

Opinion

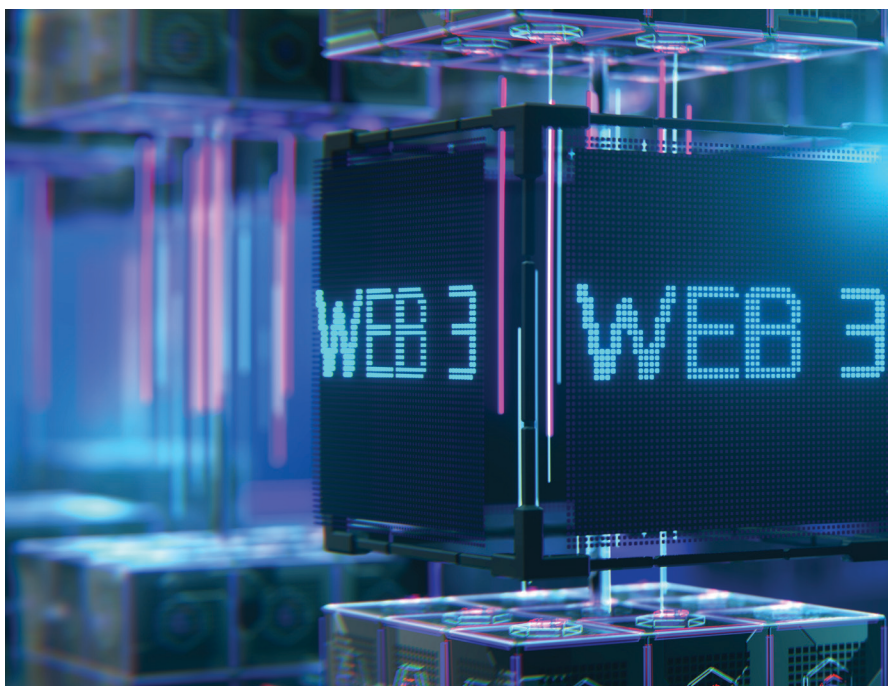
Web 3.0 Requires Data Integrity

New integrity-focused standards are necessary to enable the trusted AI services of tomorrow.

IF YOU'VE EVER taken a computer security class, you've probably learned about the three legs of computer security—confidentiality, integrity, and availability—known as the CIA triad.^a When we talk about a system being secure, that's what we're referring to. All are important, but to different degrees in different contexts. In a world populated by artificial intelligence (AI) systems and artificial intelligent agents, integrity will be paramount.

What is data integrity? It's ensuring that no one can modify data—that's the security angle—but it's much more than that. It encompasses accuracy, completeness, and quality of data—all over both time and space. It's preventing accidental data loss; the “undo” button is a primitive integrity measure. It's also making sure that data is accurate when it's collected—that it comes from a trustworthy source, that nothing important is missing, and that it doesn't change as it moves from format to format. The ability to restart your computer is another integrity measure.

The CIA triad has evolved with the Internet. The first iteration of the Web—Web 1.0 of the 1990s and early 2000s—prioritized availability. This era saw organizations and individuals rush to digitize their content, creating what has become an unprecedented repository of human knowledge. Orga-



nizations worldwide established their digital presence, leading to massive digitization projects where quantity took precedence over quality. The emphasis on making information available overshadowed other concerns.

As Web technologies matured, the focus shifted to protecting the vast amounts of data flowing through online systems. This is Web 2.0: the Internet of today. Interactive features and user-generated content transformed the Web from a read-only medium to a participatory platform. The increase in personal data, and the emergence of interactive platforms for

e-commerce, social media, and online everything demanded both data protection and user privacy. Confidentiality became paramount.

We stand at the threshold of a new Web paradigm: Web 3.0. This is a distributed, decentralized, intelligent Web. Peer-to-peer social-networking systems promise to break the tech monopolies' control on how we interact with each other. Tim Berners-Lee's open W3C protocol, Solid, represents a fundamental shift in how we think about data ownership and control. A future filled with AI agents requires verifiable, trustworthy personal data

^a <https://tinyurl.com/248uyhq6>

and computation. In this world, data integrity takes center stage.

For example, the 5G communications revolution isn't just about faster access to videos; it's about Internet-connected things talking to other Internet-connected things without our intervention. Without data integrity, for example, there's no real-time car-to-car communications about road movements and conditions. There's no drone swarm coordination, smart power grid, or reliable mesh networking. And there's no way to securely empower AI agents.

In particular, AI systems require robust integrity controls because of how they process data. This means technical controls to ensure data is accurate, that its meaning is preserved as it is processed, that it produces reliable results, and that humans can reliably alter it when it's wrong. Just as a scientific instrument must be calibrated to measure reality accurately, AI systems need integrity controls that preserve the connection between their data and ground truth.

This goes beyond preventing data tampering. It means building systems that maintain verifiable chains of trust between their inputs, processing, and outputs, so humans can understand and validate what the AI is doing. AI systems need clean, consistent, and verifiable control processes to learn and make decisions effectively. Without this foundation of verifiable truth, AI systems risk becoming a series of opaque boxes.

Recent history provides many sobering examples of integrity failures that naturally undermine public trust in AI systems. Machine-learning (ML) models trained without thought on expansive datasets have produced predictably biased results in hiring systems. Autonomous vehicles with incorrect data have made incorrect—and fatal—decisions. Medical diagnosis systems have given flawed recommendations without being able to explain themselves. A lack of integrity controls undermines AI systems and harms people who depend on them.

They also highlight how AI integrity failures can manifest at multiple levels of system operation. At the training level, data may be subtly corrupted or biased even before model

AI systems need clean, consistent, and verifiable control processes to learn and make decisions effectively.

development begins. At the model level, mathematical foundations and training processes can introduce new integrity issues even with clean data. During execution, environmental changes and runtime modifications can corrupt previously valid models. And at the output level, the challenge of verifying AI-generated content and tracking it through system chains creates new integrity concerns. Each level compounds the challenges of the ones before it, ultimately manifesting in human costs, such as reinforced biases and diminished agency.

Think of it like protecting a house. You don't just lock a door; you also use safe concrete foundations, sturdy framing, a durable roof, secure double-pane windows, and maybe motion-sensor cameras. Similarly, we need digital security at every layer to ensure the whole system can be trusted.

This layered approach to understanding security becomes increasingly critical as AI systems grow in complexity and autonomy, particularly with large language models (LLMs) and deep-learning systems making high-stakes decisions. We need to verify the integrity of each layer when

A lack of integrity controls undermines AI systems and harms people who depend on them.

building and deploying digital systems that impact human lives and societal outcomes.

At the foundation level, bits are stored in computer hardware. This represents the most basic encoding of our data, model weights, and computational instructions. The next layer up is the file system architecture: the way those binary sequences are organized into structured files and directories that a computer can efficiently access and process. In AI systems, this includes how we store and organize training data, model checkpoints, and hyperparameter configurations.

On top of that are the application layers—the programs and frameworks, such as PyTorch and TensorFlow, that allow us to train models, process data, and generate outputs. This layer handles the complex mathematics of neural networks, gradient descent, and other ML operations.

Finally, at the user-interface level, we have visualization and interaction systems—what humans actually see and engage with. For AI systems, this could be everything from confidence scores and prediction probabilities to generated text and images or autonomous robot movements.

Why does this layered perspective matter? Vulnerabilities and integrity issues can manifest at any level, so understanding these layers helps security experts and AI researchers perform comprehensive threat modeling. This enables the implementation of defense-in-depth strategies—from cryptographic verification of training data to robust model architectures to interpretable outputs. This multi-layered security approach becomes especially crucial as AI systems take on more autonomous decision-making roles in critical domains such as healthcare, finance, and public safety. We must ensure integrity and reliability at every level of the stack.

The risks of deploying AI without proper integrity control measures are severe and often underappreciated. When AI systems operate without sufficient security measures to handle corrupted or manipulated data, they can produce subtly flawed outputs that appear valid on the surface. The failures can cascade through interconnected systems, amplifying errors



Association for Computing Machinery
Advancing Computing as a Science & Profession

ACM Student Research Competition

Attention: Undergraduate and Graduate Computing Students

The ACM Student Research Competition (SRC) offers a unique forum for undergraduate and graduate students to present their original research before a panel of judges and attendees at well-known ACM-sponsored and co-sponsored conferences. The SRC is an internationally recognized venue enabling students to earn many tangible and intangible rewards from participating:

- **Awards:** cash prizes, medals, and ACM student memberships
- **Prestige:** Grand Finalists receive a monetary award and a Grand Finalist certificate that can be framed and displayed
- **Visibility:** meet with researchers in their field of interest and make important connections
- **Experience:** sharpen communication, visual, organizational, and presentation skills

Learn more:

<https://src.acm.org>

Integrity controls will fundamentally shape AI development by moving from optional features to core architectural requirements.

and biases. Without proper integrity controls, an AI system might train on polluted data, make decisions based on misleading assumptions, or have outputs altered without detection. The results of this can range from degraded performance to catastrophic failures.

We see four areas where integrity is paramount in this Web 3.0 world. The first is granular access, which allows users and organizations to maintain precise control over who can access and modify what information and for what purposes. The second is authentication—much more nuanced than the simple “Who are you?” authentication mechanisms of today—which ensures that data access is properly verified and authorized at every step. The third is transparent data ownership, which allows data owners to know when and how their data is used and creates an auditable trail of data providence. Finally, the fourth is access standardization: common interfaces and protocols that enable consistent data access while maintaining security.

Luckily, we’re not starting from scratch. There are open W3C protocols that address some of this: decentralized identifiers^b for verifiable digital identity, the verifiable credentials data model^c for expressing digital credentials, ActivityPub^d for decentralized social networking (that’s what Mastodon uses), Solid^e for distributed data storage and retrieval, and WebAuthn^f

b <https://tinyurl.com/2adqpbta>

c <https://tinyurl.com/25lk298r>

d <https://tinyurl.com/y5ut5tng>


e <https://tinyurl.com/y58xg4bc>

f <https://tinyurl.com/y3hxrkol>

for strong authentication standards. By providing standardized ways to verify data provenance and maintain data integrity throughout its lifecycle, Web 3.0 creates the trusted environment that AI systems require to operate reliably. This architectural leap for integrity control in the hands of users helps ensure that data remains trustworthy from generation and collection through processing and storage.

Integrity is essential to trust, on both technical and human levels. Looking forward, integrity controls will fundamentally shape AI development by moving from optional features to core architectural requirements, much as SSL certificates evolved from a banking luxury to a baseline expectation for any Web service.

Web 3.0 protocols can build integrity controls into their foundation, creating a more reliable infrastructure for AI systems. Today, we take availability for granted; anything less than 100% uptime for critical websites is intolerable. In the future, we will need the same assurances for integrity. Success will require following practical guidelines for maintaining data integrity throughout the AI lifecycle—from data collection through model training and finally to deployment, use, and evolution. These guidelines will address not just technical controls but also governance structures and human oversight, similar to how privacy policies evolved from legal boilerplate into comprehensive frameworks for data stewardship. Common standards and protocols, developed through industry collaboration and regulatory frameworks, will ensure consistent integrity controls across different AI systems and applications.

Just as the HTTPS protocol created a foundation for trusted e-commerce, it’s time for new integrity-focused standards to enable the trusted AI services of tomorrow. 

Bruce Schneier is a public-interest technologist, working at the intersection of security, technology, and people. He is currently chief of Security Architecture at Inrupt Inc. and has been writing about security issues since 1998.

Davi Ottenheimer is vice president of Trust and Digital Ethics at Inrupt Inc., serving as a security strategist and implementer working at the intersection of technology, ethics, and governance. Ottenheimer has been leading large-scale enterprise security innovations and operations across multiple industries for more than 30 years.