are commonly grouped into two categories:

1. *Lagging SPIs* track the occurrence of safety-related adverse events such as property damage, injury, or death. These are useful long-term indicators that can show the efficacy of new safety measures and provide impetus for change. However, they suffer from *lagging* behind events: They provide information about safety-related events only after they have already happened.
2. *Leading SPIs* track other values and events that are not directly related to harm but are correlated with adverse events or can provide insight into the likelihood of such

events occurring. This prominently includes near-miss events but can also include less direct measures of safety performance.

Both types of SPI are useful and they complement each other [13] by providing insight into different aspects of the system's behaviour. For pipeline integrity one lagging SPI would be the frequency of loss of containment, which directly tracks an adverse event. The frequency at which defects threatening pipeline integrity are discovered would be a leading SPI, because the frequency of discovered defects implies the existence of further, as-yet undiscovered defects. Critical defects are not in themselves adverse events, so long as they are mitigated before harm is caused, but they can be thought of as a near-miss event. For the trustworthiness of an ILI system, the adverse event is not loss of containment, but a failure to detect and report a critical defect in the pipeline. Thus, a lagging SPI for ILI trustworthiness could be the frequency at which critical defects are missed during inspection. A leading SPI for ILI trustworthiness might be the frequency at which data analysis work is found to be incorrect during the vendor's quality-control processes. Although the hypothetical error was caught during quality control, and thus did not lead to a defect going unreported, the error was still made and represents potential for a defect to be missed.

3. METHODOLOGY

Below, we describe the collaborative approach used to construct the assurance case with the tool vendor, methods to derive SPIs from the assurance case, and the means to use the assurance case on a run-by-run basis.

3.1 Constructing an assurance case collaboratively with the tool vendor

The core of the approach presented in this paper is for the operator to construct an AC collaboratively with the tool vendor. By involving subject matter experts (SMEs) from both the operator and the vendor, the AC reflects areas of concern that the operator has observed, challenges the vendor has faced and overcome, and the mitigations that exist. This also ensures that the AC reflects the actual practices used by the vendor and operator, rather than an idealized version described in a standard.

The AC should address everything that impacts the integrity of the final inspection results. For the developed AC, these topics included selection of an inspection tool to meet the needs of the operator for the pipeline under inspection, design of the tool itself, tool qualification, data analysis and reporting.

To develop the AC, working groups were brought together for each topic. The working groups consisted of subject matter experts and front-line workers from both the operator and the vendor who were able to discuss current practices and common sources of error. By documenting this information in a structured assurance case, it formalizes existing conversations that often occur during an inspection and records the information for later use. In this way, the assurance case can be seen as a list of information the operator and vendor should exchange before, during, and after an inspection, together with context for why that information matters and how it can affect the trustworthiness of the ILI results.

3.2 Evidence Types

Dialogue with the vendor, and within the operator's organisation, will produce numerous claims, doubts, and pieces of evidence. An EA assurance case is built of logical claims, but ultimately its foundation rests on evidence. Two types of evidence are used in this methodology:

1. **Document criteria**, which are criteria applied to specific documents, such as personnel qualifications and training records, standard operating procedures (SOPs), review records, tool design documentation, and reliability analyses such as Fault Tree Analysis (FTA) and Failure Modes and Effects Analysis (FMEA).
2. **Safety Performance Indicators (SPIs)**, which are numeric values tracked from run-to-run. Their values can be derived from the ILI results themselves, or from other sources within the vendor and operator, such as the DQA.

While document criteria tend to change only when technology or processes change and provide a baseline level of confidence in the operator and the vendor, the numerical value of the SPIs vary with each new inspection and provide specific information related to that run. Because evidence is tied to specific nodes in the assurance case tree (shown in Figure 3), variation in evidence and SPIs is traceable to its impact on the top-level claim (shown in Figure 2). Evidence should be selected so as to be verifiably correct outside the context of the assurance case. Evidence of dubious origin or correctness introduces doubt into the argument.

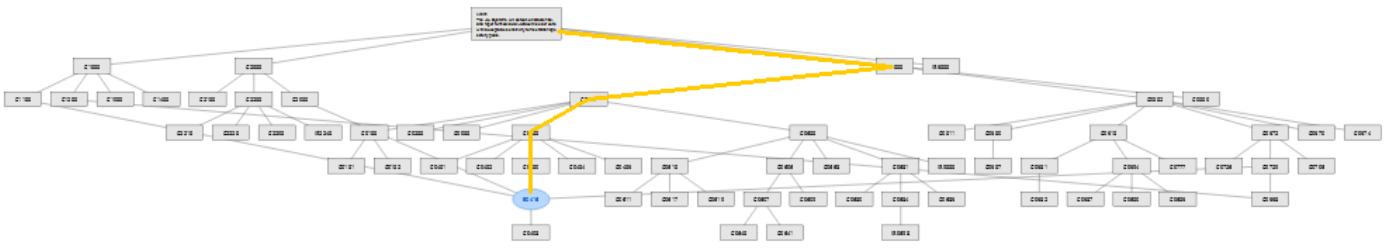

Figure 2 - Example of traceability from an evidence node to the top-level claim. Text in this figure is deliberately illegible.

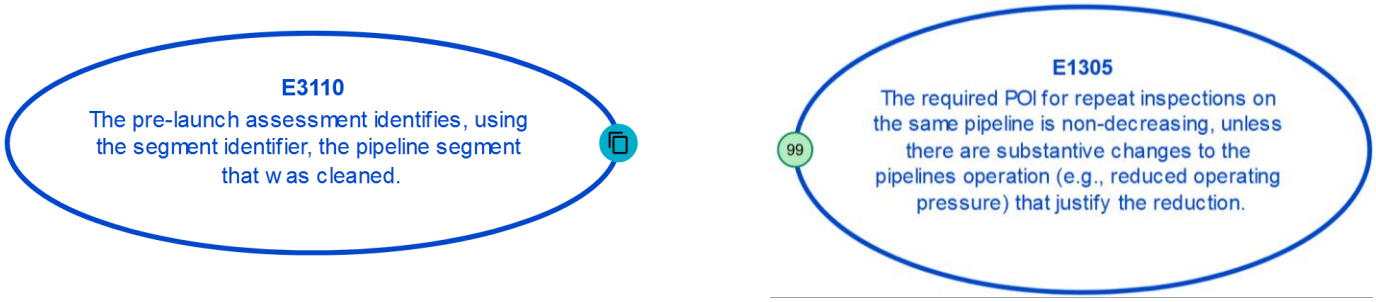

Figure 3 - Evidence nodes with associated artifact (left) and SPI (right)

It is crucial that the AC is constructed to be logically sound, meaning that if all claims at a level are true, then the parent claim is also guaranteed to be true. For the occasional case when it is not possible to construct a deductive argument, an inductive argument is made instead, where the available evidence is stated, and an inference is made as to why this evidence suggests or implies but does not guarantee the parent claim. This logical completeness is critical for the AC to link the evidence to the top-level claim.

3.3 Deriving SPIs from the assurance case

Some SPIs are already commonplace in pipeline integrity, such as essential variable metrics in the DQA. Others emerge naturally during the process of constructing the assurance case. These are typically values that the vendor or operator is already aware of but don't necessarily track or use to evaluate run success. One example of an SPI of this type could be the confidence margin in the capability of the tool to identify critical defects. Finally, some SPIs will only be elicited through deliberate discussion and examination of the assurance case. For each claim and piece of evidence in the assurance case, the practitioner should consider which numeric values represent, support, or refute that node of the argument. Not every node requires an SPI, but many nodes have natural metrics that appear upon examination.

3.4 Operationalizing the assurance case for run-by-run use

Assurance should not be a one-time activity; an assurance case that is created and then shelved is not useful. It must remain live and up to date, or it might lose touch with the system it represents. Furthermore, the assurance case should be used to guide decision-making around that system for its utility to be fully realized. In this case, it should guide decisions of accepting the results of in-line inspection, which are performed after each inspection. Because SPIs provide information on a run-to-run basis, they can be used to proactively assess the success of that run before an integrity dig is performed. Practically speaking, this involves evaluating SPIs for the run against thresholds and typical values. Evaluating run trustworthiness might also involve verifying that other documents are available and correct. Here the assurance case shows its strength – by maintaining a structured, logical connection between evidence (documents and SPIs) and the top-level safety claim (the ILI results are reliable), the operator can easily assess the impact of abnormal SPIs or missing documentation. This activity could be performed by the operator before accepting an inspection run, or while using the ILI results to perform integrity assessments. In both cases, the goal is to provide additional visibility into the run's success and the trustworthiness of the results.

No individual SPI is sufficient to prove or refute the success of an inspection. Rather, an abnormal SPI value indicates a potential problem to be investigated. SPIs must be considered collectively and in the context of other available evidence, the assurance case and the specific inspection being performed. By weighing the degree to which the SPI(s) is or are abnormal, and the risk that represents, the operator can make a better-informed assessment of the degree of trust in the inspection results than is otherwise possible.

Some SPIs are drawn directly from information already provided by the vendor, whether as part of the DQA or in the inspection results themselves. Some SPIs are not directly present in these documents but can be calculated from some combination of values already reported, such as measures of repeatability for repeat inspections of the same pipeline. However, some SPIs rely on information that is known only to the vendor, or even information that the vendor does not typically collect and record, e.g., measures of the degree of similarity between the inspected pipeline and the validation data used to generate the ILI system's performance specification. To capture these SPIs, the operator can update the templates or forms used by the vendor to generate the DQA or inspection report.

With this information available to the operator after each inspection, trust in the run can be evaluated. If all documents are available and meet requirements, and all SPIs are nominal, this lends trust to the results of the inspection. If any documents or SPIs are missing or unusual, then further analysis of the assurance case for that run is required. This relationship is illustrated in Figure 2. This provides more than context – the operator can evaluate redundancy and estimate the impact of an abnormal condition. For instance, if the vendor's data analysts perform less peer-review than usual, the assurance case might show that this is of minimal impact, since quality control works in parallel with analyst training and well-considered standard operating procedures (SOPs) as a mitigation for error in data analysis. However, if SOPs or analyst qualification were unavailable or dubious, then these abnormalities might be cause

for concern. Collectively, these abnormalities pose a risk of inspection data having been analyzed incorrectly, which poses the risk that a critical defect in the pipeline could have been missed during analysis and consequently not included in the inspection report.

SPIs can be used to assess confidence in two main ways:

1. **Thresholds**. For some SPIs, a high or low limit can be defined, outside of which the SPI's value is considered abnormal and might warrant further investigation. An example commonly reported in existing DQAs is "The percentage of the pipeline's surface area for which essential variables were outside the tool's performance bounds," with a threshold for run success of perhaps 95%.
2. **Comparison to past runs**. For other SPIs, the "correct" value could vary significantly between vendors, operators, and pipelines. In this instance, a fixed threshold is not useful. Instead, the value of the SPI can be compared against past runs with the same vendor or pipeline. A significant deviation from previous inspections could indicate a problem or abnormality with the inspection. One example would be "The required POD (probability of detection) and POI (probability of identification) for each defect class". This value is different for each pipeline and might legitimately change due to a change in operating conditions, but large changes in the PODs or POIs used to select the tool for a repeat inspection, without an accompanying change in operating conditions, should raise suspicion.

The key to run-by-run use is that the data from SPIs and document criteria appears in the assurance case where it is relevant, so that the evidence is considered in the context of its impact on run success. This "golden thread" separates the use of an assurance case with SPIs from a simple metrics dashboard which provides data without context or meaning.

3.5 Continuous improvement of the assurance case

To be useful in the long term, the assurance case must remain live and up to date. This requires ongoing discussion and collaboration between the operator and the vendor. If the vendor updates or changes their technology or processes, that must be reflected in the assurance case, and likewise if the operator discovers novel defect classes or finds that a tool struggles to redetect defects from previous inspections. The initial assurance case is a strong foundation for trust, but by improving it over time it can embody the ongoing learning and experience of all involved.

SPIs should be reviewed and revised on an ongoing basis. Over time and with use, some SPIs will demonstrate more usefulness than others, and engineers at both the vendor and the operator will discover new SPIs. These can and should be added to the assurance case as an incremental improvement.

4. RESULTS AND DISCUSSION

This section presents a summary of the contents of the assurance case that was constructed following the proposed methodology, describes a high-level view of the assurance case's structure, shows an example of traceability from bottom-level evidence and doubt to confidence in the top-level claim, summarizes the experience of creating an assurance case collaboratively with a tool vendor, and provides an example of how the assurance case could be used to decide which of two conflicting inspection reports from the same line to trust.

4.1 Assurance Case Statistics

The total size and distribution of node types in the argument is shown in Table 1. This AC is similar in size of others reported in literature for other industries [25]. The bulk of the argument is evidence, followed by claims and then defeaters. Claims and defeaters form the logic of the argument, with evidence supporting claims from the bottom. Assumptions are used sparingly, mostly to state scope. Inference and strategy nodes describing the logical connections between claims are concentrated in the higher levels of the argument, where the top-level claim is decomposed, and are less common towards the bottom of the argument, where specific pieces of evidence are discussed. Context is provided throughout the argument, as needed.

About 50 SPIs were proposed during AC development, and a subset of these were planned for operationalisation based on the availability of information. These cover the entire inspection process, from required performance and tool selection through inspection, analysis, and reporting. The assurance case presented in this paper includes a majority of leading SPIs, along with a smaller number of lagging SPIs.

Table 1- Number of nodes in the assurance case, by type. Terminating nodes are excluded from individual counts but included in the total.

Node Type	Count	Percentage
Assumption	3	0.7 %
Inference	9	2.2 %
Strategy	12	3.0 %
Context	18	4.4 %
Defeater	45	11.1 %
Claim	142	35.0 %
Evidence	165	40.6 %
Total	**406**	**100 %**

4.2 High-level overview of the ILI assurance case

The goal of this assurance case is to provide a basis for trust in the results of an ILI run. To that end, the top-level claim being argued is:

> *The ILI system will detect, characterize, and report axial, crack-like critical defects with acceptable sensitivity to meet Enbridge safety goals.*

The degree to which this claim is supported is the degree to which these results can be relied upon for use in safety-critical decision-making about pipeline integrity. This top-level claim is broken down into three branches:

1. Inspection performance requirements and reporting requirements for the inspection are defined to meet Enbridge's safety goals for pipeline inspection.
2. An appropriate ILI system(s) is selected to meet the inspection performance requirements for the pipeline to be inspected.
3. The ILI system and resulting report by the ILI vendor satisfies the system's performance specification and reporting requirements.

This top-level decomposition is shown in Figure 4. The three branches collectively support the top-level claim, and all three must be true for the top-level claim to be true. The first two branches address required performance and the selection of an ILI tool, both of which are performed primarily by the operator. In these branches it is argued that the operator has defined the minimum performance for the inspection required to meet their safety goals and has selected an ILI tool that is nominally capable of meeting or exceeding that minimum performance under the conditions of the pipeline to be inspected, based on the tool's performance specification. The third branch addresses the actual performance of the tool and related analysis by the vendor, which is collectively referred to as the "ILI System". This branch assesses the design and implementation of the ILI tool, the techniques used for analysis, human factors, and the actual conditions in the pipeline during the inspection.

Collectively, these three subclaims logically form the top-level claim. If the required inspection performance defined by the operator is sufficient for their safety goals, the selected ILI system has a performance specification that is better than the required performance and the ILI system meets or exceeds its performance specification, then the actual performance of the system is sufficient for the operator's safety goals.

As discussed in section 3.1, the scope of this AC only includes the trustworthiness of ILI results for use in decision-making. It does not address the correctness of the decision-making itself, such as whether a given defect requires excavation and repair.

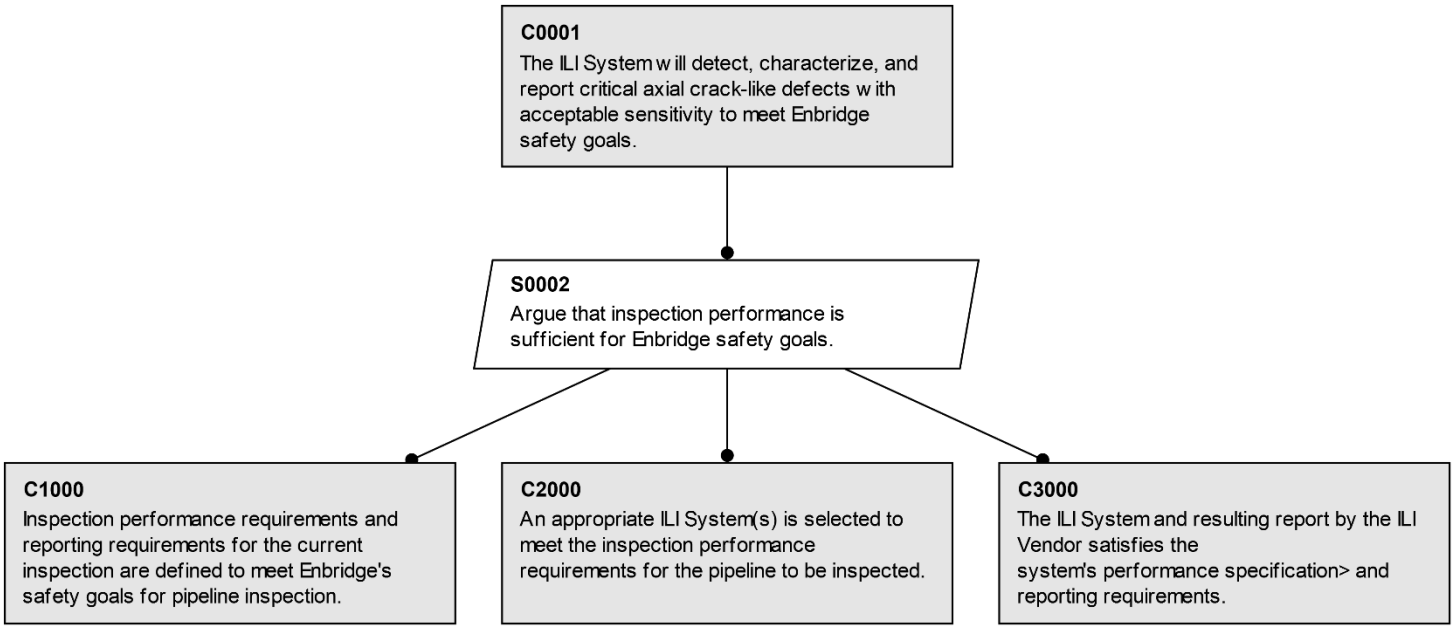

Figure 4 - Top level of the assurance case.

4.3 Argument Structure (Top-to-bottom example)

Under each of the top-level branches described above, the argument is developed down until eventually terminating with either residual doubt or confidence in the evidence presented. Figure 5 shows a single logical thread from the claim that an appropriate ILI system is selected, down to a residual doubt that the anomaly population in the pipeline might be unknown due to latent, undetectable defects. Note that Figure 5 shows only a single thread of the entire argument tree shown in Figure 2.

The initial claim of tool selection being appropriate is supported by arguing following internal tool selection processes is sufficient, given that sufficient information is available to a suitably qualified SME (human factors and process branches omitted). This claim is supported with evidence that data is available for the pipeline under inspection. However, this evidence is undermined by the possibility that the data could be inaccurate (a defeater or doubt). This is partially resolved (side branches omitted) with argumentation about the data sources used, but ultimately a residual risk is identified that latent, defect classes could exist, preventing the SME from fully understanding the anomaly population in the pipeline, which could in theory lead to them selecting an inappropriate inspection tool. This ultimately contributes some small amount of doubt in the completeness of the ILI results, because the selected ILI system might be incapable of detecting this hypothetical latent defect class.

4.4 Our experience developing the assurance case collaboratively with the vendor

A high-level assurance case was prepared by the operator and used to identify key questions and areas of interest for discussion with the vendor. These were addressed in a series of workshops over 6 months, each focused on specific aspects of the inspection process, such as tool development or data analysis. The assurance case was iteratively refined during and between workshops, so that the operator and the vendor could build a common understanding of factors contributing to inspection success. Design and process documents were exchanged, both for use in the assurance case, and to provide context that would otherwise consume face-to-face time during the workshops.

Collaborating with the tool vendor provided useful insight for the assurance case. With tools and data analysis becoming increasingly complex, and much of the new technology being proprietary to each vendor, it is crucial to have input from the vendor themselves.

This collaboration also facilitates dialogue between parties that might not otherwise be in close communication. For instance, the collaboration revealed that the vendor's data analysts often had comments about the nature of identified defects that could not fit in the operator's inspection report template. With this knowledge, the operator could update this template to explicitly include fields for comments of this type. Standardizing reporting this way could make the inspection results more consistent and understandable for decision-makers.

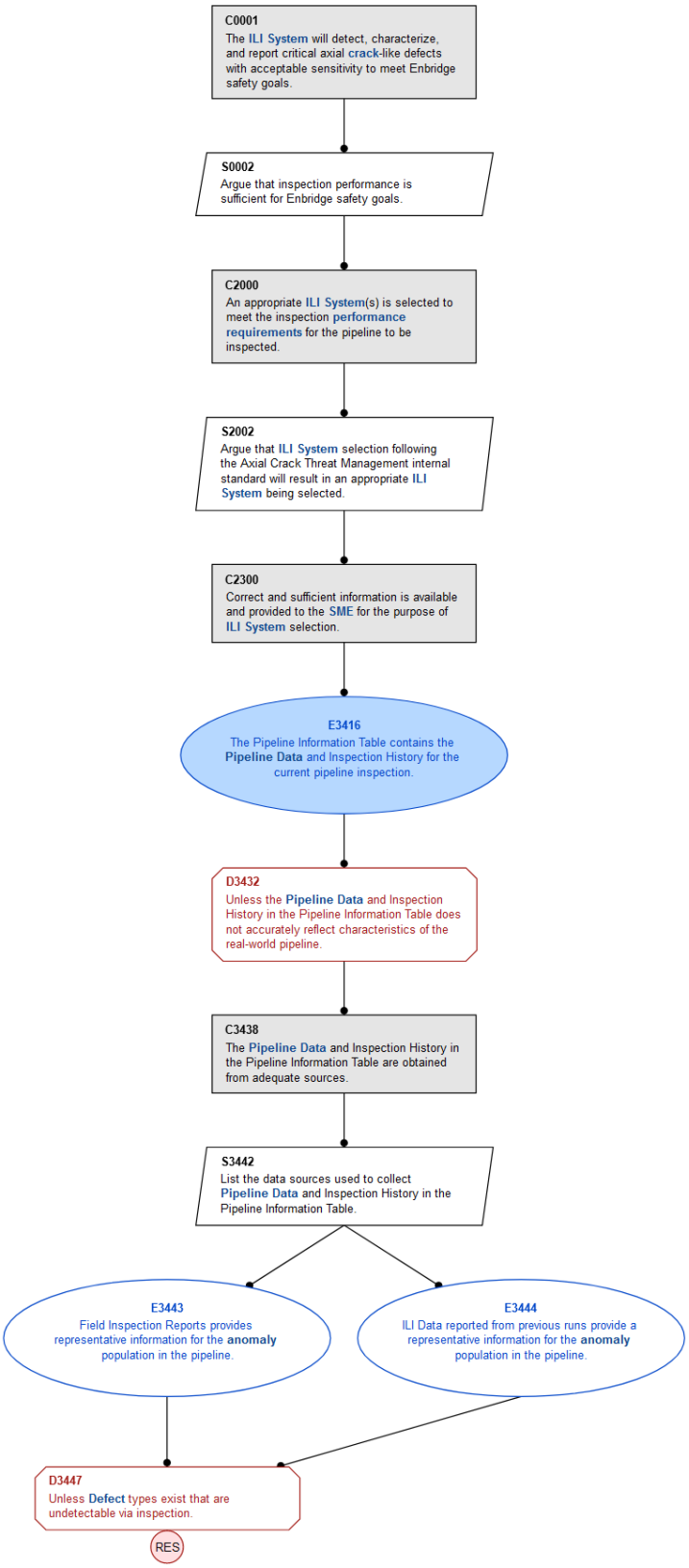

Figure 5 - Example of the logic connecting a high-level claim down to specific evidence and residual doubt. Side branches are not shown.

4.5 Example use case: Comparing two conflicting inspection reports on the same segment

Repeat inspections on the same pipeline segment can provide conflicting inspection results. Figure 6 shows hypothetical results from two crack inspections performed on the same pipeline segment over a short period of time (i.e., three months to one year). These inspections used the same technology, but the results differ in the density distribution and types of reported features, with the second inspection reporting more SCC (stress corrosion cracking) and crack-like features. Given these two confliction reports, the operator has a need to assess which of the two provides more accurate information on the integrity condition of the segment.

The ILI assurance case can be used to systematically evaluate the two runs to determine their relative trustworthiness and provide insight into the source of the discrepancy. The operator could ask:

- Was the line subject to increased pressure cycling?
- Was product composition changed?
- Was the inspection tool configured correctly for each run?
- Did the operating conditions (e.g., temperature and pressure) differ between runs?
- Was a newer version of the same tool used for the second run?
- Did the vendor change their data analysis processes?
- Did either run see significant data loss?

By answering these questions, sources of doubt are eliminated, and the run with higher confidence can be prioritized for final analysis.

5. CONCLUSION

Pipeline operators rely on the accuracy of the results of in-line inspection and benefit from structured visibility into the processes and technologies that produced those results. This paper presents an approach to construct an assurance case for the trustworthiness of ILI results in collaboration with the tool vendor.

This paper also includes a high-level overview of an assurance case constructed following this approach. About 50 SPIs were proposed during AC development, and a subset of these were planned for operationalisation based on availability of information. These cover the entire inspection process from required tool performance, tool selection, run execution, data analysis and reporting.

The constructed AC promises several benefits:

- **Trust in ILI results.** The AC gives the operator a structured view of the controls employed by the vendor to manage material risks.

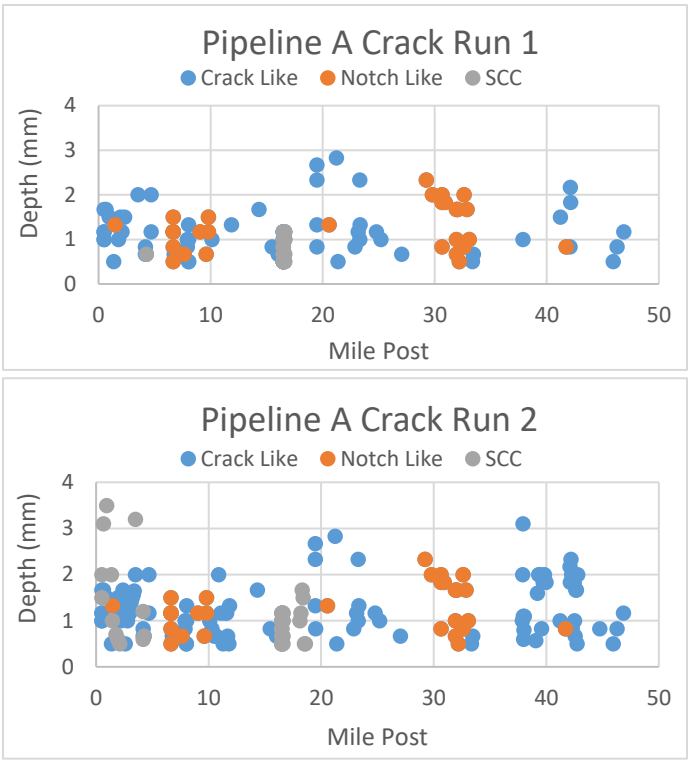

Figure 6 - Hypothetical results from repeat crack inspections on the same pipeline segment

- **Structured evaluation reflecting a pipeline's unique characteristics.** The use of SPIs highlights areas that warrant further investigation due to the properties of the specific pipeline under inspection.
- **Facilitate knowledge transfer.** The scope of the AC facilitates knowledge sharing both between the operator and the vendor, and within each of the organisations. The "golden thread" from evidence to the top-level claim allows this technical knowledge to be communicated to decision makers.
- **Amenable to technology related change management.** As ILI technology continues to evolve, the AC provides a flexible structure for operators to understand the potential impacts of these changes.

Operators can use the methodology described in this paper to prepare assurance cases for their own ILI programs, potentially using the AC described in section 0 as a template or starting point.

ACKNOWLEDGEMENTS

A cooperative relationship with the ILI vendor is key to the development and use of the assurance case. The authors thank NDT Global for their collaboration in developing the assurance case and for their feedback on the derived SPIs.

REFERENCES

[1] M. Xie and Z. Tian, "A review on pipeline integrity management utilizing in-line inspection data," *Engineering Failure Analysis,* no. 92, pp. 222-239, 2018.

[2] S. Bott, R. MacKenzie, M. Hill and T. Hennig, "At the Forefront of In-Line Crack Inspection Services: A Highly Versatile Crack Inspection Platform for Complex Flaw Morphologies and Absolute Depth Sizing," in *International Pipeline Conference*, Calgary, 2020.

[3] H. Willems and T. Hennig, "Recent improvements regarding ultrasonic crack inspection of pipelines," Pigging Products & Services Association, NDT Global, 2017.

[4] S. Timashev, "Basic Performance Metrics of Inline Inspection Tools," in *Rio Pipeline 2003 Conference and Exposition*, Rio, 2003.

[5] J. R. Walker, P. Mallburn and D. Balmer, "Inline Inspection: Both Effective Data Collection and Interpretation Needed to Achieve High Quality Reporting Results," in *International Pipeline Conference*, Calgary, 2010.

[6] S. Keane, K. Cheng and K. Korol, "Systematic Approach to Measuring ILI Tool Performance," in *International Pipeline Conference*, Calgary, 2014.

[7] Pipeline Operators Forum, "Specifications and Requirements for Inline Inspection of Pipelines, Standard Practice - POF 100," 2021," Pipeline Operators Forum, Netherlands, 2021.

[8] American Petroleum Institute, "API 1163 - In-line Inspection Systems Qualification," American Petroleum Institute, Washington, DC, 2021.

[9] A. Schartner, A. Woo, D. Puri and S. Kariyawasam, " A Prudent Approach to Evaluate Dig Effectiveness," in *International Pipeline Conference*, Calgary, 2020.

[10] S. Bott, O. Burdek, R. MacKenzie, M. Hill, T. Hennig, M. Haas and T. S. a. R. Guajardo, "Next Generation ILI crack Inspection Service - an Operator Vendor Collaboration for a 26-inch Pipeline," in *Pipeline Pigging and Integrity Management Conference*, Houston, 2020.

[11] ISO, "ISO 26262 - Road Vehicles - Functional Safety," International Organisation for Standardization, 2018.

[12] United Kingdom Ministry of Defense, "Defence Standard 00-056 - Safety Management Requirements for Defence Systems," 2023.

[13] Underwritter Laboratories, "UL4600 - Standard for Evaluation of Autonomous Products," 2022.

[14] International Organization for Standardization, "ISO 21434 - Road vehicles - Cybersecurity engineering," 2021.

[15] Canadian Nuclear Safety Commission, "REGDOC-1.1.3 - License Application Guide: License to Operate a Nuclear Power Plan," 2016.

[16] CENELEC, "EN 50128 - Railway applications - Communication, signalling and processing systems - Software for railway control and protection systems," 2011.

[17] AAMI Infusion Device Committee, "Medical device safety assurance case guidance," Association for the Advancement of Medical Instrumentation, 2019.

[18] ACWG, Goal Structuring Notation Community Standard Version 3, SCSC, 2021.

[19] J. B. Goodenough, C. B. Weinstock and A. Z. Klein, "Eliminative Argumentation: A Basis for Arguing Confidence in System Properties," Carnegie Mellon University - Software Engineering Institute, Pittsburgh, United States, 2015.

[20] C. M. Holloway, "The Friendly Argument Notation (FAN): 2023 Version," National Aeronautics and Space Administration , 2023.

[21] T. Kelly, "Arguing safety - A Systematic Approach to Safety Case Management," University of York, 1998.

[22] Assurance Case Working Group, "Assurance Case Guidance - Challenges, Common Issues and Good Practice (Version 1.1)," Safety-Critical Systems Club, 2021.

[23] P. Koopman, How Safe Is Safe Enough?: Measuring and Predicting Autonomous Vehicle Safety, Pittsburgh, Pennsylvania, 2022.

[24] C. Rees, M. Delgado, R. Lippelt, J. Joyce, S. Diemert, C. Menghi, T. Viger, M. Chechik, J. Uythoven, M. Zerlauth and L. Felsberger, "Assessing the Usefulness of Assurance Cases: an Experience with the CERN Large Hadron Collider," *Reliability Engineering & System Safety,* 2023.

[25] S. Diemert and J. Joyce, "Eliminative Argumentation for Arguing System Safety - A Practitioner's Experience," in *2020 IEEE International Systems Conference (SysCon)*, 2020.