ADELARD
part of nccgroup

CAE MINI-GUIDE 2: CAE CONCEPTS

Summary

This document is part of the Declare CAE guidance document set. It contains guidance on the main CAE concepts of Claims, Arguments, Evidence, and Defeaters.

65 Curzon Street
London
W1J 8PE

T +44 161 209 5200
E office@adelard.com
W www.adelard.com

Authors
Robin Bloomfield
Nick Chozos
Panayiota Sketsiou

Copyright © 2024
NCC Group plc.
All rights reserved.

Contents

1 Introduction 3
2 This mini-guide and the CAE document set 3
3 CAE concepts 4
3.1 Claims 4
3.1.1 The basic concept 4
3.1.2 Guidance on formulating a claim 5
3.1.3 Testing the claim formulation 8
3.1.4 Summary 8
3.2 Evidence 9
3.2.1 The basic concept 9
3.2.2 Guidance on identifying evidence 9
3.2.3 Summary 10
3.3 Arguments 11
3.3.1 The basic concept 11
3.3.2 Summary 13
3.4 Defeaters 13
3.4.1 Basic concepts 13
3.4.2 Identification of defeaters 14
3.4.3 Applying and addressing defeaters 15
3.4.4 Summary 17
4 Summary guidance 17
5 Acknowledgements 17
6 Bibliography 17
6.1 CAE main-guide 17

6.2 CAE mini-guides 17

Figures

Figure 1: Summary of the CAE process and supporting mini-guides 3
Figure 2: Example of top-down argument development (incomplete) 11
Figure 3: Example of top-down argument development (complete) 12
Figure 4: Representation of defeaters in a safety case 14
Figure 5: Exact/eliminative defeater example 15
Figure 6: Addressed defeater 16

Tables

Table 1: Relationship of this mini-guide to the CAE process 4
Table 2: Dialogue about claims 5
Table 3: Guidance on formulating a claim 7
Table 4: Examples of dependability properties 7
Table 5: Examples of other attributes 7
Table 6: Example of putting objects and properties into claims 8
Table 7: Guidance questions and commentary 10

CONFIDENTIALITY, INTELLECTUAL PROPERTY RIGHTS, AND DISCLAIMER STATEMENT
The information contained in this Report has been produced on behalf of EDF Energy Nuclear Generation Limited, Nuclear Decommissioning Authority (Sellafield Ltd.), AWE plc, Urenco UK Ltd., Devonport Royal Dockyard Limited, Rolls-Royce Submarines Ltd. and Rolls-Royce SMR ("the Parties"). This Report is the property of EDF Energy Nuclear Generation Limited ("the Lead Party") who hereby grants each of the other Parties and their successor companies, an irrevocable royalty-free, non-exclusive licence to EDF Energy Nuclear Generation Limited's rights to the Intellectual Property generated in the Report. This is in accordance with Clause 10 of the CINIF Research Agreement. This information is to be held strictly in confidence within the Parties and duly authorised recipient organisations including the Office for Nuclear Regulation, the Health and Safety Executive, Government Departments, or non-Parties with a support contract to assess a Party's safety case. No disclosure is to be made to any other third party without the written agreement of the Lead Party and is to be used solely for the purposes sanctioned by the Parties.

DISCLAIMER
The views expressed in this Report are those of the author(s) and do not necessarily represent the views of the members of the Parties. The Parties do not accept liability for any damage or loss incurred as a result of the information contained in this Report.

1 Introduction

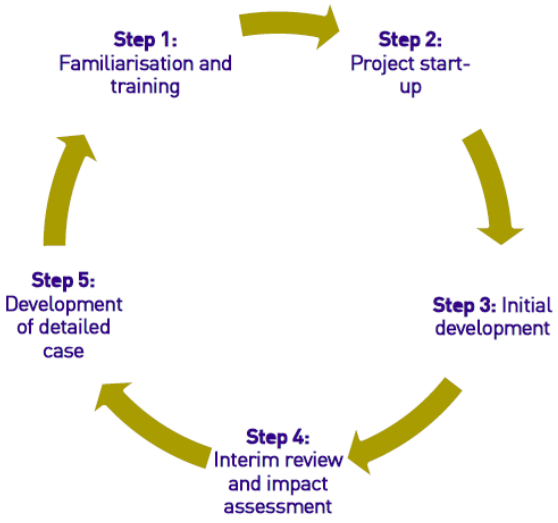
This document is part of the Declare CAE guidance document set. It contains guidance on the main CAE concepts of Claim, Argument, Evidence.

2 This mini-guide and the CAE document set

The CAE guidance can be seen as having two main components:

1. **CAE process:** The first component describes an overall process made up of five steps (the "CAE process"), explaining the evolution of a justification within an organisation and the activities involved. The approach is flexible, adaptable, and will apply differently to different scenarios of use.
2. **CAE mini-guides:** The second part provides specific technical guidance on the underlying concepts, their definition and their application. We have compartmentalised the technical guidance into "mini-guides": small, dedicated sets of guidance each focusing on a particular issue. Each mini-guide contains a concise summary with a short list of the key points and risks and challenges that need to be considered, which is then supported by more detailed guidance.

The CAE process, and the supporting mini-guides, are summarised in Figure 1 below. This document is highlighted (mini-guide 2). A list of the full guidance document set can be found in Section 6.2.



	Mini-guide
1.	One-page guidance
2.	**CAE concepts**
3.	Formulation of top-level claim
4.	Connection rules
5.	The CAE building blocks
6.	Summarising and communication
7.	Review and challenge
8.	Defeaters

Figure 1: Summary of the CAE process and supporting mini-guides

The overall CAE process is described in the main CAE guidance document [4]. The main guide explains the various scenarios of use and how the guidance may apply in different cases. The document also discusses how the mini-guides may be used in different scenarios and at different phases of a project.

Table 1 below illustrates how this mini-guide (CAE concepts) applies throughout the CAE process.

CAE steps	CAE concepts
Step 1: Familiarisation and training	Review and preparation.

CAE steps	CAE concepts
Step 2: Project start-up	Top-level claim definition. Focus on capturing context and assumptions for top-level claim.
Step 3: Overview and initial development	Application during development of outline case – refinement is expected in next steps.
Step 4: Interim review and safety case impact assessment	Full application on all nodes for purposes of review Review upwards as evidence is being linked to claims
Step 5: Development of a more detailed case	Iteration of Step 4 until all evidence has been linked to case Guidance may be used for review and challenge

Table 1: Relationship of this mini-guide to the CAE process

3 CAE concepts

In this section we provide guidance on the individual concepts of claims, arguments and evidence. Using the CAE concepts correctly is part of the basic engineering SQEPness that is required.

3.1 Claims

3.1.1 The basic concept

A **claim** is a true/false statement about a property of a particular object.

A claim is exactly what you might consider it to be from common usage of the term; an idea that someone is trying to convince somebody else is true. We are concerned with claims about an engineered system[1] that are in principle demonstrable or falsifiable: either true or false.

In the example of a crane, Clare might claim, "The crane is safe". As the dialogue in Table 2 shows, this concise and simple claim may be interpreted in a number of different ways depending on the audience.

Claim	How the claim is interpreted
Clare says: *"The crane is safe"*	Angus hears: *"The nuclear crane in building A is safe and risks are 10^{-3} pa"*
	Edward hears: *"The tower crane near building A is safe for unloading a flask"*
	Alice hears: *"The nuclear crane in building A is safe for this particular operation because the risks of not doing what we plan are so high it is worth accepting higher crane-related risks"*

[1] Not, for example, with claims that all Martians are green.

Table 2: Dialogue about claims

Here, only Alice interprets correctly the full intent of the claim. Perhaps because they have been involved in the project, only Alice and Clare both know the project context so it does not need spelling out in this dialogue.

The challenge is that a balance must be struck between making a claim that is so precise, detailed and caveated by assumptions it is incommunicable, and so short and memorable that it is easily misheard or misinterpreted out of the specific context. In practice, we may consider different but consistent versions of the case for different stakeholders.

Indeed, in communicating the claim, Clare will refine her ideas and the claim may change. After some trial and error and peer review, we may end up with a claim that is indeed a claim (properly defined in CAE terms) and which we wish to demonstrate or challenge.

There are some claims that may be self-evident but normally the point of formulating a claim is so that we can debate and argue about it with as little potential for misunderstanding as possible, rather than to state the obvious.

It may be that we find evidence that directly supports or refutes the claim (we discuss arguments later and the definition of evidence) or it may be that the claim is not readily demonstrable. This may be because we have been too vague or too general in the claim we are making. We may therefore find that we wish to make the claim more precise in terms of the property, the assumptions, or the claims that are being made. We call making a claim more precise and less abstract, *concretion*.

Alternatively, we may find that the claim is too complex to readily demonstrate so that we need a "divide and conquer approach" in which we expand the claim into constituent subclaims – we call this *decomposition* (see mini-guide on the CAE building blocks [9] for more on decomposition). For example, we may have a general property such as dependability and split this into relevant constituent properties of reliability and maintainability. Similarly, we might decompose the system architecture into sub-components (e.g. input, processing, output) leading to a claim expansion. We could also expand the claim to deal with different environments or periods of time. This is the topic addressed by the CAE blocks (see mini-guide [9]).

3.1.2 Guidance on formulating a claim

The table below contains guidance on claim formulation, by considering some of the questions that the authors themselves must address with CAE.

Question	Guidance
Who will be interpreting the claim and what is the CAE for?	In initiating the case (see Step 2: project start-up in Table 1) the purpose of the CAE and the target audience will be determined. This will help shape the scope and the high-level claim of the case. For instance, the purpose of the CAE could be "An internal working summary between co-workers with limited circulation and lifetime" or a "Major project that will have 20 stakeholders in many different organisations, in different business and safety cultures and last 30 years".

Question	Guidance
Is the claim a statement that can be true or false?	As discussed in other parts of the guidance, statements like "the test report" are not claims (there is no property to be true or false and those such as "all Martians like blue carrots" are hard to demonstrate although philosophers might try).
What system or object does it refer to?	The object of the claim can be a component, a system, an organisation or an activity (e.g. transportation of nuclear fuel), in fact anything that is real and can have a property. Consider whether we need to be more precise about the state or operating mode of the object
What property does it address?	The properties that we wish to make a claim about are often dependability related. Table 4 lists some of the types of dependability properties that we may wish to make claims about and Table 5 lists other attributes that might be of interest for discussion.
Over what time period is the claim being made?	The validity of claims (as well as arguments and evidence) can be challenged outside their scope and context. If the lifetime of a system is ten years, and the safety case is not bound by the same timeframe, then after ten years these claims may not hold true (e.g. due to system aging).
Is the environment of the system clear? Does it need to be more/less explicit?	We may consider a claim "the reliability of widget X is adequate". However, widget X does not live in isolation so the property (reliability in this example) is only valid in a particular context and for a period of time. So we have to clarify the claim "the reliability of widget X is adequate in a particular environment over a period of time".
Does the context/environment need further explanation?	Is the operating mode of the system (e.g. the crane) sufficiently defined? Is the phase of the project clear? Is the state of the rest of the plant (e.g. normal or fault conditions) relevant and detailed sufficiently?
Are there any common terms that might be "overloaded" or used to make the definition more precise	Every single word within a claim can have a considerable impact on the safety case. Terms such as "safe", "all hazards" and "adequate" need to be defined and reviewed in terms of accuracy and completeness, and they may require further justification within the safety case themselves.
Do we need to make "confidence" explicit in the phrasing?	In any real situation we will have doubts about the claims we make; we will have uncertainty about our knowledge of the world. Implicit in a claim is the notion "I am confident that ..." and sometimes we might make this more explicit and even have a measure for our confidence. "Confidence building" is discussed later.

Question	Guidance
Are assumptions sufficiently documented and detailed?	Any claim is likely to be based on assumptions and these may need to be stated explicitly. The longer term the project and the more stakeholders that are involved the greater attention should be given to making assumptions explicit. The judgement over which assumptions to make explicit in a claim can be crucial; a lack of shared assumptions can be the root of many problems (assumptions are the "mother of all accidents"). Yet, if we documented every assumption, we would be swamped by details: that the sun rises tomorrow might generally be irrelevant, but crucial in some contexts such as space exploration.

Table 3: Guidance on formulating a claim

The properties that we wish to make a claim about are often dependability related. Table 4 lists some of the types of dependability properties that we may wish to make claims about and Table 5 lists other attributes that might be of interest for discussion, depending on the system and application.

Reliability	Time response	Accuracy
Availability	Maintainability	Robustness to overload
Security (from external attack)	Usability (by the operator)	Modifiability
Functional correctness	Fail-safety	Safety

Table 4: Examples of dependability properties

Competency	Was completed (successfully)
Effectiveness	Was started (on time)
Compliance	
ALARP	

Table 5: Examples of other attributes

Note that the attributes listed in Table 5 are only examples and further attributes may be relevant. Conversely, for some applications not all attributes need be relevant, e.g. time response would not be safety-relevant for off-line stress analysis programs, but it would be necessary to have accuracy and functional correctness.

We therefore might put these objects and properties together into claims, as outlined in Table 6.

	Object		Property qualifier (if applicable)	Property
I am confident that	System (X) Component (C)	is	Sufficiently Adequately Acceptably	Reliable Secure Available Responds in t sec
	Activity (A)	is	-	Compliant with standard clause x.y Completed
	Organisation (O)	is	Sufficiently Adequately Acceptably	Competent
	Risks	are	-	ALARP

Table 6: Example of putting objects and properties into claims

In many cases, a qualifier is needed to complete the property of the object. For instance, it is not realistic to phrase a claim in the format "All hazards have been identified". This depends on the property and the claim made.

Any CAE claim will need to be bounded by its context. Claims made about a system will only be valid during its intended lifetime. It is important that this is articulated – this is often missed as the authors often assume that this is obvious or known. However, this information should be recorded, and the safety case should show that the environment of use and the lifetime of the system have been considered in the risk management and engineering.

3.1.3 Testing the claim formulation

A test of whether we have formulated a claim properly is whether we can turn it into a proposition; a statement or assertion that expresses a judgement or opinion.

One test for this is to review the claims and see if they can be mapped into the following form:

> Long form: "*I am confident that component **(C)** of system **(X)**, is acceptably **(V)** in environment **(E)** for duration **(T)**, under assumptions **(A)**".*
>
> Short form: ***P (C, X, V, E, A, T)** is true.*

3.1.4 Summary

A claim is a true/false statement about a property of a particular object. It has to include details of the exact object(s) it applies to and the circumstances under which we assert that it is true. These circumstances include details of the environment and context of the object and any other relevant assumptions. The property and the nature of the object can be all manner of things from physical components, abstract functions and organisations.

We may find that the claim is not sufficiently well defined for us to effectively argue about it. It needs *concretion* to make it more detailed and/or more precise. We might find the claim is too complex to reason about on its own, so we can expand it by decomposing some aspect of the claim (e.g. the object, property, environment, time, etc.) into constituent components.

3.2 Evidence

3.2.1 The basic concept

Evidence is an artefact that establishes facts that can be trusted and lead directly to a claim. Evidence serves as the ground and starting-point for safety arguments, from which the validity of claims can be challenged, contextualised and established.

In projects there can be many sources of information but what makes this evidence is the support or rebuttal it gives to a claim. It is therefore useful to identify the claim that is directly supported by the evidence. In order for the case to be convincingly for or against the claim, the evidence must be of sufficient quality, and it must be credible and accurate.

In practice, "Evidence" is sometimes assumed to mean both the supporting report and the claim that is directly supported by the report. We often find that there is quite a gap between the evidence being offered and the claim it supports in the case. While this may be appropriate in summary cases (see mini-guide 6: Summarising and communication [10]) we recommend the following as good practice in developing cases:

- identify the direct claim, the fact, that can be supported by the documentation
- if necessary, develop further claims and arguments to link this to the case

Some guidance on identifying evidence is provided below in Section 3.2.2.

3.2.2 Guidance on identifying evidence

Question	Guidance
What is the nature of the evidence?	This should address the type of artefact: evidence is usually in the form of documents but might be videos, a demonstration, or recordings. Does the evidence exist yet? If not, is it feasible to generate it?
What is the source of the information?	What is the organisation or project that originates the evidence? Sources of evidence are rich and varied. Evidence may include design specifications, definition of the development process and associated artefacts, analysis of prior field experience, measurements and test results, analytical calculations, (e.g. loading, safety margins), software source code analysis, analysis of compliance, documented interviews and ethnographic studies.
Why is the information evidence? What is the direct claim that it supports or rebuts?	This is important for efficiency and focus of the case. What are the key pieces of evidence that can demonstrate or refute the claim?

Question	Guidance
Is the direct claim actually supported by the offered evidence?	There is a need to investigate that this is the case. For example, does the standard actually say that? What does the test actually measure?
Is the evidence a primary source? On what other sources does the evidence depend? Are these available and evaluated as well?	There may be dependencies between evidence and this should be identified. The evidence may not be the primary source, which potentially poses dangers by basing the case on derived reports, e.g. PowerPoint based on PowerPoint based on reports. See, for example, the investigation into Nimrod and the dangers of evidence trustworthiness.
What might be counter evidence?	Has possible counter evidence been identified? Has there been a systematic effort to search for counter evidence? Have evidence sources been neglected for the claim because they are not strong enough to confirm a claim but could be useful in providing contrary evidence? (e.g. analysis of operating experience might provide evidence of failure that would negate a SIL claim but if there were no failures it would only weakly support the SIL claim).
Is the evidence reliable and to be trusted? Is an explicit claim made about the evidence trustworthiness?	The reliability and trustworthiness of evidence should be addressed. Is the evidence authentic, trustworthy, verifiable, and produced by competent organisations? We might have information about the provenance of the evidence (which could help to verify it) and this meta-data can be important in a case. For example, in a safety case we might have a set of successful test cases as evidence, but we could have some doubts that these apply to the actual system. It may be that there is some unforeseen confusion regarding the test cases, the wrong device could have been tested, or even that there has been deliberate and malicious information offered as evidence when it is not. CAE may explicitly be used for this with an argument making a bridge between say, "the report says system passed 25 tests" to "high confidence that the system passed 25 tests". Do the quoted and other third party assessments actually demonstrate what is being claimed?
Is trustworthiness dealt with for different groups of evidence or individually?	The discussion of trustworthiness could either be done for each piece of evidence or for different sources of evidence (as trust might apply to an organisation or process as a whole). The case could become unwieldy if each piece of evidence is addressed separately.

Table 7: Guidance questions and commentary

3.2.3 Summary

Evidence is an artefact that establishes facts that can be trusted and lead directly to a claim.

Evidence needs to be shown to be reliable and relevant. It can come from many sources and these and their provenance should be identified and assessed.

3.3 Arguments

3.3.1 The basic concept

We have previously introduced the concept of a claim that we are seeking to investigate (prove either true or false) and the evidence that we can confidently know (or at least prove) as being true about the world. In the context of CAE, arguments are what links claims and evidence together.

In CAE an **argument** is the way in which we investigate the validity of the claim. It is a rule that provides the bridge between what we know or are assuming (the subclaims, evidence) and the claim we are investigating.

It is worth noting that "argument" sometimes also refers to all the CAE structure between the top-level claim and evidence (as in ISO15026 and ONR talks). Potential confusion can be reduced by distinguishing between argument steps and the overall argument.

Arguments may themselves be valid or fallacious, too weak, or wrongly applied. One of the benefits of CAE is that in making them explicit, the precise evidence can be identified and nugatory work can be avoided.

We illustrate the CAE concept of argument by considering top-down (i.e. from claim to argument) and bottom-up (i.e. from evidence to argument) examples of identifying the argument.

A concrete example of a top-down approach might be:

Claim: [*Crane can safely lift a weight of X tonnes over the full travel of the Crane*] because

- Subclaim: *[Test lift success for 2X tonnes]*
- Subclaim: *[Crane type approved]*

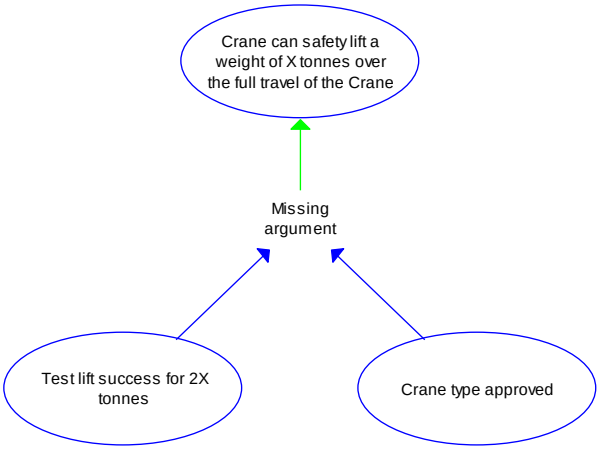


Figure 2: Example of top-down argument development (incomplete)

The argument links the subclaims to the claim

- Argument: *[For a crane with this type of approval, demonstrating a test lift of 2X tons to a limited height before use is sufficient to show Crane can lift X tonnes in operation]*

This can be seen as an application of the rule

If *"Pass test lift of 2X tons to a limited height before use"* then *"Can operate with X tons"*

We can see the completed CAE for this structure in the figure below.

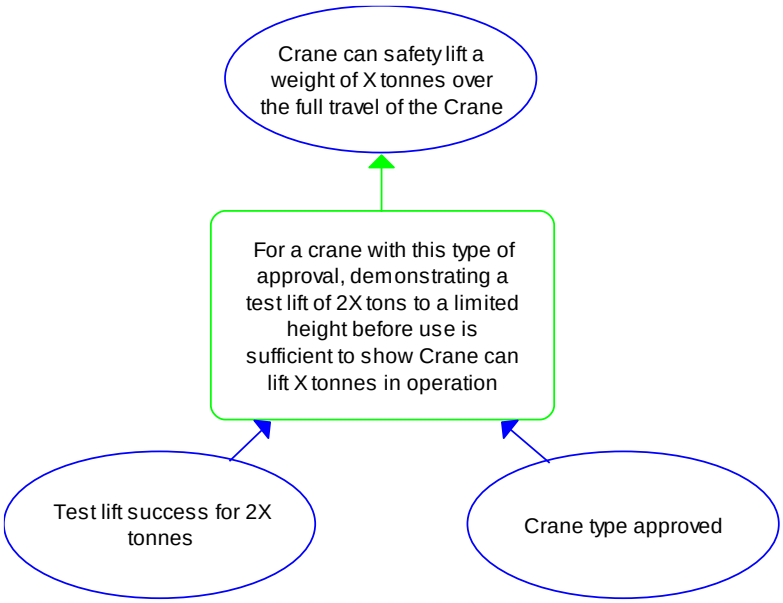


Figure 3: Example of top-down argument development (complete)

In practice, this would need detailing and making consistent: is a "test lift" in the subclaim what is referred to in the argument, what is the evidence for the crane type approval and is there any difference between "full travel" and "in operation". All this shows the type of discussion that focusing on claims and arguments brings out. Another pragmatic issue is to balance the amount of the text shown in the graphical structure and the amount in the supporting narrative. In the graphical structure, we may refer to an argument by its simpler name rather than containing the full argument (as in Figure 3).

Note that when we first ask the question "why can the crane safely lift a weight of X tonnes over the full travel?" the answer to this first step identifies the subclaims that provide "input" to the argument – what we are arguing from. We then need to probe further to get to the argument with a further "why" to identify the reasons why these support the top claim.

A simpler example is given below:

Claim: *[Chris did not commit the murder]*

The evidence offered supports a direct claim of the fact:

Subclaim: *[Chris was not at the crime scene]*

So we might argue that Chris was not in fact the murderer. We might say something like "Chris did not commit the murder because he was not at the crime scene". In CAE terms this leaves the argument implicit. We need to ask the follow up question of "why does this mean he didn't commit the murder" and this might elicit the argument:

One can only commit murder if at the scene or that "if you are not at the scene then you cannot commit a crime"

So again the argument provides a general rule we might use and supports review and challenge. What about action at a distance? Getting someone else to do it? What about delayed poison? And this would lead to a discussion of what "commit" means and to what extent this rule is valid in this particular example.

Alternatively, we might identify an argument by finding the substitution in a bottom-up phrase starting from the evidence, such as

> *If* (this fact, the evidence and these subclaims are true) *then* (this claim is true) *because* (the argument).

For example,

> *If* "there are leaves on the line" *then* "running trains is hazardous" *because* "leaves might prevent a train braking in time, which is hazardous".
>
> *If* "the system passes 46k failure-free tests and assumptions are satisfied (e.g. about representativeness of operational profile)" *then* "the pfd is better than 10^{-4} with 95% confidence" *because* "reliability model XYZ shows this".

In any bottom up argumentation we should avoid traps such as assuming that because we have identified one cause or explanation we have identified them all.

Arguments are only valid if their assumptions are also satisfied. The phrase we might then use to describe the CAE will become slightly more complicated:

> *If* ((this fact, the evidence) *and* these assumptions are true) *then* (this claim is true) *because* (of the argument).

3.3.2 Summary

An argument, in the context of CAE, links evidence, assumptions and subclaims to justify, or to challenge, a claim. The argument used depends on the type, trustworthiness and extent of available evidence and the nature of the claim.

3.4 Defeaters

3.4.1 Basic concepts

A **defeater** potentially refutes or undermines our confidence in a case: it attacks the justification being made by identifying possible problems with the claims, arguments or evidence that make up the case.

Doubts in safety cases, particularly in complex and challenging contexts such as developing a CAE justification for a C&I system, can be reasonably expected to exist. They can be due to emergent changes to the system, lack of justification for the C&I development approaches used, gaps in our scientific knowledge, simple omissions, or even bias. The application of defeaters can increase our confidence in the safety case by either demonstrating that weaknesses are not present, or by improving the CAE, or the system, by addressing weaknesses identified.

A defeater usually attacks a claim in a safety case , and it can be illustrated as shown below.

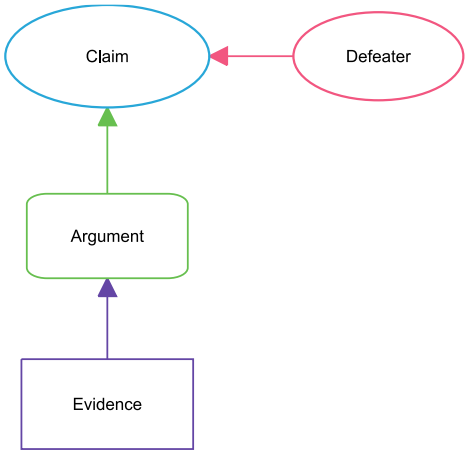


Figure 4: Representation of defeaters in a safety case

There are many types of defeaters proposed in the literature but for safety cases we consider two main types based on Pollock [1]:

- Rebutting. Reasons for believing a directly opposite conclusion, i.e., "claim is false".
- Undercutting. An undermining position, reasons for doubting that the evidence or inference does indeed support the claim e.g., "argument does not apply in this situation".

In practice we might not know at initial identification which of these classes the defeater is in.

In the final assurance case all known defeaters should be addressed, and no new defeaters can be imagined: the case is then said to be indefeasible [2][3]. To achieve this, a systematic search for defeaters needs to applied and documented throughout the development of the assurance case. The identified defeaters need to be defeated themselves through further analysis.

3.4.2 Identification of defeaters

An initial doubt can give rise to a defeater, and it is useful to consider three stages of defeat:

1. initial doubts, when an area of concern is identified
2. exploratory defeater, where a specific issue has been identified
3. addressed or accepted defeater, when a defeater has been addressed

There are several analyses that can be used to identify defeaters, and they can be grouped in three classes:

- hazard analysis and brainstorming
    - brainstorming particular topics and harvesting experience with cases and engineering justifications
    - supported by a form of preliminary hazard analysis using a guideword-based, Hazop type, approach applied to the justification process both generically and for specific cases
- knowledge of standards and relevant guidance

- knowledge in standards and guidelines where lack of compliance can be a source of defeat
- standards and guidance documents that explicitly consider vulnerabilities and things to avoid which can be directly considered as defeaters

- the literature on biases and argument fallacies and whether these are present in the case

3.4.3 Applying and addressing defeaters

Initial doubts and non-formalised defeaters in a case could be addressed by adding claims, arguments, evidence, and extending the narration of the case to account for identified gaps, omissions, etc.

To document the addressing of a formalised defeater in the assurance case, a second defeater addressing the initial defeater is added. This second defeater is followed by new claims, arguments and evidence explaining why and how the initial defeater has been addressed, along with the supporting narrative, which may also be required.

The structure of defeaters within a case can be exploratory or eliminative/exact. An example of an exact defeater in a real case is given in Figure 5, where the defeater negates the claim and proves it false.

In this example, if the field data were collected in a different operating environment, then their analysis is not applicable to the intended operating environment.

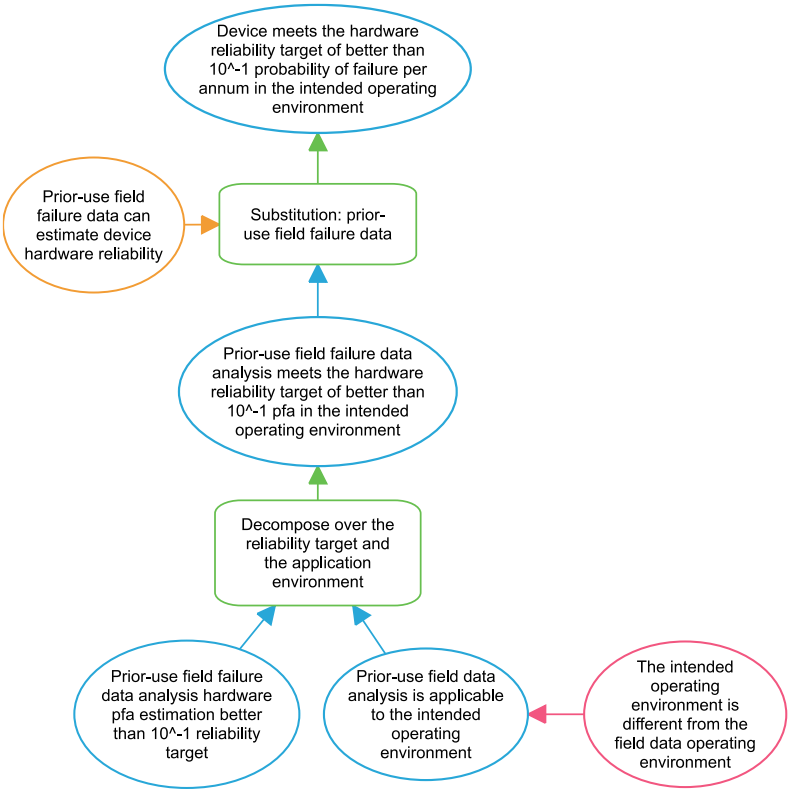


Figure 5: Exact/eliminative defeater example

An example of an addressed defeater is shown below in Figure 6.

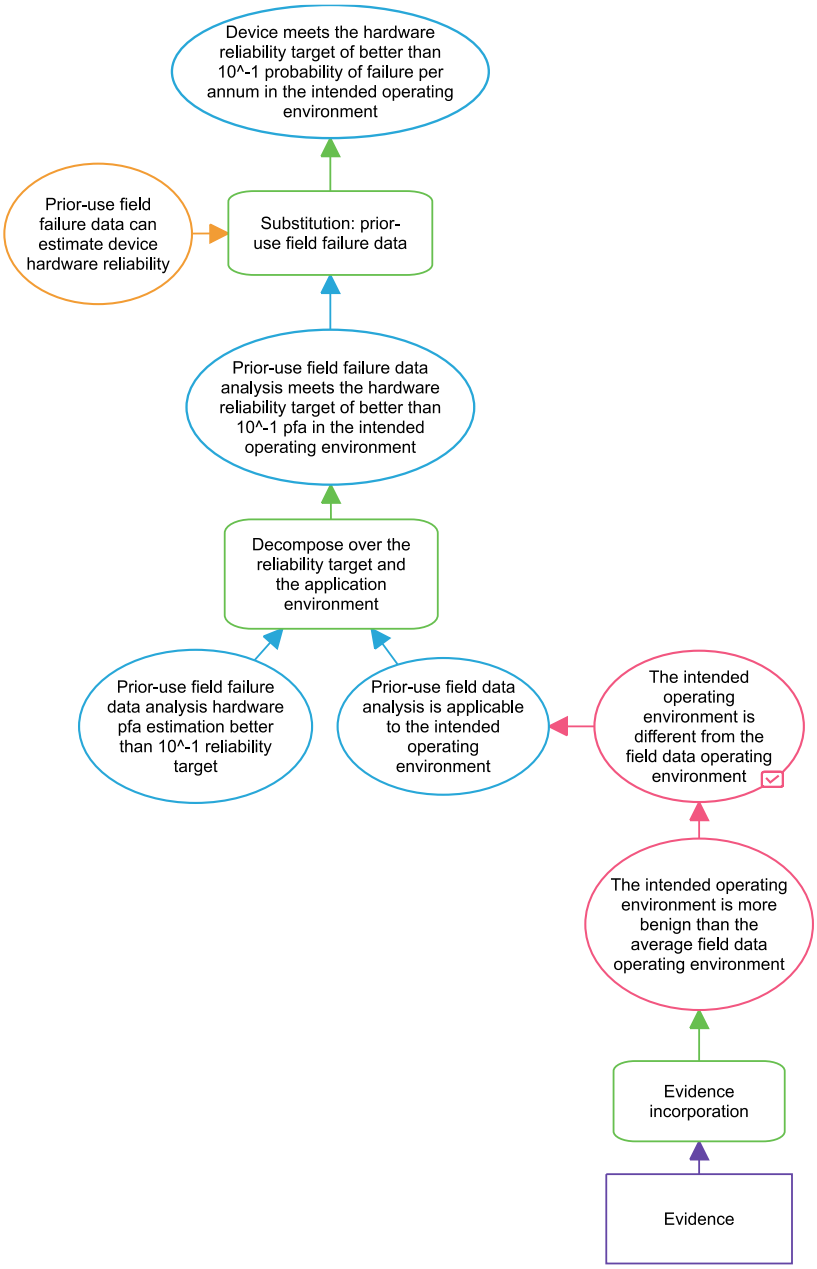


Figure 6: Addressed defeater

The different ways of addressing defeaters and expressing that in CAE form are subject to ongoing research. More details on defeaters can be found in the Defeaters guidance [12].

3.4.4 Summary

A defeater attacks the justification of a safety case by identifying possible problems with its claims, arguments or evidence. A systematic search for defeaters needs to applied and documented throughout the development of the assurance case. The final assurance case can be characterised as indefeasible when all known defeaters are addressed, and no new defeaters can be imagined.

4 Summary guidance

- A claim is a true/false statement about a property of a particular object.
- Evidence is an artefact that establishes facts that can be trusted and lead directly to a claim.
- An argument is the way in which we investigate the validity of the claim. It is a rule that provides the bridge between what we know or are assuming (the subclaims, evidence) and the claim we are investigating.
- An argument links evidence, assumptions and subclaims to justify, or to challenge, a claim. The argument used depends on the type, trustworthiness and extent of available evidence and the nature of the claim.
- Evidence needs to be shown to be reliable and relevant. It can come from many sources, and these and their provenance should be identified and assessed.
- Defeaters attack a case and its justification by identifying possible problems with the claims, arguments or evidence that make up the case.
- The application of defeaters can increase our confidence in the safety case by either demonstrating that weaknesses are not present, or by improving the CAE, or the system, by addressing weaknesses identified.

5 Acknowledgements

We would like to thank Sellafield Ltd and ONR for their high level of engagement with the project, and particularly Sellafield Ltd for their support and involvement in the project workshops.

This deliverable draws on a number of sources developed in earlier Cinif, SSM and Adelard projects.

6 Bibliography

[1] J. Pollock, "Defeasible Reasoning", in J E Adler & L J Rips (Eds.), Reasoning: Studies of Human Inference and its Foundations (pp. 451-470), Cambridge: Cambridge University Press, 2008. R.

[2] R. Bloomfield, J. Rushby. "Assurance 2.0.", 2020.

[3] R. Bloomfield, K. Netkachova, J Rushby (SRI), Defeaters and Eliminative Argumentation In Assurance (to be published Jan 2024).

6.1 CAE main-guide

[4] R. Bloomfield, N. Chozos, Declare: CAE Main Guidance and Process, D/1284/43195/2 v1.0, April 2020.

6.2 CAE mini-guides

[5] R. Bloomfield, N. Chozos, P. Sketsiou, CAE Mini-Guide 1: One-Page Guidance, W/3601/43232/6 v2.0, December 2024.

[6] R. Bloomfield, N. Chozos, P. Sketsiou, CAE Mini-Guide 2: CAE Concepts, W/3602/43232/7 v2.0, December 2024 (this document).

[7] R. Bloomfield, N. Chozos, CAE Mini-Guide 3: Formulation of Top-Level Claim, W/3104/43195/1 v1.0, April 2020.

[8] R. Bloomfield, N. Chozos, P. Sketsiou, CAE Mini-Guide 4: Connection Rules, W/3604/43232/9 v2.0, December 2024.

[9] R. Bloomfield, N. Chozos, P. Sketsiou, CAE Mini-Guide 5: The CAE Building Blocks, W/3605/43232/10 v2.0, December 2024.

[10] R. Bloomfield, N. Chozos, CAE Mini-Guide 6: Summarising and Communication, W/3104/43195/1 v1.0, April 2020.

[11] R. Bloomfield, N. Chozos, P. Sketsiou, CAE Mini-Guide 7: Review and Challenge, W/3607/43232/12 v1.0, December 2024.

[12] R. Bloomfield, E. Karadotcheva, B. Phillips, P. Sketsiou, CAE Mini-Guide 8: Defeaters, W/3608/43232/13 v1.0, December 2024.