AI and the Common Criteria:

Making Assurance Timely, Accessible, and Continuous

Dr. Ron Ross
CEO, RONROSSECURE

The Common Criteria for IT Security Evaluation (ISO/IEC 15408) has a credibility problem, not because the standard is wrong, but because its economics are brutal. A rigorous CC evaluation at the upper assurance levels can consume eighteen to twenty-four months and hundreds of thousands to millions of dollars. By the time a certificate is issued, the product may have been through two or three releases. The pool of qualified CC evaluators globally is small, specialized, and expensive. And in an era when software is developed and deployed in days and updated in hours, a framework built around multi-year evaluation cycles struggles to stay relevant against the pace at which the threat landscape evolves.

None of this means the Common Criteria approach is wrong. The version of the standard that is currently available at the Common Criteria portal [1][2][3][1] embeds one of the most rigorous and coherent frameworks ever developed for characterizing IT product security. This includes: (1) a structured security problem definition, (2) traceable security objectives, (3) formally specified functional and assurance requirements, and (4) an independent evaluation against those requirements that produces a documented, evidence-based determination of security claims. The logic is sound, but the impediments to date, have been practical. Artificial intelligence is beginning to change that calculus. Not by replacing the rigor of the CC framework, but by attacking the time, cost, and expertise constraints that have limited its reach.

What Made CC Evaluations Expensive in the First Place

To understand where AI can help, it is worth being precise about where the costs accumulate. A CC evaluation is structured around three ascending dimensions of effort: scope, depth, and rigor. At higher Evaluation Assurance Levels (EALs), all three expand. Scope grows to include more of the product's design and implementation. Depth increases to encompass finer levels of architectural detail, internal interfaces, and implementation internals. Rigor demands more structured, formal methods of analysis and documentation.

Each of these dimensions historically required significant human labor. An evaluator at a CC testing laboratory would need to read and analyze extensive Security Target (ST)[2] documentation, trace security functional requirements through design documents to implementation, develop and execute test cases against the Target of Evaluation (TOE), examine development environment controls, and produce formal evaluation reports documenting every finding. For a product of meaningful complexity evaluated at EAL 4 or above, this work could involve significant numbers of qualified evaluators across months of effort.

[1] ISO/IEC 15408:2026 is currently available but is not yet on the Common Criteria Portal [4].

[2] A Security Target (ST) is the document that defines the security problem and solution for a specific product being evaluated, known as the Target of Evaluation (TOE) [1].

Three structural problems compounded the cost.

- **Documentation burden:** Vendors were required to produce extensive, precisely formatted documentation including Security Targets, functional specifications, design documents, and implementation representations, that consumed significant resources to create and even more to maintain across product versions.
- **Scarcity of expertise:** The number of evaluators with the combination of security engineering depth, CC framework knowledge, and domain-specific product expertise needed to conduct a rigorous evaluation is limited globally, making qualified evaluator time genuinely scarce and expensive.
- **Point-in-time rigidity:** Once an evaluation was underway, changes to the evaluated product triggered re-evaluation activities, creating a powerful disincentive to update products during the evaluation window, and resulting in evaluated configurations drifting from the deployed configurations.

The result was a framework with high integrity and low accessibility. Organizations with significant resources and regulatory mandates pursued CC evaluations for specific products in specific markets. Most of the rest of the ecosystem did not.

Where AI Changes the Equation

AI does not solve the core problem of assurance, that is, the confidence in security claims must be grounded in evidence, not assertion. But it can dramatically accelerate and broaden the evidence-generating activities that CC evaluation depends on, across all three of the dimensions that drive cost.

Automated vulnerability and conformance analysis. The most immediately applicable AI contribution is in the *depth* dimension of evaluation. Modern AI systems, including large language models paired with static analysis tools, fuzzing engines, and formal verification frameworks, can examine implementation artifacts at a level of detail and speed that no human team can match. An AI system can analyze a codebase for patterns associated with known vulnerability classes, check that documented interfaces accurately reflect implementation behavior, and generate test cases derived from the functional requirements in the Security Target, with coverage orders of magnitude greater than manual test development.

This directly addresses one of the most labor-intensive aspects of higher-EAL evaluations, that is, the verification that the product does what its design documentation says it does, and nothing more. Evaluators spend enormous level of effort on exactly this kind of assessment activity. AI can handle a substantial portion of it, freeing human evaluators to focus on the interpretive judgments and structural reasoning that remain genuinely difficult for automated systems.

Security Target and documentation generation and review. Preparing a conformant Security Target, one that accurately and completely specifies the security problem definition, security objectives, security functional requirements, and security assurance requirements in CC-compliant format, is a significant undertaking for even experienced vendors. AI can substantially accelerate this process by analyzing existing product documentation, prior Security Targets for similar product types, and generating draft content that evaluators and vendors can review, refine, and verify rather than produce from scratch.

More significantly, AI can continuously check Security Target consistency, verifying that the security objectives completely address the threats and assumptions in the security problem definition, that the security functional requirements are correctly traced from objectives, and that design documentation accurately reflects the claimed security architecture. These consistency checks, currently performed manually by evaluators with significant effort, become a continuous automated activity rather than a periodic milestone.

AI-augmented penetration testing and attack surface analysis. The Common Criteria specifies vulnerability assessment requirements that grow in depth and rigor at higher assurance levels, from basic independent testing at lower EALs to methodical vulnerability analysis and penetration testing at the highest levels. Historically, this work required highly skilled evaluators with both CC framework knowledge and attack expertise.

AI-enabled offensive security tools have matured to the point where they can systematically probe attack surfaces, chain vulnerabilities, and attempt exploitation patterns derived from threat intelligence, at a scale and consistency that human evaluators cannot match. For CC evaluation, this means the vulnerability assessment activities that anchor the highest-EAL evaluations can be conducted with greater coverage and repeatability, and the evidence they produce can be more comprehensive and auditable than manual testing reports alone.

This matters for something more than cost reduction. The strength of an assurance claim is directly proportional to the comprehensiveness and rigor of the evidence behind it. Broader attack surface coverage means more credible assurance, not just faster assurance.

Continuous evaluation and incremental re-evaluation. Perhaps the most structurally significant opportunity AI creates is the shift from point-in-time evaluation to continuous evaluation. Under the current evaluation model, a CC certificate reflects a product at a specific version, evaluated in a specific configuration, at a specific point in time. Product updates require re-evaluation activities that can take months and create strong incentives to delay updates, producing the counterproductive outcome that more mature, security-improved versions of products remain unevaluated while older evaluated configurations persist in deployment.

AI enables a fundamentally different model. If the analysis that underlies CC evaluation, including conformance checking, vulnerability analysis, documentation review, and test execution, can be automated and run continuously, then re-evaluation of product updates becomes a rapid, incremental activity rather than a multi-month project. Changes to the TOE trigger automated analysis of the delta, including what changed, how those changes affect the security architecture, whether the security functional requirements remain satisfied, and whether new vulnerabilities were introduced. Human evaluators review AI-generated delta assessments and make the consequential judgments, rather than rebuilding the entire evidence base from scratch.

This shift from periodic to continuous evaluation is the most significant structural change AI enables for the CC ecosystem. It aligns the evaluation cadence with the development cadence, something the current framework cannot do.

What AI Cannot Replace

The core of the CC framework is not testing; it is structured reasoning. An assurance case is a reasoned, auditable argument that a top-level security claim is satisfied by the available evidence. That argument requires human judgment about the relevance, credibility, and sufficiency of evidence; about the significance of findings; and about whether the security problem has been correctly and completely characterized. AI can generate evidence. It cannot *currently* make the consequential interpretive judgments that determine whether that evidence is sufficient to warrant a trustworthiness claim.

There is also a recursive challenge that the field must confront. If AI tools are used to conduct CC evaluation activities, the trustworthiness of those tools becomes a relevant question. An AI system that generates flawed test cases, misidentifies conformance, or fails to surface relevant vulnerabilities can produce a body of evidence that looks comprehensive while being systematically incomplete. The same assurance discipline the CC applies to IT products must be applied to the AI systems used in evaluation. This is not a reason to avoid AI in CC evaluation, but it is a reason to be deliberate about how AI-generated evidence is reviewed, validated, and incorporated into assurance arguments.

The CC evaluation assurance scale, from scope through depth through rigor, is not just a description of how much work is done. It is a description of how much confidence the resulting evidence can support. AI expands what is achievable at each level. It does not eliminate the need for the levels.

From Component Assurance to System Trustworthiness: The SP 800-160 Connection

The argument for accelerating CC evaluation through AI is not only about making individual products cheaper or faster to certify. It is about what those evaluated products make possible at a fundamentally different level of abstraction: the *trustworthiness* of the systems built from them.

NIST SP 800-160 Vol. 1 Rev. 1, Engineering Trustworthy Secure Systems, establishes a principle that systems engineers must confront directly: building trustworthy secure systems depends on the development of trustworthy secure system components [5]. This is the principle of compositional trustworthiness, and it carries a corollary with significant practical weight: a system can be no more trustworthy than the components from which it is composed, and the trust relationships among them.

This creates a foundational problem for systems engineers using SP 800-160. When they select commercial components (e.g., processors, operating systems, cryptographic modules, network devices) they are making implicit trustworthiness judgments with whatever evidence is available. Too often, that evidence consists of vendor documentation, compliance attestations, and the absence of publicly known vulnerabilities. In the language of SP 800-160, this is axiomatic assurance at best: belief based on assertion, not on evidence-grounded demonstration. The systems engineer integrates the component, inherits its unknown characteristics and behavior, and builds a trustworthiness case at the system level that rests, in part, on foundations the engineer cannot fully see.

CC-evaluated components change this equation. A CC-evaluated product at a meaningful level of assurance carries something that an unevaluated product cannot provide: an independently produced, documented, evidence-based characterization of its security properties, including its security problem definition, the objectives it addresses, the functional security requirements it satisfies, and the

assurance evidence that substantiates those claims. For a systems engineer constructing an SP 800-160 trustworthiness case, that evidence is not merely useful, it is essential.

SP 800-160 requires that assurance judgments be compositional, that is, the trustworthiness of the whole be grounded in the trustworthiness of the parts [5]. A CC certificate and its underlying evaluation evidence provide exactly the component-level assurance artifacts that feed into the system-level assurance case. The security functional requirements satisfied by the component become inputs to the system's security architecture. The vulnerability analysis conducted during a CC evaluation informs the risk at the system level. The assurance evidence produced during evaluation can be incorporated directly into the structured reasoning that SP 800-160's trustworthiness context requires.

The implication for AI-accelerated CC evaluation is significant. Every additional product category that becomes economically feasible to evaluate, including every component type that moves from unevaluated to CC-certified, expands the library of trustworthy building blocks available to systems engineers. If AI can bring evaluation timelines and costs within reach of the broader commercial ecosystem, systems engineers gain access to a fundamentally richer supply of components with documented, independently verified security properties. The trustworthiness argument at the system level becomes correspondingly stronger, because it rests on demonstrated component trustworthiness rather than assumed trustworthiness.

> *We need more trustworthy secure products so that we can build more trustworthy secure systems. That is not a statement about compliance. It is a statement about the foundational dependency between component-level assurance and system-level trustworthiness that SP 800-160 makes explicit and CC evaluation is uniquely positioned to satisfy.*

AI that makes CC evaluation more accessible is not just improving a certification program. It is strengthening the engineering foundation on which trustworthy systems are built.

An Opportunity Worth Taking Seriously

The practical consequence of making CC evaluations timelier and less resource intensive is not marginal. It is transformative for the security of the broader U.S. critical infrastructure and supply chain.

Today, the products most likely to undergo rigorous independent security evaluation are those in regulated markets where evaluation is mandated or strongly incentivized. The products in widest general use — commercial operating systems, development tools, and network infrastructure, are evaluated selectively if at all, and often at assurance levels well below what their criticality warrants. Systems engineers building on these components have no choice but to accept the associated uncertainty into their system-level trustworthiness arguments.

If AI can reduce the cost and time of a rigorous CC evaluation by an order of magnitude, and there is genuine reason to believe it can for many product categories, the addressable scope of independent security evaluation expands dramatically. More products can be evaluated and evaluations can be conducted more frequently. Higher assurance levels become economically viable for a broader range of products and organizations. And systems engineers gain access to a library of components whose security properties are not asserted but demonstrated.

This is the full promise of AI applied to CC evaluation: not just faster certificates, but a more trustworthy ecosystem from the component up — with more building blocks carrying credible, evidence-based security claims, and systems built from them carrying correspondingly stronger trustworthiness cases.

> *The Common Criteria framework provides the evaluation structure. SP 800-160 provides the systems engineering context that gives component-level assurance its full meaning. AI provides the economic engine that can make both frameworks accessible at the scale the problem demands.*

References

[1] Common Criteria for Information Technology Security Evaluation, Part 1: Introduction and General Model, CC:2022 Revision 1, CCMB-2022-11-001, November 2022. https://www.commoncriteriaportal.org/files/ccfiles/CC2022PART1R1.pdf

[2] Common Criteria for Information Technology Security Evaluation, Part 2: Security Functional Components, CC:2022 Revision 1, November 2022. https://www.commoncriteriaportal.org/files/ccfiles/CC2022PART2R1.pdf

[3] Common Criteria for Information Technology Security Evaluation, Part 3: Security Assurance Components, CC:2022 Revision 1, November 2022. https://www.commoncriteriaportal.org/files/ccfiles/CC2022PART3R1.pdf

[4] Common Criteria for Information Technology Security Evaluation Portal. https://www.commoncriteriaportal.org

[5] Ross R, McEvilley M, Winstead M (2022) Engineering Trustworthy Secure Systems. National Institute of Standards and Technology, NIST Special Publication 800-160 Vol. 1 Rev. 1. https://doi.org/10.6028/NIST.SP.800-160v1r1