HOW SHOULD WE REGULATE AI?

Practical Strategies for Regulation and Risk Management from the IEEE 1012 Standard for System, Software, and Hardware Verification and Validation

AUTHOR

Jeanna Matthews
Vice Chair, IEEE-USA AI Policy Committee
Professor of Computer Science, Clarkson University

IEEE USA

INTRODUCTION

Autonomous and intelligent systems (A/IS) are being increasingly deployed for the purpose of making consequential decisions that impact the lives, liberty, and well-being of individuals in areas such as hiring, housing, credit, and criminal justice. There is growing concern about generation of disinformation and toxic content, threats to critical infrastructure, and even existential risks to humanity. Policy makers around the world are debating regulatory approaches to control automated systems, especially in response to growing concern over generative AI technologies like ChatGPT and DALL-E. This article shows how and why we should apply the practical and time-tested strategies for risk management from IEEE Standard 1012 for System, Software, and Hardware Verification and Validation[1] (IEEE 1012) to the current hotly-debated conversation about AI regulation.

For policy makers discussing regulation, it can be difficult to know where to start.

- Is it possible to manage the most severe risks without dampening innovation?
- Are there certain technologies like facial recognition or large language models that should be regulated in any context?
- Are there certain application areas like criminal justice or health care where any A/IS should be regulated?
- Should small businesses face a lower regulatory burden than large businesses?
- What types of penalties or liabilities should governments use to enforce compliance?
- What avenues to redress should be provided to potential victims of A/IS failures?

IEEE 1012 can help to answer these questions and has a lot to offer the current debates about controlling the risks of emerging AI systems. Specifically, it has a long history of practical use in critical environments.[2] It has been used to verify and validate many critical systems including United States Department of Defense weapons systems, nuclear weapons systems power control systems, NASA manned space vehicles, and medical devices. Notably, the principles described by IEEE 1012 apply to any software and hardware system, including new systems based on emerging generative AI technologies that have

1 The most recent version of IEEE 1012 is 1012-2016 that was published on September 29 2017 and is available in its entirety at DOI 10.1109/IEEESTD.2017.8055462 and at persistent URL https://ieeexplore.ieee.org/servlet/opac?punumber=8055460. The full IEEE 1012-206 standard is over 250 pages. With permission from IEEE-SA, this article contains excerpts from Annex B "A risk-based integrity level schema" and Annex C, "Definition of independent verification and validation (IV&V)". This article presents information from these sections in a context designed to be more accessible to policy makers desiring to make use of IEEE 1012 as an example. Portions of IEEE 1012-12016 - Reprinted with permission from IEEE, Copyright IEEE 2016. All Rights Reserved.

2 IEEE 1012 was first introduced in 1998 by the IEEE Standards Association (IEEE-SA), the leading developer of global technical standards used in power and energy, telecommunications, biomedical and healthcare, information technology, transportation, and information assurance products and services. The most recent version of IEEE 1012 is 1012-2016 that was published on September 29 2017 and is available in its entirety at DOI 10.1109/IEEESTD.2017.8055462.

been the topic of so much recent discussion and concern. It is a broadly accepted process for ensuring that the right product is correctly built for its intended use.

Policy makers are understandably concerned with controlling the most serious consequences of AI systems without harming innovation by introducing overly onerous requirements in applications where the risks are lower. Fortunately, there is no need to start from scratch and reinvent the wheel. IEEE 1012 already offers a roadmap for focusing regulation and other risk management actions directly where they will be most impactful.

In current discussions of AI risk management and regulation, many different approaches are actively being considered, some based on specific technologies or application areas, while others consider the size of the company or its user base. Many of these approaches either sweep up low-risk systems into the same category as high-risk systems or leave gaps where regulation would not apply to some high-risk systems.

In contrast, the approach in IEEE 1012 focuses risk management resources directly on the systems with the most risk, regardless of any other factor. It does so by 1) determining risk as a function of both the severity of consequences and their likelihood of occurring and then 2) assigning the most intense levels of risk management activity to the highest risk systems and lower levels of activity to systems with lower risk. In this way, it would, for example, distinguish between a facial recognition system used to unlock your own cell phone, where the worst consequence might be relatively light (e.g. the need to type in a passcode or a false match allowing unauthorized access) and a facial recognition system used to identify suspects in a criminal justice application where the worst consequence might be severe (e.g. improper arrest or use of lethal force).

Applying a similar risk-based approach would offer policy makers an effective and flexible framework for arriving at governance alternatives that are well-calibrated to the specifics of a system and the nature and level of the risks it entails. Specifically, policy makers interested in using IEEE 1012 as an example could use the following six steps, each discussed in this article:

1. **Establish consequence levels from high to low (Section 2)**
2. **Establish likelihood levels from high to low (Section 2)**
3. **Establish integrity levels using a map between consequence levels and likelihood levels (Section 3)**
4. **Assign requirements through the lifecycle of the system to the highest level and then reduce those requirements as appropriate for lower levels (Section 4)**
5. **Require risk management throughout the full system lifecycle (Section 5)**
6. **Require independent review of risk management activities that is appropriate to each integrity level (Section 6)**

IEEE 1012 presents a specific set of activities for the verification and validation of any system, software or hardware. The intensity and depth of these activities prescribed varies based on how the system falls along a range of integrity levels (1-4). Systems at integrity level 1 have the lowest risks and the recommended verification and validation efforts are correspondingly lighter than higher levels. Systems at integrity level 4, on the other hand, may occasionally have catastrophic consequences and therefore warrant substantial verification and validation efforts throughout the lifecycle of the system. Some examples of these verification and validation efforts include documentation of requirements, design, testing, and maintenance activities.

Policy makers could follow IEEE 1012 directly to prescribe verification and validation requirements or they could use the same high-level framework for guiding regulation and other risk management efforts. The primary focus of this article is not to discuss the specific verification and validation efforts recommended by IEEE 1012. Instead, the goal is to give policy makers a set of clear steps they can use to identify integrity levels to which their own desired regulatory approaches can be applied.

2. PREREQUISITES: ESTABLISH CONSEQUENCE LEVELS AND LIKELIHOOD LEVELS

The first two steps are quite simple: Establish consequence and likelihood levels from high to low. Consequence levels refer to the severity of possible negative outcomes. Likelihood levels refer to how probable it is that a particular consequence will actually occur. Combined, the likelihood of particular consequences determines the risk of a potential system failure. During these first two steps, we are not yet specifying any particular regulations or risk management activities. Rather, we are establishing a framework into which systems can be classified according to the severity and probability of consequences.

IEEE 1012 defines a range of four consequence levels from catastrophic (high) to negligible (low) (Table 1).[3] Policy makers could use this same set of consequence levels or could instead vary the number of levels or their exact definitions. The key idea is to establish a range of consequence levels from high to low that provides enough levels of distinction to express the range of regulatory options desired.

Table 1: Table of Consequences from High to Low

Consequences	Definition
Catastrophic	Loss of human life, complete mission failure, loss of system security and safety, or extensive financial or social loss.
Critical	Major and permanent injury, partial loss of mission, major system damage, or major financial or social loss.
Marginal	Severe injury or illness, degradation of secondary mission, or some financial or social loss.
Negligible	Minor injury or illness, minor impact on system performance, or operator inconvenience.

3 IEEE 1012, Annex B, Table B.2, “Definition of consequences”, p. 196.

All systems fall somewhere on this range of consequence levels. If we are using a system for any purpose and it fails to function correctly, then the consequences are at least negligible because there will be operator inconvenience. Risk management actions prescribed for systems with at most negligible consequences can be correspondingly light. For policy makers using IEEE 1012 as an example, it could be reasonable to specify no actions for low-risk systems or to point developers of those systems to voluntary guidance and best practices.

Similarly, we need a scale to express the probability of outcomes. This can be as simple as a list of qualitative terms expressing likelihoods from high to low. IEEE 1012 uses 4 levels, from reasonable to infrequent, as shown in Table 2. IEEE 1012 does not define these likelihood terms in more detail, nor does it map them on to concrete numerical percentages. As with consequence levels, policy makers could choose a different range of likelihoods if desired.

Table 2: Table of Likelihoods used by IEEE 1012 from High to Low

reasonable
probable
occasional
infrequent

NIST Special publication 800-30 offers an example of associating additional qualitative and quantitative information to likelihood levels. For example, they use a range of five likelihood levels - Very High, High, Moderate, Low and Very Low Likelihoods. To each of these levels, they associate numbers between 0 and 10 (e.g. Very High is mapped to 10, Moderate to 5, and Very Low to 0). They also associate numbers between 0 and 100 (e.g. Very High is mapped to the range 96-100, Moderate to 21-79 and Very Low to 0-4).

3. FILL IN THE MAP: INTEGRITY LEVELS

With the foundation of consequence levels and likelihood levels in place, the third step is to establish integrity levels[4] by establishing a map between these consequence and likelihood levels. This will take the form of a 2-D grid with consequence levels on one axis and likelihood levels on the other.

Table 3[5] offers a concrete example of this. It is a two-dimensional matrix with the 4 levels of likelihood (reasonable, probable, occasional, infrequent) listed horizontally and the 4 levels of consequence (catastrophic, critical, marginal, negligible) listed vertically. The interior cells of Table 3 contain the numbers from 1 to 4 corresponding to the 4 integrity levels. As one might expect, the highest integrity level (4) appears in the upper-left corner of the table corresponding to high consequence and high likelihood. Similarly, the lowest integrity level (1) appears in the lower-right corner of the table corresponding to low consequence and low likelihood. IEEE 1012 includes some overlaps between the integrity levels to allow for individual interpretations of acceptable risk depending on the application. For example, the cell corresponding to occasional likelihood of catastrophic consequences can map onto integrity level 3 or 4.

For policy makers using IEEE 1012 as an example, step 3 involves specifying a similar matrix. The number of integrity levels could vary as well as the exact way in which they are mapped onto consequence levels and likelihood levels. Policy makers could start with a fewer number of levels. For example, even two tiers, regulated and non-regulated could be quite effective or three tiers, high risk/highly-regulated, medium risk/lightly-regulated and low risk/voluntary risk management guidance. The key idea is to establish a range of consequence levels from high to low that provides enough levels of distinction to express the range of regulatory options desired.

Table 3: Map of Integrity Levels onto a Combination of Consequence and Likelihood Levels

	Likelihood of occurrence of an operating state that contributes to the error (decreasing order of likelihood)			
Error consequence	**Reasonable**	**Probable**	**Occasional**	**Infrequent**
Catastrophic	4	4	4 or 3	3
Critical	4	4 or 3	3	2 or 1
Marginal	3	3 or 2	2 or 1	1
Negligible	2	2 or 1	1	1

Table 4[6] is another way to present the same information as in Table 3. Rather than being organized by the intersection of consequence level and likelihood level, it is simply ordered by integrity level. Notice that the description in Table 4 is a narrative representation of where each integrity level appears in Table 3.

4 Integrity levels are sometimes called risk tiers in other contexts. It is important to note that risk tiers can be based on many things. IEEE integrity levels or risk tiers focus specifically on the severity of consequences and likelihood of those consequences, unlike other risk tiering systems that identify risk tiers based on the presence of specific technologies (e.g. facial recognition) or application-areas (e.g. health care).

5 IEEE 1012, Annex B, Table B.3, “Graphic illustration of the assignment of integrity levels”, p. 196.

6 IEEE 1012, Annex B, Table B.1, “Assignment of integrity levels”, p. 196.

Table 4: Table of Integrity Levels from High to Low

Integrity Level	Description
4	Behavior of the system, in combination with its environment, causes the following: • Catastrophic consequences for which the likelihood of the behavior occurring is at most occasional or • Critical consequences for which the likelihood of the behavior occurring is at most probable
3	Behavior of the system, in combination with its environment, causes the following: • Catastrophic consequences for which the likelihood of the behavior occurring is at most infrequent or • Critical consequences for which the likelihood of the behavior occurring is at most occasional or • Marginal consequences for which the likelihood of the behavior occurring is at most probable
2	Behavior of the system, in combination with its environment, causes the following: • Critical consequences for which the likelihood of the behavior occurring is at most infrequent or • Marginal consequences for which the likelihood of the behavior occurring is at most probably or • Negligible consequences for which the likelihood of the behavior occurring is at most reasonable
1	Behavior of the system, in combination with its environment, causes the following: • Critical consequences for which the likelihood of the behavior occurring is at most infrequent or • Marginal consequences for which the likelihood of the behavior occurring is at most occasional or • Negligible consequences for which the likelihood of the behavior occurring is at most probable

For many types of autonomous and intelligent systems, there has not been a thorough public conversation to fully reach a consensus on appropriate integrity level designations. The lack of such a consensus, however, does not limit the value of IEEE 1012 or preclude its use. For example, IEEE has previously stated that while there is no consensus on an integrity level for DNA software, digital forensics, and many other forensic techniques, when software and hardware is used to generate evidence in the criminal legal system, there is the possibility of catastrophic failure (including loss of life and liberty), and therefore, should be governed at the highest integrity level with classical IV&V (the highest technical, managerial, and financial independence).[7] Similarly, for generative AI and large language models (LLMs), regulators still ought to determine the initial integrity level and proceed accordingly, even as a consensus is developed for each specific application.

7 IEEE-USA AI Policy Committee, Letter to the National Institute of Standards and Technology (NIST) providing comments on NIST Internal Report 8351-DRAFT DNA Mixture Interpretation: A NIST Scientific Foundation Review, November 18 2021.

4. THE CORE WORK: ASSIGN REQUIREMENTS TO EACH INTEGRITY LEVEL

Steps 1-3, as discussed in Sections 2 and 3 of this document, set a framework for risk management with consequence levels, probability levels and integrity levels. Step 4, described in this section, actually uses this framework to specify actions and thus represents the most substantive work for policy makers. This is also where it may make most sense for policy makers to depart from IEEE 1012 itself if desired. IEEE 1012 focuses on verification and validation (V&V) activities and policy makers should consider including some of those for risk management purposes, but policy makers also have a much broader range of possible intervention alternatives available to them from education and voluntary guidance, to requirements for disclosure and documentation and oversight, to prohibitions and penalties.

In Section 3, we discussed how one way to think about integrity levels is how they are mapped onto the combination of consequences and likelihoods using a matrix like Table 3. This mapping is one key way the different integrity levels are defined, but integrity levels are also crucially defined by the actions or interventions specified for each level. In IEEE 1012, this is the precise set of verification and validation (V&V) activities prescribed for each level. For policy makers following IEEE 1012 as a framework, the range of possible activities is quite wide - regulations, oversight, documentation, certification, penalties, tracking, reporting, education, voluntary guidance, or any other action desired.

When considering the activities to assign to each integrity level, one common sense place to begin is by assigning actions to the highest integrity level where there is the most risk and then proceeding to reduce the intensity of those actions as appropriate for lower levels. Policy makers could begin by answering the question of what actions they wish to require for the highest integrity levels or what is the most intense set of governance activities they wish to describe. This might include pre-deployment approval or oversight, or even prohibitions on some technologies. If the current state of official AI risk management policy is voluntary guidance for industry then policy makers should ask themselves whether they believe that is sufficient for the highest risk systems. If not, they have an opportunity to specify a tier of required action for these highest risk systems, as identified by the consequence levels and probability levels discussed earlier. They can specify the tier of systems for which risk management activities are required without a concern that they will inadvertently introduce barriers for all AI systems, even low risk, internal systems. This is a great way to balance both concern for public welfare and management of severe risks with the desire not to stifle innovation.

Once policy makers have specified a set of required activities or interventions for the highest integrity level, they could move to the lowest level and ask if any of these actions are still appropriate. They could, for example, specify voluntary risk management best practices that are wise for all systems, but not require any actions or they could decide that they want some required actions, even if less extensive, for the lowest integrity level. Once the highest and lowest levels have been filled in, policy makers can identify if there is a need for intermediate levels. They can fill in as many additional intermediate levels as they deem necessary to represent an appropriate range of response across the full spectrum of integrity levels.

This framework of consequence levels, likelihood levels and integrity levels from IEEE 1012 gives policy makers a framework that allows them to target interventions they deem necessary for high-risk systems without adding an undue burden to lower risk systems. Systems where more severe consequences occur with higher likelihood deserve a greater investment in risk management. Systems where consequences are negligible and/or where consequences occur with lower frequency do not warrant as great an investment.

For policy makers interested in standardizing and augmenting commercial practices of V&V activities with oversight and regulation, the US Food and Drug Administration (FDA) Classification of Medical Devices offers an interesting example of a system with 3 risk tiers [4]. Specifically, it defines 3 classes of medical devices. Class I devices involve the lowest risk, defined as presenting minimal potential for harm to patients. Class III devices involve the highest risk. Devices in this class sustain or support life, are implanted, or present potential unreasonable risk of illness and injury. There are a set of general controls described for all devices, including Class 1, by the Federal Food, Drug, and Cosmetic Act (FD&C) Act [5]. There are additional requirements (special controls) for Class II devices such as performance standards, postmarket surveillance, patient registries, special labeling requirements, and premarket data requirements. Class III devices are further subject to approval of a Premarket Approval Application.

5. INCLUDING RISK MANAGEMENT THROUGHOUT THE FULL SYSTEM LIFECYCLE

IEEE 1012 recognizes that managing risk effectively means requiring action throughout the lifecycle of the system, not simply focusing on the final operation of a deployed system.[8] Similarly, policy makers need not be limited to placing requirements on the final deployment of a system. They can require actions throughout the entire process of considering, developing and deploying a system.

IEEE 1012 recognizes the following phases of a system's life cycle and specifies V&V activities throughout these phases, with more intensive activities required at each phase for higher integrity levels.

- Concept
- Requirements
- Design
- Construction
- Integration
- Qualification Testing
- Acceptance Testing
- Verification
- Installation and Checkout
- Validation
- Operation
- Maintenance
- Disposal

The full life cycle of a system begins when it is initially conceived (the concept phase). This initial concept is expanded and formalized as stakeholders are consulted for a full set of system requirements describing detailed success criteria to be met by the system. The design phase takes those formal requirements and proposes a detailed plan and structure for how the requirements can be met. Construction then involves the actual implementation of the system and in particular, the implementation of all pieces of the system specified by the design. Integration brings all the pieces of the system that have been constructed together into a working system. Individual pieces are typically tested in isolation, but integration testing often reveals problems or flaws that were not identified when looking only at individual components. The system is then tested first internally and then by stakeholders in increasing real-world scenarios. Validation finally determines if the correct system has actually been constructed to meet the requirements laid out by all stakeholders. Once the system is deployed, there are many opportunities for the system to evolve. During operation of the system, failures and unintended consequences are often identified and repairs must be made. Similarly, additional features may be requested and added. Changes to the system once it is deployed are considered system maintenance. Finally, there may come a time when the system is no longer needed and is decommissioned.

Policy makers may not choose to make such fine grained distinctions as these 13 lifecycle phases from IEEE 1012, but it is still important and effective to consider interventions that do not simply target the system post-deployment. For example, policy makers would be wise to consider requirements on how stakeholders are consulted/involved at the early phases of concept and design or requirements for testing before deployment. Specifically, involving a wide variety of stakeholders in identifying the severity of possible consequences and their likelihood is important to effective risk management [2]. They could also consider requirements related to iterative improvement and debugging of the system after deployment including processes for collecting and responding to reports of errors or other negative impacts, and processes for redress for those impacted by the system. Specifically, policy makers could require that data is collected and published to update consequence severity and likelihood estimates made pre-deployment. For example, system designers might estimate the likelihood of a particular consequence as low but then find evidence after deployment that it occurs frequently.

8 The NIST AI Risk Management Framework also recommends risk management activities throughout the life cycle of an AI system [3].

6. DEFINING AN APPROPRIATE LEVEL OF INDEPENDENT REVIEW TO REQUIRE

1. Require a level independent review of risk management activities that is appropriate to each integrity level

When the people who develop a system are the same ones to evaluate its integrity and safety, they have both a difficulty thinking out of the box about what problems may remain and a vested interest in a positive outcome. A proven way to improve outcomes is to require independent review of the verification and validation activities. Therefore, policy makers should strongly consider requirements for independent review of risk management activities through the life cycle of AI/S systems in any proposed regulation. This is a final and crucial step that policy makers can take following the example of IEEE 1012.

IEEE 1012 recognizes that independent review is crucial to the reliability and integrity of outcomes and the management of risk. It further tackles the question of what really constitutes independent review; policy makers can include aspects of these definitions in their own requirements. Specifically, IEEE 1012 describes three crucial aspects of independent review: technical independence, managerial independence, and financial independence as shown in Table 5[9] and requires more intense levels of independent review for higher integrity levels as shown in Tables 6 and 7.

Table 5: Definitions of Technical, Managerial and Financial Independence

Type of Independence	Requirement
Technical	Verification and Validation effort uses personnel who are not involved in the development of the system or its elements.
Managerial	Responsibility for the Verification and Validation effort be vested in an organization separate from the development and program management organizations.
Financial	Control of the Verification and Validation budget be vested in an organization independent of the development organization.

Table 5 provides concise definitions of technical, managerial and financial independence. IEEE 1012 further elaborates on these definitions as follows:

Technical Independence

Technical independence requires personnel who are not involved in the development of the system or its elements. Reviewers should formulate their own understanding of the problem and how the proposed system is solving the problem. Technical independence ("fresh viewpoint") is an important method to detect subtle errors overlooked by those too close to the solution.

9 IEEE 1012 Annex C, "Definition of independent verification and validation (IV&V)."

Managerial Independence

This requires that the responsibility for independent review be vested in an organization separate from the development and program management organizations. Managerial independence also means that reviewers independently select the segments of the software, hardware, and system to analyze and test, choose the techniques, define the schedule of activities, and select the specific technical issues and problems to act on. Reviewers should be allowed to submit results, anomalies, and findings without any restrictions (e.g., without requiring prior approval from the development group) or adverse pressures, direct or indirect, from the development group.

Financial Independence

This requires that control of the review budget be vested in an organization independent of the development organization. This independence prevents situations where the review effort cannot complete its analysis or deliver timely results because funds have been diverted or adverse financial pressures or influences have been exerted.

There are also different ways organizations could attempt to conduct reviews. The extent to which each of the three independence parameters (technical, managerial, and financial) is vested in a review effort determines the degree of independence achieved. Table 6 lists five prevalent forms of independent review: 1) classical, 2) modified, 3) integrated, 4) internal, and 5) embedded along with a row representing no attempt at independent verification and validation. Table 6 also rates the degree of independence achieved by each of these forms from High to None. Forms rated HIGH offer more benefits from independent review than forms rated CONDITIONAL, MINIMAL or NONE. Policy makers following IEEE 1012 as an example should consider requiring higher forms of independent review (like Classical or Modified) for the highest risk systems.

IEEE 1012 further elaborates on the contents of Table 6.

Table 6: Rating forms of Independent Verification and Validation By Technical, Managerial and Financial Parameters

IV&V Form	Technical	Management	Financial
Classical	HIGH	HIGH	HIGH
Modified	HIGH	CONDITIONAL	HIGH
Integrated	CONDITIONAL	HIGH	HIGH
Internal	CONDITIONAL	CONDITIONAL	CONDITIONAL
Embedded	MINIMAL	MINIMAL	MINIMAL
No IV&V	NONE	NONE	NONE

Classical IV&V

Classical IV&V embodies all three independence parameters. The IV&V responsibility is vested in an organization that is separate from the development organization. The IV&V effort establishes a close working relationship with the development organization to assure that IV&V findings and recommendations are integrated rapidly back into the development process. Typically, classical IV&V is performed by one organization (e.g., supplier) and the development is performed by a separate organization (i.e., another vendor). IEEE recommends Classical IV&V for systems at integrity level 4 (i.e., loss of life, loss of mission, significant social loss, or financial loss) through regulations and standards imposed on the system development.

Modified IV&V

Modified IV&V is used in many large programs where a prime integrator is selected to manage the entire system development including the IV&V. The prime integrator selects organizations to assist in the development of the system and to perform the IV&V. In the modified IV&V form, the acquirer reduces its own acquisition time by passing this responsibility to the prime integrator. Because the prime integrator performs all or some of the development, the managerial independence is compromised by having the IV&V effort report to the prime integrator. Technical independence is preserved because the IV&V effort formulates an unbiased opinion of the system solution and uses an independent staff to perform the IV&V. Financial independence is preserved because a separate budget is set aside for the IV&V effort. IEEE 1012 recommends that a Modified IV&V effort would be appropriate for systems with integrity level 3 (i.e., an important mission and purpose).

Integrated IV&V

This form is focused on providing rapid feedback of V&V results into the development process and is performed by an organization that is financially and managerially independent of the development organization to minimize compromises with respect to independence. The rapid feedback of V&V results into the development process is facilitated by the integrated IV&V effort: working side by side with the development organization, reviewing interim work products, and providing V&V feedback during inspections, walkthroughs, and reviews conducted by the development staff (potential impact on technical independence). Impacts to technical independence are counterbalanced by benefits associated with a focus on interdependence between the integrated IV&V effort and the development organization. Interdependence means that the successes of the organizations are closely coupled, ensuring that they work together in a cooperative fashion.

Internal IV&V

Internal IV&V exists when the developer conducts the IV&V with personnel from within its own organization, although preferably not the same personnel involved directly in the development effort. Technical, managerial, and financial independence are compromised. Technical independence is compromised because the IV&V analysis and test is vulnerable to overlooking errors by using the same assumptions or development environment that masked the error from the developers. Managerial independence is compromised because the internal IV&V effort uses the same common tools and corporate analysis procedures as the development group. Peer pressure from the development group may adversely influence how aggressively the system is analyzed and tested by the IV&V effort. Financial independence is compromised because the development group controls the IV&V budget. IV&V funds, resources, and schedules may be reduced as development pressures and needs redirect the IV&V funds into solving development problems. The benefit of an internal IV&V effort is access to staff who know the system and its software. This form of IV&V could be used when a specific degree of independence is not explicitly stated and the benefits of preexisting staff knowledge outweigh the benefits of objectivity. Internal IV&V is not recommended for high integrity levels, but it can still be beneficial at lower integrity levels.

Embedded IV&V

This form is similar to internal IV&V in that it uses personnel from the development organization who should not be involved directly in the development effort. Embedded V&V is focused on ensuring conformance to the development procedures and processes. The embedded V&V effort works side by side with the development organization and attends the same inspections, walkthroughs, and reviews as the development staff (i.e., compromise of technical independence). Embedded V&V is not tasked specifically to assess independently the original solution or conduct independent tests (i.e., compromise of managerial independence). Financial independence is compromised because the V&V staff resource assignments are controlled by the development group. Embedded V&V allows rapid feedback of V&V results into the development process, but compromises the technical, managerial, and financial independence of the V&V effort. Embedded IV&V is not recommended for high integrity levels, but it can still be beneficial at lower integrity levels.

No IV&V

If members of the development team are the only ones engaging in verification and validation efforts, then there is no independence to those efforts. It is still possible for the development team to spend more resources or fewer resources on verification and validation efforts, but regardless, they would not be independent verification and validation efforts. It is specifically worth noting that peer reviewed publications authored by the development team are also not a form of IV&V [6].

Following the proven example of IEEE 1012, policy makers can and should consider both including requirements for independent review and specifying clearly what type of independent review is required at each integrity level. Table 7 summarizes the recommendations of IEEE 1012 for each integrity level. A higher level of independent review would always be acceptable but never lower. IEEE 1012 recommends some form of independent review at all integrity levels (e.g. at least embedded) rather than review only by the development team itself.

Table 7: Type of Independent Review Recommended for each Integrity Level

	Classical	Modified	Integrated	Internal	Embedded
4	✓	X	X	X	X
3	✓	✓	X	X	X
2			✓	✓	✓
1			✓	✓	✓

7. CONCLUSIONS

IEEE 1012 is a time-tested, broadly accepted, and universally applicable process for ensuring that the right product is correctly built for its intended use. It offers both wise guidance and practical strategies for policy makers seeking to navigate the confusing debates about how to regulate new AI systems. IEEE 1012 could be adopted as-is for verification and validation of software systems, including the new systems based on emerging generative AI technologies that have been the topic of so much recent discussion and concern. However, the focus here is how IEEE 1012 can also serve a high-level framework that policy makers can use, modifying the exact details of consequence levels, likelihood levels, integrity levels, and requirements to better suit their own regulatory intent.

This article has presented a roadmap of six steps to follow:

1. **Establish consequence levels from high to low**
2. **Establish likelihood levels from high to low**
3. **Establish integrity levels using a map between consequence levels and likelihood levels**
4. **Assign requirements through the lifecycle of the system to the highest level and then reduce those requirements as appropriate for lower levels**
5. **Require risk management throughout the full system lifecycle**
6. **Require independent review of risk management activities that is appropriate to each integrity level**

Together, these steps provide clear guidance for policy makers concerned with controlling the most serious consequences of AI systems without harming innovation by introducing overly onerous requirements in applications where the risks are lower.

The verification and validation activities in IEEE 1012 build on decades of both research and practical experience in managing risk in constructing complex software and hardware systems. In IEEE 1012, what determines the intensity of risk management activity prescribed is the severity of consequences from the deployed system and the probability of those risks occurring. IEEE 1012 specifies a set of risk management tasks throughout the lifecycle of the system with higher integrity levels involving a more intensive set of tasks, including more intense forms of independent review. This time-tested and practical framing of risk management from IEEE 1012 can serve as a roadmap and example to policy makers who need concrete guidance on how to effectively manage risk in today's rapidly changing AI landscape.

8. ACKNOWLEDGEMENTS

Thank you to IEEE for allowing extracts of IEEE 1012 to appear in this report and to all the volunteers who contributed to the construction of the 1012 standard since 1998. Thank you also to Marc Canellas, Carlos Ignacio Gutierrez, Anne Toomey McKenna, Matthew O'Shaughnessy, Bruce Hedin, Nathan Adams, and Erica Wissolik for your support and helpful discussions that contributed to this article. Any errors are my own. Thank you to Greg Hill for his beautiful work on formatting this document.

9. REFERENCES

[1] IEEE, IEEE 1012-2016 Standard for System, Software, and Hardware Verification and Validation, https://ieeexplore.ieee.org/document/8055462

[2] National Institute of Standards and Technology (NIST), AI Risk Management Framework 1.0, NIST AI 100-1, January 2023. https://doi.org/10.6028/NIST.AI.100-1

[3] National Institute of Standards and Technology (NIST), Guide for Conducting Risk Assessments, NIST Special Publication 800-30.
https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-30r1.pdf

[4] US Food and Drug Administration, "Classify Your Medical Device,"
https://www.fda.gov/medical-devices/overview-device-regulation/classify-your-medical-device

[5] *Federal Food, Drug, and Cosmetic Act*, https://www.fda.gov/regulatory-information/laws-enforced-fda/federal-food-drug-and-cosmetic-act-fdc-act.

[6] Jeanna Matthews, Bruce Hedin, Marc Canellas
Trustworthy Evidence for Trustworthy Technology: An Overview of Evidence for Assessing the Trustworthiness of Autonomous and Intelligent Systems
IEEE-USA, September 29 2022.

10. APPENDIX - IEEE 1012 ANNEX B AND C

Annex B

(informative)

A risk-based integrity level schema

Table B.1 defines four integrity levels used for reference purposes by this standard. Table B.2 describes the consequences of errors for each of the four integrity levels. There are overlaps between the integrity levels to allow for individual interpretations of acceptable risk depending on the application.

Table B.1—Assignment of integrity levels

Integrity level	Description
4	Behavior of the system, in combination with its environment, causes the following: — Catastrophic consequences for which the likelihood of the behavior occurring is at most occasional or — Critical consequences for which the likelihood of the behavior occurring is at most probable
3	Behavior of the system, in combination with its environment, causes the following: — Catastrophic consequences for which the likelihood of the behavior occurring is at most infrequent or — Critical consequences for which the likelihood of the behavior occurring is at most occasional or — Marginal consequences for which the likelihood of the behavior occurring is at most probable
2	Behavior of the system, in combination with its environment, causes the following: — Critical consequences for which the likelihood of the behavior occurring is at most infrequent or — Marginal consequences for which the likelihood of the behavior occurring is at most probably or — Negligible consequences for which the likelihood of the behavior occurring is at most reasonable
1	Behavior of the system, in combination with its environment, causes the following: — Critical consequences for which the likelihood of the behavior occurring is at most infrequent or — Marginal consequences for which the likelihood of the behavior occurring is at most occasional or — Negligible consequences for which the likelihood of the behavior occurring is at most probable

Table B.2—Definition of consequences

Consequence	Definition
Catastrophic	Loss of human life, complete mission failure, loss of system security and safety, or extensive financial or social loss.
Critical	Major and permanent injury, partial loss of mission, major system damage, or major financial or social loss.
Marginal	Severe injury or illness, degradation of secondary mission, or some financial or social loss.
Negligible	Minor injury or illness, minor impact on system performance, or operator inconvenience.

IEEE 1012-2016 - Reprinted with permission from IEEE, Copyright IEEE 2016. All Rights Reserved.

Table B.3 illustrates the risk-based schema shown in Table B.1 and Table B.2. Each cell in the table assigns an integrity level based on the combination of an error consequence and the likelihood of occurrence of an operating state that contributes to the error. Some table cells reflect more than one integrity level, indicating that the final assignment of the integrity level can be selected to address the system application and risk mitigation recommendations. For some industry applications, the definition of likelihood of occurrence categories may be expressed as probability figures derived by analysis or from system requirements.

Table B.3—Graphic illustration of the assignment of integrity levels

Error	Likelihood of occurrence of an operating state that contributes to the error (decreasing order of likelihood)			
Consequence	**Reasonable**	**Probable**	**Occasional**	**Infrequent**
Catastrophic	4	4	4 or 3	3
Critical	4	4 or 3	3	2 or 1
Marginal	3	3 or 2	2 or 1	1
Negligible	2	2 or 1	1	1

Annex C

(informative)

Definition of independent verification and validation (IV&V)

C.1 Independence parameters

C.1.1 Introduction

Independent V&V (IV&V) is defined by three parameters: technical independence, managerial independence, and financial independence.

C.1.2 Technical independence

Technical independence requires the V&V effort to use personnel who are not involved in the development of the system or its elements. The IV&V effort should formulate its own understanding of the problem and how the proposed system is solving the problem. Technical independence ("fresh viewpoint") is an important method to detect subtle errors overlooked by those too close to the solution.

For system tools, technical independence means that the IV&V effort uses or develops its own set of test and analysis tools separate from the developer's tools. Sharing of tools is allowable for computer support environments (e.g., compilers, assemblers, and utilities) or for system simulations where an independent version would be too costly. For shared tools, IV&V conducts qualification tests on tools to assure that the common tools do not contain errors that may mask errors in the system being analyzed and tested. Off-the-shelf tools that have extensive history of use do not require qualification testing. The most important aspect for the use of these tools is to verify the input data used.

C.1.3 Managerial independence

This requires that the responsibility for the IV&V effort be vested in an organization separate from the development and program management organizations. Managerial independence also means that the IV&V effort independently selects the segments of the software, hardware, and system to analyze and test, chooses the IV&V techniques, defines the schedule of IV&V activities, and selects the specific technical issues and problems to act on. The IV&V effort provides its findings in a timely fashion simultaneously to both the development and program management organizations. The IV&V effort is allowed to submit to program management the IV&V results, anomalies, and findings without any restrictions (e.g., without requiring prior approval from the development group) or adverse pressures, direct or indirect, from the development group.

C.1.4 Financial independence

This requires that control of the IV&V budget be vested in an organization independent of the development organization. This independence prevents situations where the IV&V effort cannot complete its analysis or test or deliver timely results because funds have been diverted or adverse financial pressures or influences have been exerted.

C.2 Forms of independence

C.2.1 Introduction

The extent to which each of the three independence parameters (technical, managerial, and financial) is vested in a V&V effort determines the degree of independence achieved.

Many forms of independence can be adopted for a V&V effort. The five most prevalent are as follows: 1) classical, 2) modified, 3) integrated, 4) internal, and 5) embedded. Table C.1 illustrates the degree of independence achieved by these five forms.

Table C.1—Forms of IV&V

IV&V form	Technical	Management	Financial
Classical	I	I	I
Modified	I	i	I
Integrated	i	I	I
Internal	i	i	i
Embedded	e	e	e

NOTE—I = rigorous independence; i = conditional independence; e = minimal independence.

C.2.2 Classical IV&V

Classical IV&V embodies all three independence parameters. The IV&V responsibility is vested in an organization that is separate from the development organization. The IV&V effort establishes a close working relationship with the development organization to assure that IV&V findings and recommendations are integrated rapidly back into the development process. Typically, classical IV&V is performed by one organization (e.g., supplier) and the development is performed by a separate organization (i.e., another vendor). Classical IV&V is generally required for integrity level 4 (i.e., loss of life, loss of mission, significant social loss, or financial loss) through regulations and standards imposed on the system development.

C.2.3 Modified IV&V

Modified IV&V is used in many large programs where the system prime integrator is selected to manage the entire system development including the IV&V. The prime integrator selects organizations to assist in the development of the system and to perform the IV&V. In the modified IV&V form, the acquirer reduces its own acquisition time by passing this responsibility to the prime integrator. Because the prime integrator performs all or some of the development, the managerial independence is compromised by having the IV&V effort report to the prime integrator. Technical independence is preserved because the IV&V effort formulates an unbiased opinion of the system solution and uses an independent staff to perform the IV&V. Financial independence is preserved because a separate budget is set aside for the IV&V effort. Modified IV&V effort would be appropriate for systems with integrity level 3 (i.e., an important mission and purpose).

C.2.4 Integrated IV&V

This form is focused on providing rapid feedback of V&V results into the development process and is performed by an organization that is financially and managerially independent of the development organization to minimize compromises with respect to independence. The rapid feedback of V&V results into the development process is facilitated by the integrated IV&V effort: working side by side with the development organization, reviewing interim work products, and providing V&V feedback during inspections, walkthroughs, and reviews conducted by the development staff (potential impact on technical independence). Impacts to technical independence are counterbalanced by benefits associated with a focus on interdependence between the integrated IV&V effort and the development organization. Interdependence means that the successes of the organizations are closely coupled, ensuring that they work together in a cooperative fashion.

C.2.5 Internal IV&V

Internal IV&V exists when the developer conducts the IV&V with personnel from within its own organization, although preferably not the same personnel involved directly in the development effort. Technical, managerial, and financial independence are compromised. Technical independence is compromised because the IV&V analysis and test is vulnerable to overlooking errors by using the same assumptions or development environment that masked the error from the developers. Managerial independence is compromised because the internal IV&V effort uses the same common tools and corporate analysis procedures as the development group. Peer pressure from the development group may adversely influence how aggressively the system is analyzed and tested by the IV&V effort. Financial independence is compromised because the development group controls the IV&V budget. IV&V funds, resources, and schedules may be reduced as development pressures and needs redirect the IV&V funds into solving development problems. The benefit of an internal IV&V effort is access to staff who know the system and its software. This form of IV&V is used when the degree of independence is not explicitly stated and the benefits of preexisting staff knowledge outweigh the benefits of objectivity.

C.2.6 Embedded V&V

This form is similar to internal IV&V in that it uses personnel from the development organization who should not be involved directly in the development effort. Embedded V&V is focused on ensuring conformance to the development procedures and processes. The embedded V&V effort works side by side with the development organization and attends the same inspections, walkthroughs, and reviews as the development staff (i.e., compromise of technical independence). Embedded V&V is not tasked specifically to assess independently the original solution or conduct independent tests (i.e., compromise of managerial independence). Financial independence is compromised because the V&V staff resource assignments are controlled by the development group. Embedded V&V allows rapid feedback of V&V results into the development process, but compromises the technical, managerial, and financial independence of the V&V effort.


IEEE
USA