National Protective Security Authority

CAE:

CONCEPTS

CONTENTS

FIGURES

TABLES

01. INTRODUCTION

This document presents generic guidance on the individual concepts of Claims, Arguments, and Evidence (CAE). It provides definitions of the three components, discusses how to formulate them, and offers simple examples of their application. It also introduces the concept of defeaters.

02. SIGNPOSTING

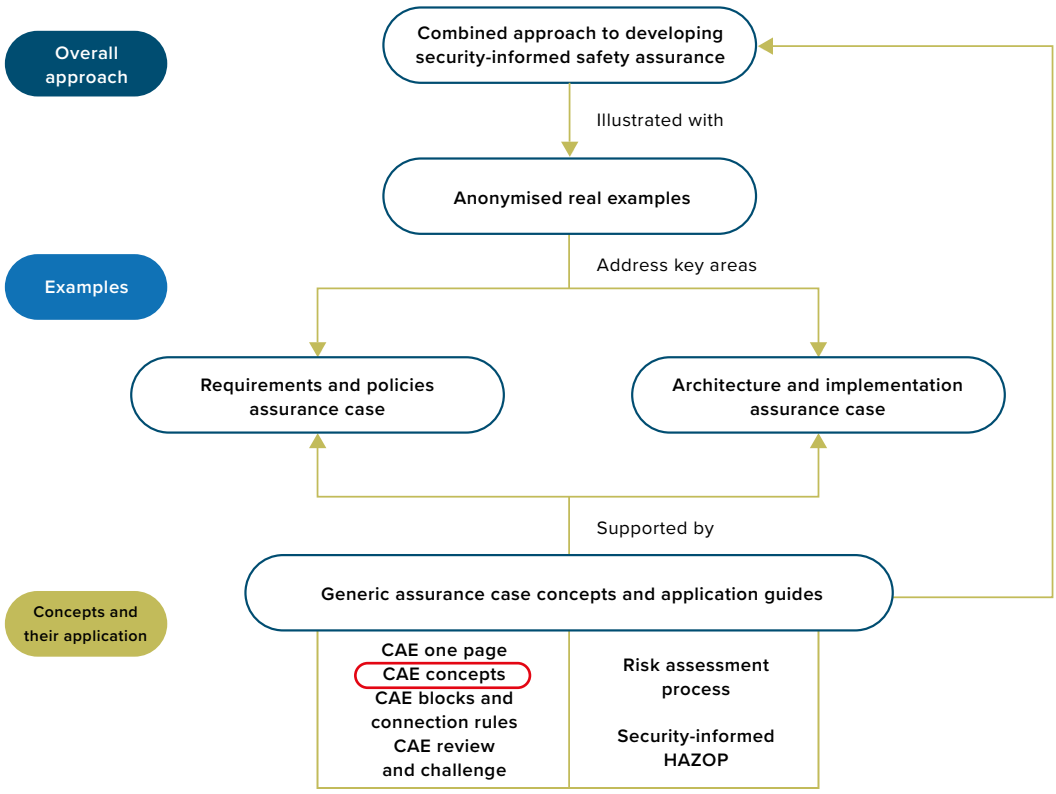

Figure 1: Location of this guide in the set of resources

03. GUIDANCE

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; the idea that someone is trying to convince somebody something else is true. Security-informed safety assurance is concerned with claims about an engineered system that are in principle demonstrable or falsifiable; either true or false. 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.

A simple example of a claim may be: "The x-ray scanner can detect the presence of weapons on attendees of an event with acceptable precision."

As the dialogue in Table 1 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 x-ray scanner can detect the presence of weapons on attendees of the event with acceptable precision"	Angus hears: "The scanner eliminates the threat of dangerous items being brought into the venue."
	Edward hears: "The scanner is suitable for reducing the risk of knife-crime at event X."
	Alice hears: "The scanner, when deployed with trained staff in the correct configuration, is capable of detecting automatic weapons, large bladed weapons, and backpack-sized explosive devices. The precision is sufficient to reduce the risk as far as reasonably practicable."

Table 1: Dialogue about claims

Here, only Alice correctly interprets 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, different but consistent versions of the case for different stakeholders may be required.

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

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

It may be that evidence is found that directly supports or refutes the claim (arguments and the definition of evidence are covered later in this document) or it may be that the claim is not readily demonstrable. This may be because the claim that is being made is too vague or too general. This may result in the claim being made more precise in terms of the property, the assumptions, or the claims that are being made. Making a claim more precise and less abstract is termed concretion.

Alternatively, it may be found that the claim is too complex to readily demonstrate, so a ‘divide and conquer approach’, in which the claim is expanded into constituent subclaims, is required. This is termed decomposition and is discussed in detail in ‘CAE Blocks and Connection Rules’. For example, a general property such as dependability can be split into relevant constituent properties of reliability and maintainability. Similarly, a system architecture might be decomposed into sub-components (e.g. input, processing, output) leading to a claim expansion. A claim can also be expanded to deal with different environments or periods of time.

3.1.2 GUIDANCE ON FORMULATING A CLAIM

The table below contains guidance on claim formulation.

Question	Guidance
Who will be interpreting the claim and what is the CAE for?	In initiating the case, 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’.
Is the claim a statement that can be true or false?	As discussed earlier in this guide, statements like ‘the test report’ are not claims (there is no property that is true or false(, and claims 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 there is a 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 3 lists some of the types of dependability properties that we may wish to make claims about, and Table 4 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 scanner) sufficiently defined? Is phase of 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?	We should always have doubts about the claims we make; we will have uncertainty about our knowledge of the world. Sometimes we might make this more explicit and even have a measure for our confidence. Confidence-building is discussed in 'CAE Review and Challenge'.

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 2: Guidance on formulating a claim

The properties that we wish to make a claim about are often dependability related. Table 3 lists some of the types of dependability properties that we may wish to make claims about and Table 4 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 3: Examples of dependability properties

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

Table 4: Examples of other attributes

Note that the attributes listed in Table 4 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 offline stress analysis programs, but it would be necessary to have accuracy and functional correctness.

These objects and properties might therefore be put together into claims, as outlined in Table 5.

	Object		Property qualified (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 5: Example of putting objects and properties into claims

	Property		Object		Property qualifier (if applicable)
I am confident that	Integrity	of	Component (C)	is	Acceptable
	Compliance	of	Activity (A)	is	Acceptable
	Risks	from	System (X)	are	ALARP

Table 6: Example of putting properties and objects 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. In justifying the claim, thought will be needed to develop a definition of what these qualifiers mean.

Any CAE claim will need to be bounded by its context. Claims made about a system will only be valid during its intended lifetime and in a particular environment. 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 a claim has been formulated properly is whether it can be turned 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 it is asserted 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 to abstract functions and organisations.

It may be found that the claim is not sufficiently well-defined to effectively argue about it. It needs concretion to make it more detailed and/or more precise. It might be found the claim is too complex to reason about on its own, so it can be expanded 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 convincing 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. It is often found 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, it is recommended that the following is used as good practice in developing cases:

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

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.
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?
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 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? There might be 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 there might be a set of successful test cases as evidence, but there could be 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 has 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

This document has already introduced the concept of a claim that is being investigated (to prove it as 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 is investigated. It is a rule that provides the bridge between what we know or are assuming (the subclaims, evidence) and the claim that is being investigated.

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.

The CAE concept of argument is illustrated 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: [The scanner can detect hidden knives with sufficient accuracy] because:

- Subclaim: [The scanner shows 95% precision on a test sample.]
- Subclaim: [The event at which the scanner will be deployed is classed as low-risk.]

This example is shown in Figure 2 below.

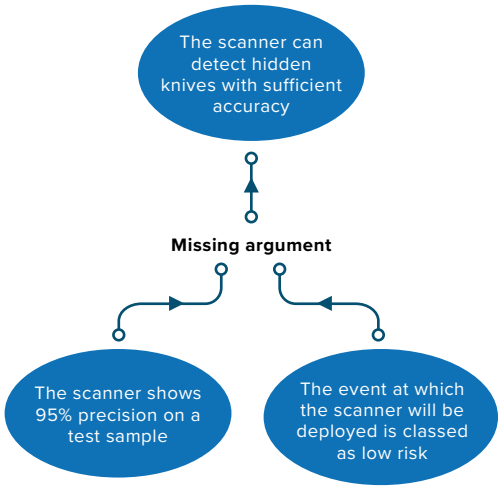

Figure 2: Example of top-down argument development (incomplete)

The argument links the subclaims to the claim:

- Argument: [For a low-risk event, a precision of greater than 90% is sufficient to reduce risk to an acceptable level.]

This can be seen as an application of the rule:

> If 'the scanner shows 95% precision on a test sample' then 'the scanner can detect hidden knives with sufficient accuracy'.

The completed CAE for this structure is shown in Figure 3 below.

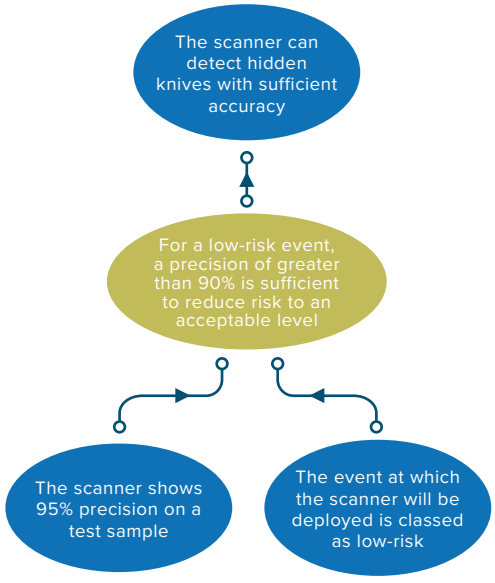

Figure 3: Example of top-down argument development (complete)

In practice, this would need detailing and making consistent: is a "test sample" in the subclaim what is referred to in the argument, what is the evidence for the event being classed as low-risk, and is there any difference between 'test sample' and the population that will be scanned at the event? All this shows the type of discussion that focusing on claims and arguments encourages.

Another pragmatic issue is to balance the amount of text shown in the graphical structure and the amount in the supporting narrative. In the graphical structure, an argument may be referred to by its simpler name rather than containing the full argument (as in Figure 3).

Note that when the question "why is the scanner sufficiently reliable to reduce risk to acceptable levels" is first asked, the answer to this first step identifies the subclaims that provide 'input' to the argument – what is being argued from. It is then necessary 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 it might be argued that Chris was not in fact the murderer. This could lead to something like 'Chris did not commit murder because he was not at the crime scene'. In CAE terms this leaves the argument implicit. It is necessary 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 that might be used 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, an argument might be identified 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, traps such as assuming that because one cause or explanation has been identified they have all been identified should be avoided.

Arguments are only valid if their assumptions are also satisfied. The phrase that might then be used 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

A defeater – as the name suggests – is something that undermines the justification that is being put forward. It is a counter-belief or an objection to the CAE justification.

Any doubt whether a claim is true or whether the evidence adequately supports a claim can be expressed by a defeater. Such doubts can originate from confirmation bias, simple omissions, gaps in knowledge, and errors in reasoning. Identifying and addressing defeaters systematically helps build confidence in the CAE argumentation.

The explicit use of defeaters in assurance cases is less mature than the central CAE concepts, and tools and guidance to support the capture, systematic search and elimination of defeaters are still under development.

04. ACKNOWLEDGEMENTS

This document is based on material developed in earlier projects partially funded by the UK Control and Instrumentation Nuclear Industry Forum (CINIF) and guidance from previous NPSA projects and published research by Adelard.

Disclaimer

This guide has been prepared by NPSA and is intended to support the implementation of security-informed safety assurance and the Claims, Arguments and Evidence (CAE) methodology. This document is provided on an information basis only, and whilst NPSA has used all reasonable care in producing it, NPSA provides no warranty as to its accuracy or completeness.

To the fullest extent permitted by law, NPSA accepts no liability whatsoever for any expense, liability, loss, damage, claim or proceedings incurred or arising as a result of any error or omission in the report or arising from any person acting, refraining from acting, relying upon or otherwise using the [report]. You should make your own judgment with regard to the use of this document and seek independent professional advice on your particular circumstances.

No Endorsement

Reference to any specific commercial product, process or service by trade name, trademark, manufacturer or otherwise, does not constitute or imply its endorsement, recommendation or favour by NPSA. The views and opinions of authors expressed within this document shall not be used for advertising or product endorsement purposes.

National Protective Security Authority