Towards an ISO/PAS 8800:2024 Compliant Assurance Argument: Assurance Case Development for Artificial Intelligence (AI) and Machine Learning (ML) Systems

Jonathan Groves, Critical Systems Labs Inc., Vancouver, BC, Canada
Ehsan Ghahremani, Critical Systems Labs Inc., Vancouver, BC, Canada
Jeff Joyce, Critical Systems Labs Inc., Vancouver, BC, Canada

Copyright Notice:

Critical
Systems
Labs

WWW.CRITICALSYSTEMSLABS.COM

Towards an ISO/PAS 8800:2024 Compliant Assurance Argument: Assurance Case Development for Artificial Intelligence (AI) and Machine Learning (ML) Systems

Revision 1.0

Jonathan Groves, Critical Systems Labs, Vancouver, BC, Canada
Ehsan Ghahremani, Critical Systems Labs, Vancouver, BC, Canada
Jeff Joyce, Critical Systems Labs, Vancouver, BC, Canada

Abstract

ISO/PAS 8800:2024 provides guidance on the use of artificial intelligence (AI) systems in safety-related functions for road vehicles, which can be extended to AI systems in other industries. The standard provides guidance on producing an AI assurance case argument, which contains logical argumentation and evidence to defend the position that the AI system meets specific AI safety requirements. Although the expectation that an assurance case will be produced is not new, this new ISO/PAS standard for AI Safety shifts emphasis to supporting an argument that *process or product characteristics of the system achieve an acceptably low, although non-zero, level of residual risk associated with the use of the AI system.* This shift should prompt those responsible for managing AI safety risks to reconsider their approach to safety assurance arguments.

Introduction

ISO/PAS 8800 (“Road vehicles — Safety and artificial intelligence”) provides guidance for extending ISO 26262 (“Road vehicles – Functional safety”) and ISO 21448 (“Road vehicles — Safety of the intended functionality”), and is focused on providing automotive-specific guidance with respect to AI systems that either provide functionality or operate as a safety mechanism. Although the examples in ISO 8800 are focused on automotive AI systems, the actual content of this standard is mostly industry-agnostic, i.e., the standard can be easily tailored to other industries (e.g., rail, energy, software development).

ISO 8800 allows for certification of AI systems from QM to ASIL D (i.e., all ASIL levels) for ISO/SAE PAS 22736 levels 1 to 5 of autonomous driving (i.e., all levels of autonomous driving), and it is applicable regardless of the actual AI system being used. The standard addresses the risk of undesired safety-related behaviour at the vehicle level due to output insufficiencies, systematic errors, and random hardware errors of AI elements within the system.

This white paper focuses on ISO 8800’s need for the creation of an AI-specific assurance argument. Among other aspects, this includes demonstrating that AI-specific safety requirements with respect to the expected functionality of the AI system are met to an ideally quantitative (e.g., statistical-confidence-related) manner.

One important difference in thinking from ISO 26262 is that instead of the assurance argument providing evidence that the top claim is true all of the time, an ISO 8800 AI assurance argument provides evidence that the top claim is true to a specified (likely high) percentage of the time; this is due to the statistical nature of many AI/ML systems, where there is a residual risk of the AI/ML system producing an unsafe output. For example, an ISO 8800 assurance argument allows for a percentage of time that the system is unsafe, but argues that that percentage of time is tolerably low with respect to the AI safety requirements and the target functionality of the greater system.

Overview of ISO/PAS 8800

Unlike ISO 26262 and ISO 21448, ISO/PAS 8800 provides a dedicated framework to address the unique challenges of AI-based systems. Previous safety standards in the automotive sector do not fully address some critical issues related to AI systems, like model bias, deficiencies in datasets, and the need for ongoing monitoring as the AI systems or the environment they are in change. ISO/PAS 8800 bridges these gaps by introducing a safety lifecycle tailored specifically to AI systems by building on traditional safety principles adapted to the iterative and data-driven nature of machine learning.

An aspect of ISO/PAS 8800 that stands out is the defining and updating of AI safety requirements throughout the development process. An objective is to ensure that the goal of the AI system behavior, like hitting accuracy targets or keeping failure rates low, aligns with the overall safety goals of the vehicle. This means that care and attention must go into the data, e.g., training data, testing data, and validation data must be sufficient in breadth and variance. The standard also recognizes that operating conditions of an AI-based system are not static. Things like sensor recalibration, out-of-sync updates to other components, or changes to the operational design domain such as roads, traffic patterns, etc., can affect performance, and as such, this standard pushes for regular re-evaluation of the AI system in real-world conditions.

Another important difference of ISO/PAS 8800 is how it considers software testing and validation. Earlier standards, such as ISO 26262 and ISO 21448, treat software as deterministic, but this standard acknowledges that AI/ML software is more complex and might produce unpredictable or novel behavior. It emphasizes testing for edge and corner cases and promotes the use of synthetic or hybrid data to account for this. It also calls for ongoing monitoring of system behavior after deployment to identify any emergent risks or failures.

As mentioned above, a critical addition to ISO/PAS 8800 is the need for an AI-specific assurance argument. This is different from the safety case required within ISO 26262, as it is specifically aimed at AI challenges such as dependence on training data or the black-box nature of AI/ML models and their decision-making process. ISO/PAS 8800 requires a formal safety assurance argument that shows how the system meets safety requirements and how any remaining risks are mitigated or accepted. This safety case can be updated on an ongoing basis with new monitoring data that can be collected with respect to the implementation of safety performance indicators (SPIs) and key performance indicators (KPIs).

In general, ISO/PAS 8800 provides a framework that covers the entire lifecycle of an AI system, from defining initial requirements, selecting an appropriate model, identifying sufficient testing and validation strategies, and deployment strategies that include continuous monitoring and improvements. While this standard is aimed at road vehicles, the principles within it are also applicable to other industries, such as rail, aerospace, energy, or defense.

Building a Comprehensive AI Assurance argument with Socrates

Starting Early and the Iterative Process:

Assurance arguments are most effective when they are started early in the development lifecycle and are updated throughout the rest of the development lifecycle. They guide early design choices[1] and reflect ongoing safety analysis rather than a one-time event[2]. They capture safety work products and evidence on a continuous basis throughout all development phases and the argument itself should be updated to incorporate new data, new findings, or design changes, effectively driving the development process iteratively.

Clause 8.4.e of ISO/PAS 8800 states that “*Conditions can occur during operation that invalidate the assurance argument due to the complex nature of the environment in which vehicles containing AI systems are deployed. These conditions might include distributional shifts of the input space (e.g. new types of road vehicles, changes in road infrastructure), changes to the technical system (e.g. replacement or upgrade of sensors) or previously undiscovered unknown triggering conditions. A continual, periodic re-evaluation and adaptation of the assurance argument is therefore performed, including an impact analysis of which parts of the assurance argument and associated evidence are to be re-evaluated.*”

Regular updating of the AI assurance argument can be used to identify when conditions affecting its correctness occur, allowing the product requirements to be adjusted to mitigate these new identified sources of harm.

There are a number of features inside Socrates[3] that assist with making this ongoing iterative process simpler and more manageable. For example, the heatmap functionality shows which branches have changed over time at a high level, the issue tracking feature that enables analysts to highlight and address concerns, and the filtering feature which enables analysts to isolate specific branches of the argument that require attention. Respective screenshots of these features are provided below as Figure 1, 2, and 3.

[1] See https://www.criticalsystemslabs.com/resources-hub/2022ISSCDiemert/2022ISSCDiemert.pdf “Incremental Assurance Through Eliminative Argumentation” – 2023 – S. Diemert, J. Goodenough, J. Joyce, C. B. Weinstock

[2] See https://criticalsystemslabs.com/wp-content/uploads/2024/09/Driving-Development-from-the-Safety-Case-merged.pdf “Driving the Development Process from the Safety Case” – 2024 – C. Hobbs, S. Diemert, J. Joyce

[3] See https://criticalsystemslabs.com/socrates “Socrates Assurance Cases”

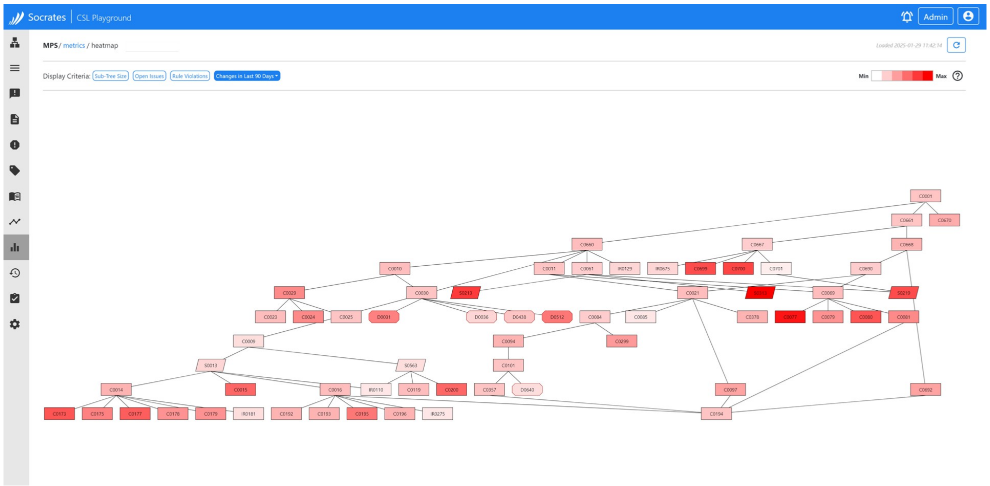
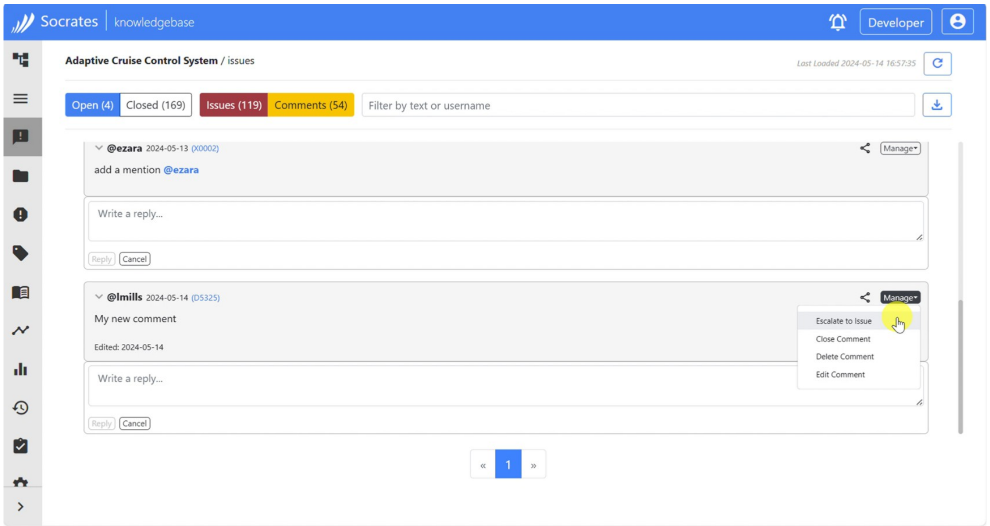


Figure 1: Heatmap Feature of Socrates



@ezara 2024-05-13 (X0002)

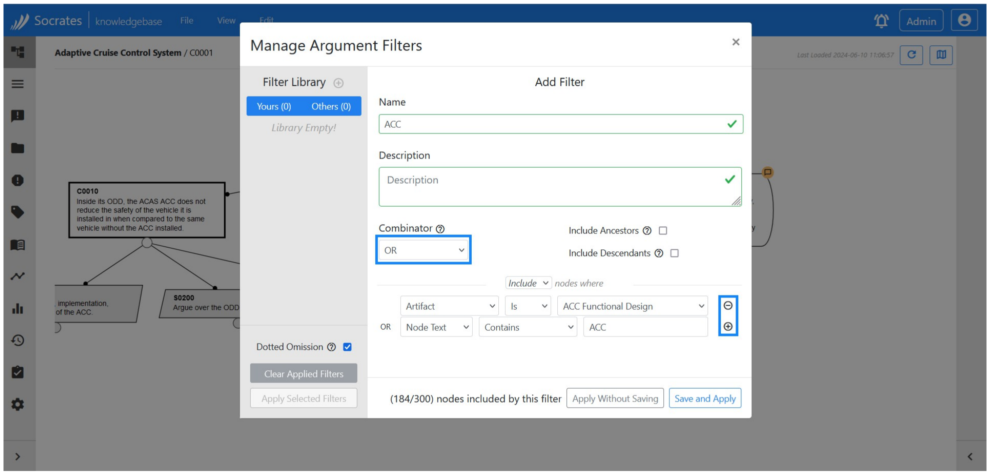
add a mention @ezara

@lmills 2024-05-14 (D5325)

My new comment

Edited: 2024-05-14

Figure 2: Issue Tracking Feature of Socrates



C0010
Inside its ODD, the ACAS ACC does not reduce the safety of the vehicle if it is installed in when compared to the same vehicle without the ACC installed

implementation, of the ACC.

S0200
Argue over the ODD

Name: ACC

Description: Description

Combinator: OR

Include Ancestors ☐

Include Descendants ☐

Include nodes where

Artifact Is ACC Functional Design

OR Node Text Contains ACC

Dotted Omission ☑

(184/300) nodes included by this filter

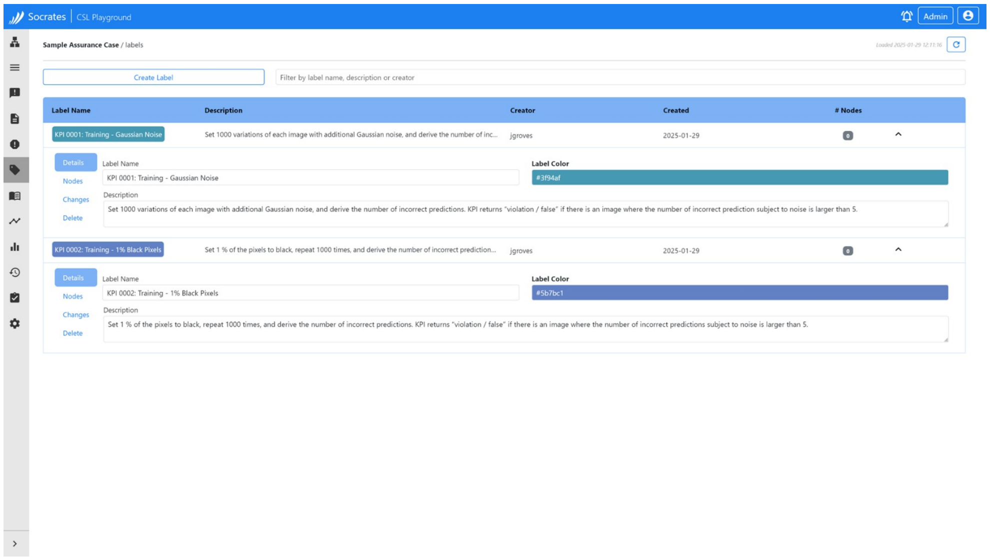
Figure 3: Filtering Feature of Socrates

KPIs and Continuous Monitoring:

Key performance indicators (KPIs), including safety-related KPIs, are quantifiable metrics used to define various properties of a system. They can use used to benchmark the progress of various properties of AI systems. For example, this could include overall AI safety, or other properties such as those listed in Table D-1 of ISO/PAS 8800 (e.g., AI robustness, AI generalization capability, AI reliability, …).

ISO/PAS 8800 references the use of safety-related KPIs, in reference to previous standards (e.g., ISO 21448 and ISO 26262) and in the context of continuous monitoring (see section 6.2 of ISO/PAS 8800), as well as a method for defining pass/fail criteria for testing and validation of requirements, safety related properties of the AI (see Table D-1 of ISO/PAS 8800), model evaluation, etc. See Table 9-4 of ISO/PAS 8800 for examples of two suggested KPIs.

Socrates has built-in features to integrate KPIs into an assurance argument. By tracking KPIs within an assurance argument (e.g., test result data), charts and other data visualization techniques could be used within Socrates to allow users to gain an understanding of to what extent KPIs are met. For example, by tracking KPIs such as perception system camera-based classification accuracy, Socrates users could compare the progress (or regression) of KPIs in real-time across multiple software revisions, without having to manually update the data and visualizations. Since Socrates is developed in-house, Critical Systems Labs can continue to extend Socrates to include features and modifications to suit customer needs.



Sample Assurance Case / labels

Loaded 2025-01-29 12:11:16

Label Name	Description	Creator	Created	# Nodes
KPI 0001: Training - Gaussian Noise	Set 1000 variations of each image with additional Gaussian noise, and derive the number of inc...	jgroves	2025-01-29	0
KPI 0002: Training - 1% Black Pixels	Set 1 % of the pixels to black, repeat 1000 times, and derive the number of incorrect prediction...	jgroves	2025-01-29	0

Label Name: KPI 0001: Training - Gaussian Noise
Label Color: #3f94af
Description: Set 1000 variations of each image with additional Gaussian noise, and derive the number of incorrect predictions. KPI returns "violation / false" if there is an image where the number of incorrect prediction subject to noise is larger than 5.

Label Name: KPI 0002: Training - 1% Black Pixels
Label Color: #5b7bc1
Description: Set 1 % of the pixels to black, repeat 1000 times, and derive the number of incorrect predictions. KPI returns "violation / false" if there is an image where the number of incorrect predictions subject to noise is larger than 5.

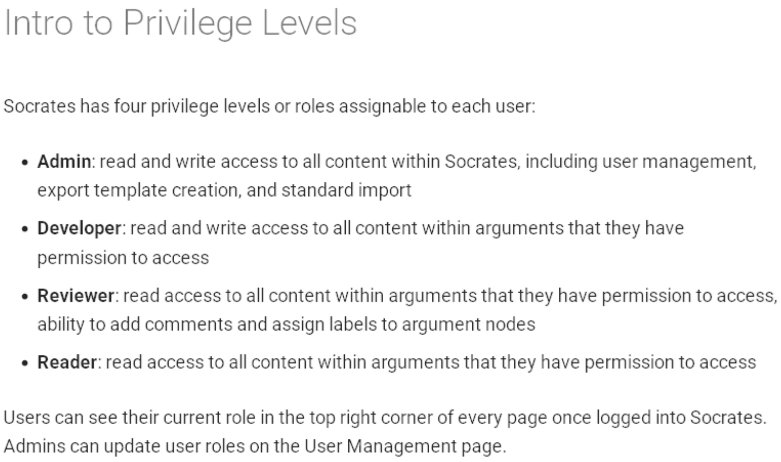
Figure 4: KPIs using the Labelling Feature of Socrates

Stakeholder Involvement and Governance:

Effective use of an assurance argument allows stakeholders at all levels (e.g., design/development/validation teams, vendors, organizational leadership, investors, auditors, …) to simultaneously view, review, and edit the assurance argument content in a manner that is effective for their role.

Clause 8.3.3's note within ISO/PAS 8800 discusses confirmation measures for the assurance argument itself, including the independent or organizationally separate review that could bring together multiple stakeholders from safety engineers and auditors to leadership who must confirm the validity of the argument's claims and evidence. ISO/PAS 8800 points out that the assurance argument can be part of a Safety Element out of Context (SEooC) approach or integrated into a distributed development model which implies that the governance process for creating, reviewing, and approving the various parts of the argument should be delineated to ensure that key stakeholders all contribute to and endorse its content.

Socrates supports **Role-based access control** (RBAC) to assurance arguments, such that read, write, and reviewing permissions for each user can be specified. Currently, there are four privilege levels: Admin, Developer, Reviewer, and Reader, whose roles are detailed in Error: Reference source not found below.



Intro to Privilege Levels

Socrates has four privilege levels or roles assignable to each user:

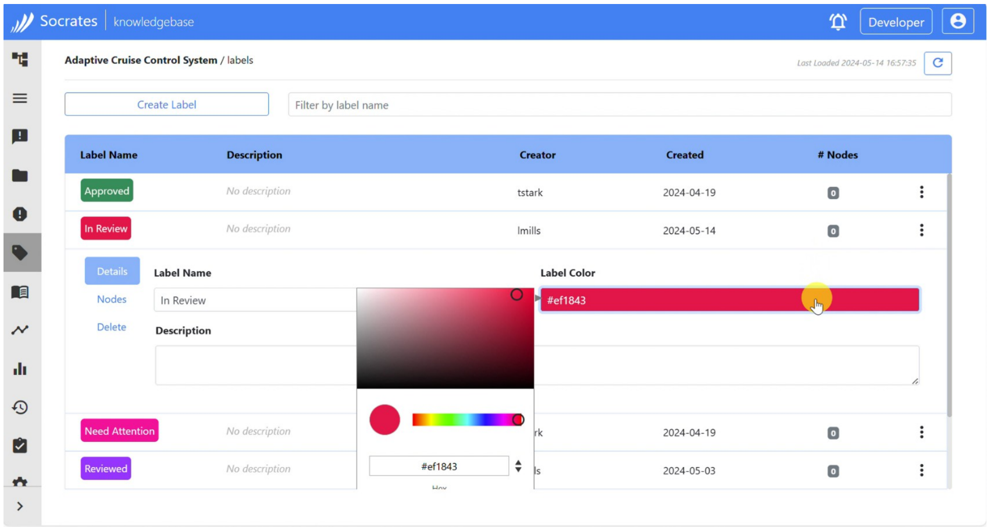
- **Admin**: read and write access to all content within Socrates, including user management, export template creation, and standard import
- **Developer**: read and write access to all content within arguments that they have permission to access
- **Reviewer**: read access to all content within arguments that they have permission to access, ability to add comments and assign labels to argument nodes
- **Reader**: read access to all content within arguments that they have permission to access

Users can see their current role in the top right corner of every page once logged into Socrates. Admins can update user roles on the User Management page.

Figure 5: Role-Based Access Feature of Socrates

Socrates also provides various features to better enable users to perform this collaborative design and review effort across the various stakeholders. These efforts can all be performed simultaneously within Socrates by multiple users, with additional other features used to keep track of efforts and communication:

- **Label management:** Users can add labels to nodes to better guide and coordinate progress and reviews, e.g.,   a section as a work in progress, or communicating which parts of the assurance argument have been reviewed.



Adaptive Cruise Control System / labels

Last Loaded 2024-05-14 16:57:35

Label Name	Description	Creator	Created	# Nodes
Approved	No description	tstark	2024-04-19	0
In Review	No description	lmills	2024-05-14	0
Need Attention	No description	rk	2024-04-19	0
Reviewed	No description	ls	2024-05-03	0

Label Name: In Review

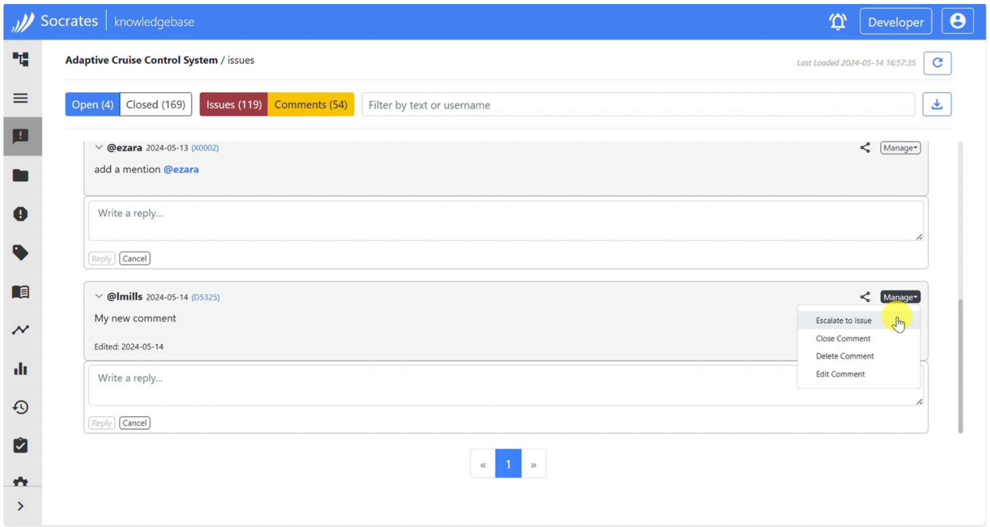
Label Color: #ef1843

Description

#ef1843

Figure 6: Label Management Feature of Socrates

- **Issue tracking**: Users can raise issues about specific nodes within the assurance argument, which automatically email relevant team members. Replies to issues can be written within Socrates, and these issues can be closed out when completed while retaining a record of the discussion.



Adaptive Cruise Control System / issues

Last Loaded 2024-05-14 16:57:35

@ezara 2024-05-13 (X0002)

add a mention @ezara

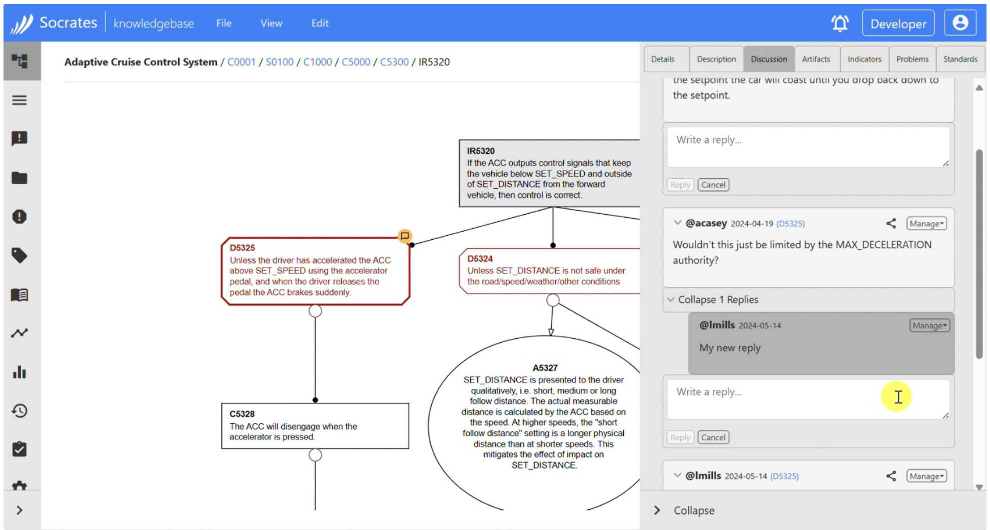
@lmills 2024-05-14 (D5325)

My new comment

Edited: 2024-05-14

Figure 7: Issue Tracking Feature of Socrates

- **Commenting:** Users can add and respond to comments within Socrates, allowing users a documented and organized method to discuss specific parts of the assurance argument back and forth.



Adaptive Cruise Control System / C0001 / S0100 / C1000 / C5000 / C5300 / IR5320

IR5320
If the ACC outputs control signals that keep the vehicle below SET_SPEED and outside of SET_DISTANCE from the forward vehicle, then control is correct.

D5325
Unless the driver has accelerated the ACC above SET_SPEED using the accelerator pedal, and when the driver releases the pedal the ACC brakes suddenly.

D5324
Unless SET_DISTANCE is not safe under the road/speed/weather/other conditions

C5328
The ACC will disengage when the accelerator is pressed.

A5327
SET_DISTANCE is presented to the driver qualitatively, i.e. short, medium or long follow distance. The actual measurable distance is calculated by the ACC based on the speed. At higher speeds, the "short follow distance" setting is a longer physical distance than at shorter speeds. This mitigates the effect of impact on SET_DISTANCE.

the setpoint the car will coast until you drop back down to the setpoint.

Write a reply...

@acasey 2024-04-19 (D5325)
Wouldn't this just be limited by the MAX_DECELERATION authority?

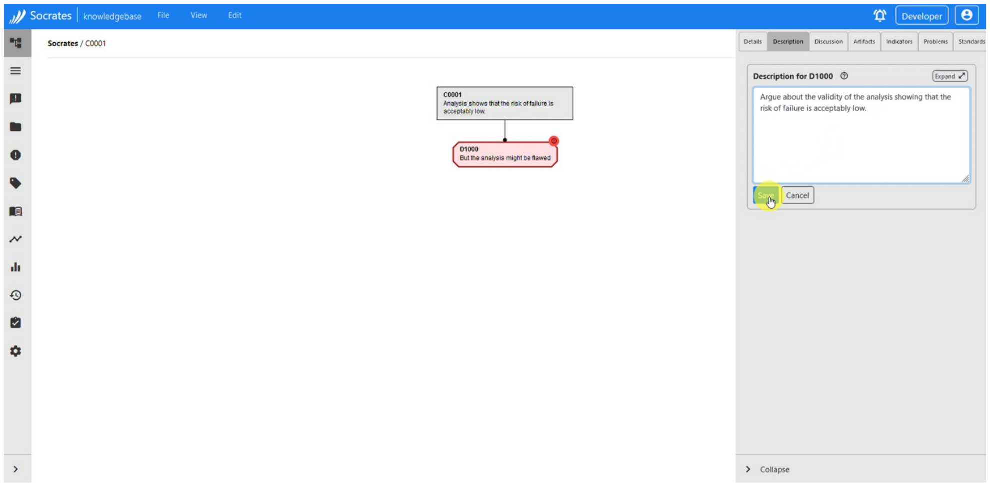
@lmills 2024-05-14
My new reply

Write a reply...

@lmills 2024-05-14 (D5325)

Figure 8: Commenting Feature of Socrates

- **Narrative summary/Description:** Users can add narrative summaries to nodes and subtrees, allowing other users to better understand the content within a subtree of the assurance argument at a high-level.



Socrates / C0001

C0001
Analysis shows that the risk of failure is acceptably low.

D1000
But the analysis might be flawed

Description for D1000

Argue about the validity of the analysis showing that the risk of failure is acceptably low.

Figure 9: Narrative Summary/Description Feature of Socrates

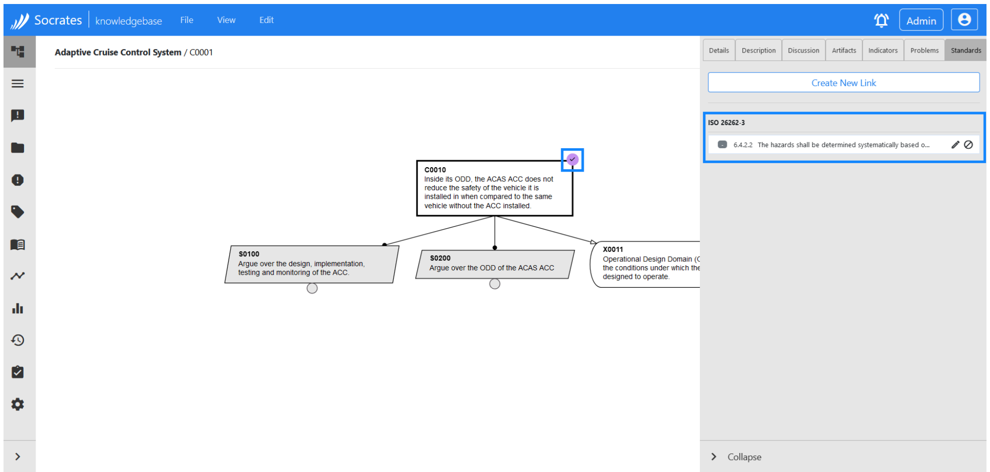
Traceability and Evidence Management:

Within an assurance argument, traceability of requirements and standards to specific portions of an assurance argument increase the argument's usability and effectiveness. Traceability between requirements and corresponding nodes within the argument allow users to quickly see how specific safety requirements are met, hazards are mitigated, and what residual risks might be present. Similarly, traceability from standards (e.g., ISO/PAS 8800) to specific nodes of the assurance argument show how various requirements of the standard are achieved within the assurance argument quickly and easily.

Clause 8.3.2 of ISO/PAS 8800 states that “*The assurance argument shall use the relevant work products generated during the AI safety lifecycle to support the assurance claims*.” Clause 8.5.2 details categories of evidence that can be linked to the safety argument claims.

As these categories of evidence are cross-functional, each stakeholder can use Socrates to provide evidence for parts of the claim that pertains to their specific role, allowing for Socrates to act as a centralized repository of evidence and documentation. Work products that produce evidence as their output can be linked from a live assurance argument that allows traceability and management of the products.

Socrates provides a “standards mapping” functionality to cross-reference any standard (imported via a .csv file) to nodes within the assurance argument. This feature can also be used to map requirements, or any other list-based information. Once complete, Socrates can also export lists of standards/requirements/etc. mappings into Excel .csv files for analysis outside of Socrates, if required.



Adaptive Cruise Control System / C0001

C0010
Inside its ODD, the ACAS ACC does not reduce the safety of the vehicle it is installed in when compared to the same vehicle without the ACC installed.

S0100
Argue over the design, implementation, testing and monitoring of the ACC.

S0200
Argue over the ODD of the ACAS ACC

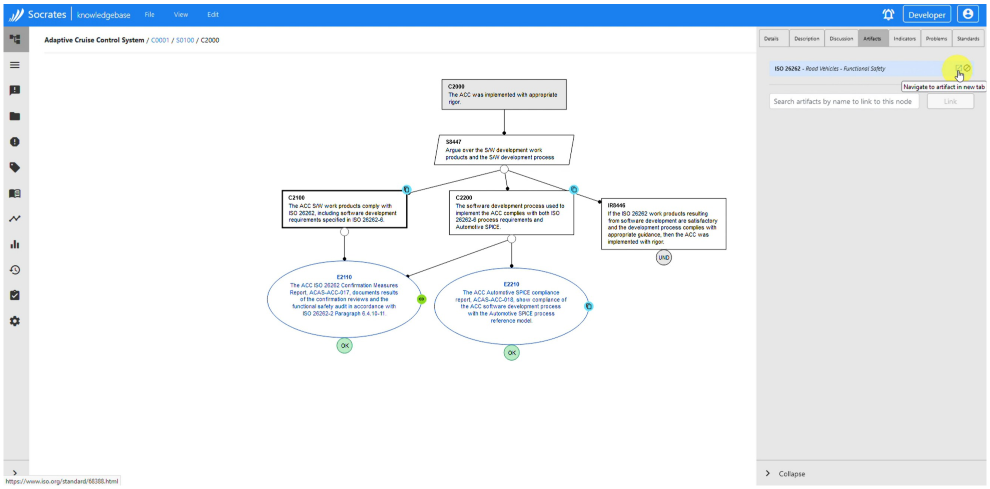
X0011
Operational Design Domain (C
the conditions under which the
designed to operate.

ISO 26262-3

6.4.2.2 The hazards shall be determined systematically based o...

Figure 10: Standards Mapping Feature of Socrates

Additionally, documents or other information can be linked for quick access from a node via hyperlinks; these hyperlinks can be used to reference online documents or documents on client servers for 1-click access. Socrates does not store documents within itself, and instead uses normal hyperlinks to provide access to documents. These hyperlinks will have the permissions of the user’s computer (e.g., a user would not be granted additional permissions such that they could view documents securely stored on an organization’s SharePoint that they would not normally be able to view).


Socrates
knowledgebase
File
View
Edit
Developer
Adaptive Cruise Control System / C0001 / S0100 / C2000
Details
Description
Discussion
Artifacts
Indicators
Problems
Standards
ISO 26262 - Road Vehicles - Functional Safety
Navigate to artifact in new tab
Search artifacts by name to link to this node
Link
UND
OK
OK
Collapse
https://www.iso.org/standard/68388.html

C2000
The ACC was implemented with appropriate rigor.

S8447
Argue over the S/W development work products and the S/W development process

C2100
The ACC S/W work products comply with ISO 26262, including software development requirements specified in ISO 26262-6.

C2200
The software development process used to implement the ACC complies with both ISO 26262-6 process requirements and Automotive SPICE.

IR8446
If the ISO 26262 work products resulting from software development are satisfactory and the development process complies with appropriate guidance, then the ACC was implemented with rigor.

E2110
The ACC ISO 26262 Confirmation Measures Report, ACAS-ACC-017, documents results of the confirmation reviews and the functional safety audit in accordance with ISO 26262-2 Paragraph 6.4.10-11.

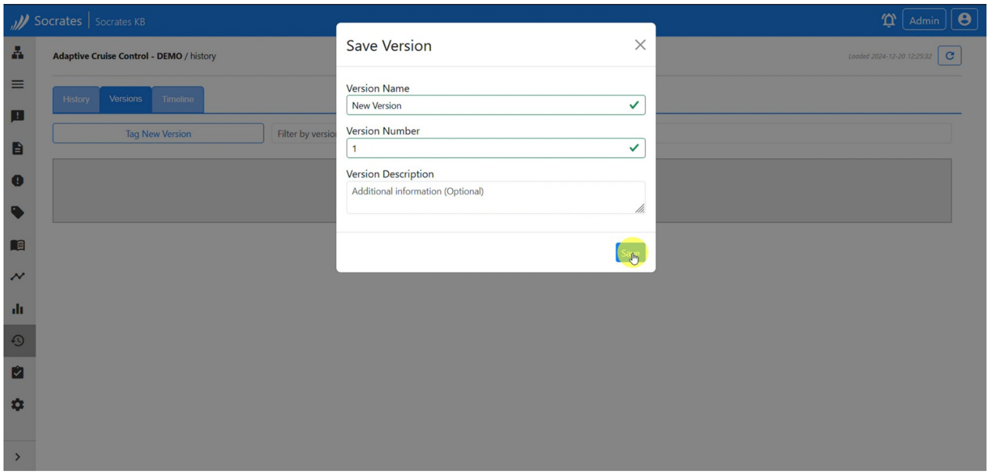
E2210
The ACC Automotive SPICE compliance report, ACAS-ACC-018, show compliance of the ACC software development process with the Automotive SPICE process reference model.

Figure 11: Hyperlinking Feature of Socrates

Change Management:
ISO/PAS 8800 addresses change management primarily by emphasizing that any modifications to the AI system, its datasets, or its operating context require a “re-evaluation” of safety evidence and “re-confirmation” of safety claims.

Clause 7 of ISO/PAS 8800 discusses the iterative process of the AI safety lifecycle and how whenever design or implementation changes occur, the AI assurance case must be revisited and updated based on new findings. For example, this includes retraining a model or introducing new capabilities (new sensors, new smart algorithms, etc.) or due to changes to claims regarding the relevant work products (safety analysis, datasets, V&V reports, etc.). This is supported by a note under clause 8.3.2 which states “Changes to the work products and their impact on the assurance argument are considered as part of change management throughout the AI safety life cycle.”

In general, the iterative nature of the AI safety process requires careful change management so that a history of design, development and post deployment changes can be captured. This must be done in an organized manner to support the organization’s safety claims about their AI system and to provide auditors with a clear history of these changes and their impact on the safety of the system. Socrates allows for both versioning and history/rollback of assurance arguments, which form part of this change management process.



Adaptive Cruise Control - DEMO / history

Loaded 2024-12-20 12:25:32

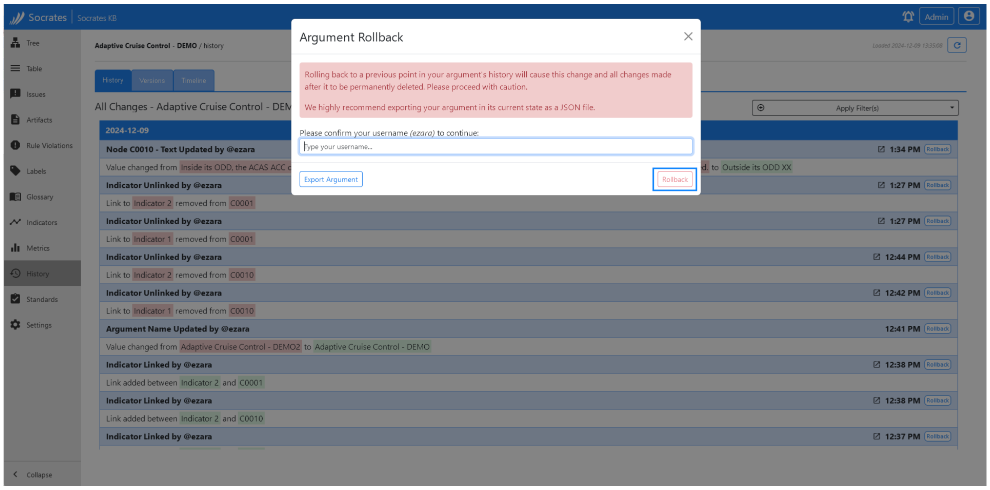
Save Version

Version Name: New Version

Version Number: 1

Version Description: Additional information (Optional)

Figure 12: Versioning Feature of Socrates



Adaptive Cruise Control - DEMO / history

Loaded 2024-12-09 13:35:08

All Changes - Adaptive Cruise Control - DEM

Argument Rollback

Rolling back to a previous point in your argument's history will cause this change and all changes made after it to be permanently deleted. Please proceed with caution.

We highly recommend exporting your argument in its current state as a JSON file.

Please confirm your username *(ezara)* to continue:

Type your username...

2024-12-09

Node C0010 - Text Updated by @ezara — 1:34 PM Rollback

Value changed from Inside its ODD, the ACAS ACC c … d. to Outside its ODD XX

Indicator Unlinked by @ezara — 1:27 PM Rollback

Link to Indicator 2 removed from C0001

Indicator Unlinked by @ezara — 1:27 PM Rollback

Link to Indicator 1 removed from C0001

Indicator Unlinked by @ezara — 12:44 PM Rollback

Link to Indicator 2 removed from C0010

Indicator Unlinked by @ezara — 12:42 PM Rollback

Link to Indicator 1 removed from C0010

Argument Name Updated by @ezara — 12:41 PM Rollback

Value changed from Adaptive Cruise Control - DEMO2 to Adaptive Cruise Control - DEMO

Indicator Linked by @ezara — 12:38 PM Rollback

Link added between Indicator 2 and C0001

Indicator Linked by @ezara — 12:38 PM Rollback

Link added between Indicator 2 and C0010

Indicator Linked by @ezara — 12:37 PM Rollback

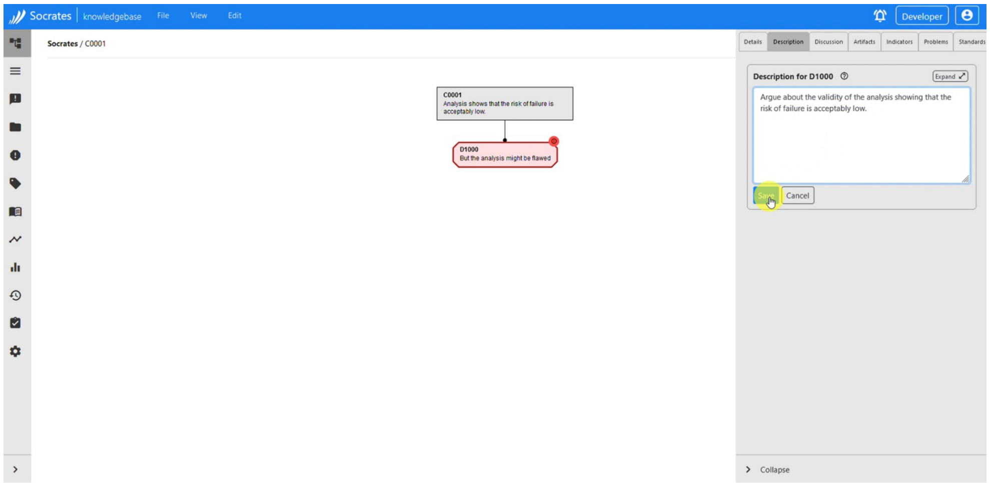
Figure 13: History/Rollback Feature of Socrates

Reporting Features:

ISO/PAS 8800 details that organizations need to maintain a structured body of evidence to demonstrate how AI safety requirements are met and how residual risks are managed over time.

The assurance argument is a central narrative for this reporting, providing a set of safety claims supported by evidence gathered throughout the development and post deployment lifecycle. ISO/PAS 8800 calls for confirmation measures to validate the safety claims of the argument; the outcome of these measures feed into the body of evidence supporting the claims of the assurance argument.

Within Socrates, high-level summaries can be created for various teams as well as for organizational leadership to evaluate the quality of the claims and evidence that make up the argument. For example, see Error: Reference source not found below.



C0001
Analysis shows that the risk of failure is acceptably low.

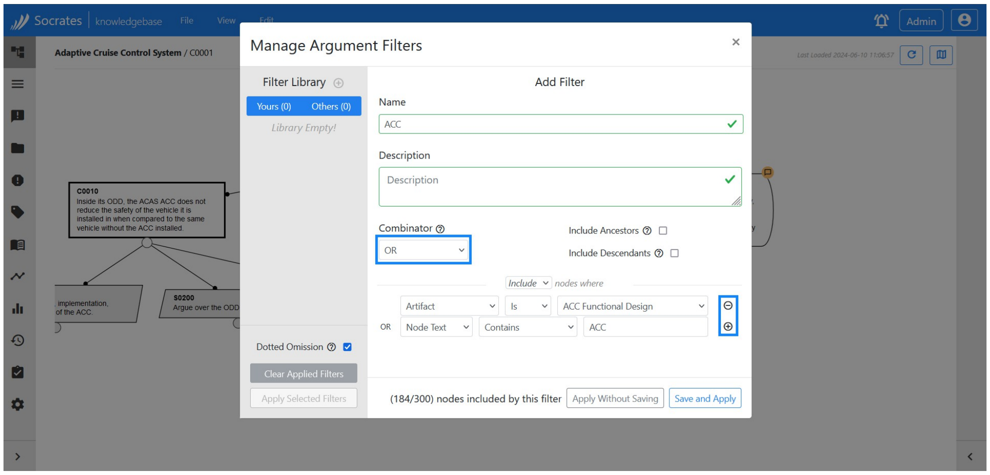
D1000
But the analysis might be flawed

Description for D1000

Argue about the validity of the analysis showing that the risk of failure is acceptably low.

Figure 14: High-Level Summary/Description Feature of Socrates

Additionally, filtering features can be used to focus the scope of reports and the display for the various stakeholders, e.g., teams that work on a specific functionality or cross functional teams, as well as for external auditors. The filtering menu is shown in Error: Reference source not found below.



Adaptive Cruise Control System / C0001

Last Loaded 2024-06-10 11:06:57

C0010
Inside its ODD, the ACAS ACC does not reduce the safety of the vehicle it is installed in when compared to the same vehicle without the ACC installed.

implementation, of the ACC.

S0200
Argue over the ODD

Name: ACC

Description: Description

Combinator: OR

Include Ancestors ☐

Include Descendants ☐

Include nodes where

Artifact | Is | ACC Functional Design

OR Node Text | Contains | ACC

Dotted Omission ☑

(184/300) nodes included by this filter

Figure 15: Filtering Feature of Socrates

Conclusion

Overall, ISO/PAS 8800 introduces guidance for creating an AI assurance case argument, from the preliminary stages of development through to updating and maintaining the AI assurance case. An assurance case, started early in the development lifecycle can inform the design in later phases of development. The use of Key Performance Indicators (KPIs), including safety-related KPIs, allows for monitoring of AI systems to ensure adequate performance throughout its lifecycle (e.g., due to the operational design domain/environment changing).

An ideal safety case benefits all stakeholders, including design teams, management, investors, and auditors. This includes quick and easy viewing of traceability between various aspects of the assurance argument and other documents, including requirements, relevant evidence, review status, etc.

Change management of an assurance case, including history, rollback, and heatmaps to indicate changes are effective tools to better understand the development of the assurance case.

Additionally, reporting features, such as Microsoft Excel and Word exports are useful tools to disseminate information to those who may not have access to Socrates' browser-based access and to limit/review/approve what information is visible, e.g., for third-parties.

Socrates is a tool specifically developed by Critical Systems Labs (CSL) to improve the way we provide critical-system safety consulting to our clients. It supports AI safety cases as described in ISO/PAS 8800, in addition to those described in other standards such as ISO 26262 and ISO 21448 in a live collaborative manner, allowing for simultaneous editing of the assurance cases by various parties. ISO/PAS 8800's guidance, in combination with Socrates, is a powerful method to develop and maintain AI/ML assurance cases; this combination improves safety and the organization/dissemination of relevant safety assurance case information across many industries for both AI/ML or conventional systems.

References

1. https://www.iso.org/standard/83303.html – ISO/PAS 8800:2024 "Road vehicles — Safety and artificial intelligence"
2. https://www.iso.org/standard/68383.html – ISO 26262:2018 "Road vehicles — Functional safety"
3. https://criticalsystemslabs.com/socrates – "Socrates Assurance Cases"
4. https://www.criticalsystemslabs.com/resources-hub/2022ISSCDiemert/2022ISSCDiemert.pdf – "Incremental Assurance Through Eliminative Argumentation" – 2023 – S. Diemert, J. Goodenough, J. Joyce, C. B. Weinstock
5. https://criticalsystemslabs.com/wp-content/uploads/2024/09/Driving-Development-from-the-Safety-Case-merged.pdf – "Driving the Development Process from the Safety Case" – 2024 – C. Hobbs, S. Diemert, J. Joyce