Federal Register Notice: 89 FR 51554, Federal Register :: Networking and Information Technology Research and Development Request for Information on Digital Twins Research and Development, June 18, 2024.

Request for Information on the National Digital Twins R&D Strategic Plan

BlockScience

Comments Submitted by BlockScience, University of Washington APL Information Risk and Synthetic Intelligence Research Initiative (IRSIRI), Cognitive Security and Education Forum (COGSEC), and the Active Inference Institute (AII) to the Networking and Information Technology Research and Development National Coordination Office's Request for Comment on The Creation of a National Digital Twins R&D Strategic Plan NITRD-2024-13379

Contents

Questions, responses, and requests related to this document may be directed to info@block.science

July 28, 2024

Submitted to:
NITRD/NCO
National Science Foundation | White House Office of Science and Technology Policy
[redacted]

The collaborating representatives applaud the Networking and Information Technology Research and Development National Coordination Office (NITRD/NCO) on its role in facilitating the development of a whole-of-government strategy for research investment on digital twin and model-based systems engineering, and appreciate the opportunity to provide recommendations and perspectives on the topics of (i) **data management infrastructure**, (ii) **trustworthiness and uncertainty quantification**, (iii) **standardization**, (iv) **responsible use**, (v) **professionalization and workforce development**, (vi) **commercial use**, and (vii) **sustainability and provisioning** in the context of digital twin implementation.

The collaborating representatives provide a unique synthesis of perspectives and recommendations on these topics, with consideration for business, operations, legal, technical, and social (BOLTS) use cases and risks, driven by their combined relationships and work within (i) universities, (ii) think-tanks, (iii) standards-setting organizations, (iv) global nonprofits and non-governmental organizations (NGOs), (v) corporations, (vi) military and government agencies, (vii) international policy and standards setting initiatives, and (viii) interdisciplinary academic, professional, and government communities of practice; and background in and prior work on (i) knowledge, information, data, and reference management and library science, (ii) intelligence and sensor fusion, (iii) interorganizational information exchange, (iv) law and legal engineering, (v) data and content verification, (vi) adtech, human factors, cognitive security, and social systems engineering, (vii) red teaming and adversarial use analysis, (viii) data poisoning and information quality control, (ix) sociotechnical systems and digital governance, (x) complexity science and dynamical systems, (xi) cognitive modeling, (xii) artificial intelligence, (xiii) cybernetics, robotics, control theory, and model-based systems engineering, and (xiv) mechanism, market, and institution design.

This response is organized into two sections: (1) background information and basis, and (2) clear, summary recommendations. The collaborating representatives have endeavored to keep this submission concise and policy-oriented, without sacrificing nuance.

BLOCK SCIENCE

IRSIRI

Contributing Organizations and Representatives

Dr. Michael Zargham [1] Dr. David Sisson [1] Scott David J.D., LL.M. [2]
Dr. Daniel Ari Friedman [3, 4] R.J. Cordes [1, 3]

1. BlockScience

BlockScience (Block.Science) is a complex systems engineering, research and development, and analytics firm focused on the development and governance of safe, ethical, and resilient socio-technical systems. Sourcing insight and expertise on technology, economics, and governance from a diverse, interdisciplinary, and international community of scientists and engineers, BlockScience provides services to a wide range of clients and contributes to working groups, standards development, communities of practice, open-source projects, and academic literature related to model-based systems engineering, artificial intelligence, operations research, market design, network science, distributed systems, and modeling and simulation.

2. Information Risk and Synthetic Intelligence Research Initiative (IRSIRI)

The Information Risk and Synthetic Intelligence Research Initiative (IRSIRI) at University of Washington's Applied Physics Laboratory is an interdisciplinary program that integrates theory and practice for information risk management across business, operating, legal, technical and social (BOLTS) domains, and engages in research and development of processes to help guide emergent distributed interaction governance structures.

3. The Cognitive Security and Education Forum (COGSEC)

The Cognitive Security and Education Forum (COGSEC.org) was formed to convene experts to contribute to knowledge management and education infrastructure within the context of Cognitive Security - which refers to practices, methodologies, and efforts made to defend against social engineering attempts or intentional and unintentional manipulations of and disruptions to cognition and sensemaking at the scale of individuals, organizations, and societies. It hosts yearly initiatives to facilitate and support interdisciplinary and interorganizational research and engineering within related fields and industries.

4. Active Inference Institute (AII)

The Active Inference Institute (activeinference.institute) is dedicated to learning, researching, and applying Active Inference. AII provides avenues for connection and integration with broad audiences and disciplines and a setting for people to aid each other in pursuit of a better understanding of Active Inference. The Institute organizes education, research, and communications to advance the progress and public awareness of frontier knowledge in Active Inference and closely related topics.

Introduction

Digital Twins are useful enough to be dangerous. US Government Agency interest in funding and facilitating research, development, engineering, and implementation of Digital Twins (alongside factors related to their safe implementation) is therefore both reassuring and urgently necessary. Factors such as trustworthiness, reliability, interoperability, stability, sustainability, and responsible use must be addressed now, as there may not be another opportunity to do so before mass proliferation. If these factors can be adequately addressed, Digital Twins hold the potential to integrate physical and digital space – sparking a renaissance of capability exploration that will expand the horizons of research and commerce. If they are not, Digital Twins will inadvertently – but inevitably – become an evergreen source of threats and frustrations that will continue to challenge future generations. Here we argue that (i) conceptually, Digital Twins are not new – and thus we can learn from the common vulnerabilities, exploits, and remedies developed by prior approaches to closely-related problems in control theory and cybernetics, (ii) stable reference and data management capabilities and provisioning considerations are the underlying (but often-overlooked) prerequisites to building reliable Digital Twins, and (iii) the functional surface of a Digital Twin is roughly identical to its threat surface. We conclude with summary recommendations.

Background

In this section we provide background information and the basis for the summary recommendations offered in the section that follows.

I. Information Twins are a Time-Tested Approach

The notion of a "Digital Twin" gains traction in the early 2000's, but ***the underlying concept of mirroring the properties and state of physical objects, systems, and organizations in information space emerges far earlier*****,** in areas such as (i) aeronautics, (ii) robotics, (iii) cybernetics, (iv) finance, accounting, and business management, (v) governance, (vi) military science and command and control, (vii) library science, and (viii) logistics. More importantly, both the theory and practice of managing and maintaining information twins has been time tested for decades in spaces with (a) high-reliability conditions (e.g., military and transport aeronautics), (b) interorganizational use cases (e.g., automotive industry), and (c) requirements related to public and environmental hazards (e.g., chemical manufacturing and nuclear power). **Digital Twins are an expansion on prior art related to model reference based forms of control**, with an eye toward inclusion of new affordances, levels of accuracy, computational and forecasting capabilities, interaction affordances, and, most importantly, new areas of implementation. Consequently. **R&D Activity in this domain will include professionals and academics from disciplines that previously did not require an engineering background, or familiarity with control systems and/or model-based systems engineering.**

The depth of prior art in related fields and the breadth of new domains of implementation creates substantial risk of "re-inventing the wheel" and redundant work. For example, best practices, case studies, and toolkits for control, sensor fusion, and requirements engineering related to modeling (and around managing expectations related to modeling) complex systems already exist, but have not necessarily been made generalizable or accessible. If researchers are unfamiliar with the state of the art, they are likely to waste time, money, and effort attempting to advance it.

- **Recommendation 1.1:** R&D Activity should investigate and be complemented by workforce, competency, and professional development related to the art, science, and practice of model-based systems engineering.
- **Recommendation 1.2:** R&D Activity should prioritize professionalization within the context of Digital Twin implementation to ensure engineering capabilities and standards can be generalized or developed as a foundation for setting, communicating, and verifying safety and other requirements.
- **Recommendation 1.3:** R&D Activity should avoid "re-inventing the wheel" by mapping the extant, conceptual terrain. Common Vulnerabilities and Exploits (CVE) and other community pattern-finding and data-basing initiatives may be a functional means of creating a bridge between extant practices, patterns, risks, and remedies, and the interdisciplinary communities necessary to advance Digital Twin methodologies in new domains.

II. Model-Reality-Specification Gap

Even "identical twins" (two human beings with the same DNA) are never *exactly* alike. Although both twins are "generated from the same genetic code," neither twin simply *is* that code; rather, each twin is a distinct *implementation* of a common *specification*, and each will undergo distinct experiences that further differentiate them over the course of their lifetimes. **For the same reasons, a Digital Twin will never be perfectly mapped with its physical counterpart – and the cyber-physical gap between a real-world system and its digital representation will inevitably grow over time. The gap between model and "reality," however, is not the only gap of concern.** Digital Twins will inevitably have to manage not only the reference-referent gap that exists between the model and the *implemented* physical system, but also the reference-referent gaps between (i) the model and the *specifications* of the physical system's subcomponents, (ii) the *specifications* of the model and the *implemented* model, (iii) the systems' sensors and the *specifications* of those sensors, etc. – gaps that all **widen over time** at "power law"-driven rates[1] due to wear-and-tear, replacements, adjustments, modifications, patches, sabotage and malfeasance, perverse incentives, and other entropic factors.

[1] A "power law" describes a functional relationship between two quantities such that a relative change in one quantity leads to a proportional relative change in the other, causing change to occur at exponential rates.

In the fields of hardware security and supply chain engineering, there is already growing concern that these factors are driving **an expanding gap between components and component specifications that has not been properly assessed.** Failure to address specification gaps proactively will substantially increase the likelihood that Digital Twins do more harm than good.

Reference-referent problems are so fundamental that problems of digital "identity" are indistinguishable from classical problems of metaphysical identity and epistemics, such as (i) the "Ship of Theseus" paradox (i.e. *how many parts of an object can be replaced before that object becomes a different object?)*, (ii) the "Sorites Paradox" (e.g., *how many like-objects can be removed from the system before that system should be given a new descriptor?*), (iii) the "River Thames Problem" as posed by Bertrand Russell (i.e. *how does one draw the boundaries of a system or object in cases where those boundaries are subjective?*), (iv) the "Frame Problem" in Artificial Intelligence (i.e. *how do we decide what is relevant or in context without considering all that is not?*), and (v) Heraclitus' "River Paradox" (i.e. *if a continuous system is never the same system, at what point do we define phase transition or assign new identities?*). **These philosophical framings are not merely of academic interest - they are the bases for our legal and commercial framings for identity**. The GDPR-based laws of the EU reflect identity notions based on Hegel and Kant. US privacy and identity laws reflect philosophies of Locke and the Utilitarians. Digital Twins will challenge existing notions of "identity" in myriad ways across business, operations, legal, technical, and social (BOLTS) domains (and their respective performance metrics); thus, the implementation of interacting Digital Twins at scale will require us to update and clarify our understanding of fundamental philosophical concepts.

If and to the extent that existing and historical notions of "identity" (as broadly conceived) can help to stabilize our organization and operation of future Digital Twin systems, it will help us to most effectively direct our attention and resources to those domains and aspects of Digital Twin infrastructure that display less linear behaviors. Digital Twins will require a "neighborhood watch" relationship with humans to maintain stable function in an exponentially-expanding information space. Existing "identity" standards efforts might be usefully and normatively cross referenced to avoid redundancy in research and to increase clarity in approach.

Further, many of the systems of interest require "multiphysics" approaches, in which there is no unified approach to modeling dynamics, but instead a collection of approaches which are fit-for-function for particular areas of the system. This means that the same systems may not only be represented using different boundaries or functions, but also different *collections* of boundaries and functions – resulting in a perceivably infinite number of valid representations and related identities. Therefore, regardless of the consistency or intensity of enforcement functions and standardization, ***Digital Twins can contain multiple overlapping representations of systems or have variable representations of objects contained within them; consequently, their use in a given situation may be ineffective – or even fundamentally misleading.*** It is important that researchers and engineers recognize that **identity, system state, and specifications are intrinsically fuzzy, whether we treat them this way technically or not.**

The intent to design and implement "ecosystems" of Digital Twins – or interaction surfaces among Digital Twins and physical systems – means that **model-reality-specification gaps can generate cascading "telephone-game" errors**. **Furthermore, affordances for digital systems to interact with or command physical counterparts means that these modeling errors can spill into the physical world.** In addition to being able to interact through physical and digital means, **Digital Twins can be represented within one another and process digital objects** (i.e. a product that has both a physical form and digital representation may move through multiple Digital Twin systems, potentially operated by different organizations). ***Lack of common reference architecture for digital objects could create inconsistencies in resulting data outputs which may undermine forecasting, training data pipelines, and intelligence fusion capabilities in ways that are not easily detected until after the damage is already done – and may even be irreversible.***[2] The absence of a common reference architecture may also result in various misuses of data. For example, in cases where Digital Twins are concerned with medical or cognitive systems, there are nontrivial requirements related to the use, storage, and anonymization of data.

Finally, **Digital Twins have maintenance requirements**. As noted, due to wear-and-tear, replacements, adjustments, modifications, and other factors, the gap between a physical system and its digital representation increases over time. Cyber-physical integration thus requires both physical and digital logistics and security considerations. The initial implementation of such systems implies requirements related to **initial provisioning** (i.e. the planning of logistics related requirements for supporting and maintaining a system for its initial period of service), and their use implies requirements related to **assured provisioning** (i.e. the planning of logistics related requirements for rendering the support and maintenance of a system sustainable and reliable for the duration of its expected service/life-cycle).

- **<u>Recommendation 2.1:</u>** Scientific and technical R&D Activity should be complemented by facilitation of the formation of professional and trade associations that can offer continuing professional development, standardization, and certification related to data and reference management and specification assertions and claims (e.g., *has this component been verified as consistent with its digital representation and/or with its reference specification?*). Such organizations can be helpfully cultivated through connection and normative cross reference to the standards, protocols, practices and policies of existing professional and trade associations at the intersection of identity and digital representations of humans across business, operating, legal, technical and social domains.

[2] As an example of irreversible inconsistency, disagreements over the validity of data related to certain kinds of systems (e.g., those which may include a canonical ledger) may result in cases where organizations disagree over overall system identity or state at a particular time-step and must therefore *fork* their representation, from which point the forked paths can never again be reconciled.

- **Recommendation 2.2:** Common data and reference management schemes should be considered critical infrastructure for Digital Twin "ecosystem" implementations. R&D Activity should explore new approaches to interorganizational reference management and data sharing, with a prioritization on use cases where content location, schema, ontology, or underlying data may be unstable or not agreed upon across organizations.
- **Recommendation 2.3:** R&D Activity related to implementations should require consideration of both initial provisioning and assured provisioning related to security, maintenance, and other logistics requirements.

III. Digital Twins are a Threat Surface

The digital representation of the physical system is a threat surface. A cyber-physical system (e.g., a physical system with a Digital Twin) is a distributed system composed of (i) network and authentication protocols, (ii) software, firmware, and hardware components, (iii) APIs and digital asset exchange mechanisms, (iv) sensor arrays, (v) specification- and asset-reference protocols, and (vi) supervisory control and data acquisition (SCADA) interfaces, all of which are points of interaction that represent potential attack vectors. In other words, **the functional surface of a cyber-physical system is essentially indistinguishable from its threat surface.** By creating a reliable, computational digital representation of a physical system, we create a highly efficient targeting apparatus and basis for disruption. ***It is important to repeatedly acknowledge that Digital Twins are useful enough to be dangerous*** – their use in public health, critical infrastructure, and supply chains represent national security risks as much as they represent opportunities for efficiency, stability, and situational awareness. Further, their use in modeling cognition should be approached with extreme caution – the potential intrusions on cognitive liberty and related cognitive security risks created by the use of such models by opportunists and threat-actors should be considered reason for very serious concern. As can be learned from prior work on information-mirroring models in aeronautics, implementation of a sensor array is effectively the implementation of a new attack vector, thus **information warfare and information security are inseparable from the introduction of reliable sensor arrays**. Targeting sensor arrays which are upstream of system action is often cheaper and more accessible than direct, disruptive action.

While the technical risks associated with reliable models and the risks they pose in various domains are reasonably well known, the human factors and cognitive security risks related to perception and use of Digital Twin and supervisory control and data acquisition (SCADA) interfaces are equally important and often overlooked. The kinds of model-reference adaptive control (MRAC), augmentation control, and SCADA control functions that Digital Twin systems promise depend entirely upon the interpretation and reliability of sensor data, which in many cases means requiring agents-in-the-loop to catch and resolve errors (e.g., discovering a broken sensor and turning off adaptive control).

Cognitive security factors related to agent perception and action, such as agents (i.e. humans and digital) engaging with the model as a canonical, unquestionable representation of the state of the system can result in catastrophe. For example, consider the Chernobyl Disaster,[3] Lion Air Flight 610 crash, or Ethiopian Airlines Flight 302 crash,[4] each of which were contributed to by variants of model-specification-reality gaps.

Part of the value of Digital Twins resides in their facilitation of measurement of certain aspects of physical systems *for purposes other than system control*, such as regulation or monitoring of output and certifying estimates of certain aspects of operations (e.g., carbon emissions). **The existence of an extrinsic incentive related to a measurement about a system attribute creates a perverse incentive to target the measurement instead of the attribute** – a phenomenon which is generalized through the lenses of Goodhart's Law[5] and Campbell's Law.[6] **There have already been multiple scandals related to the manipulation of digital representations of physical processes** in the interest of meeting regulatory criteria or qualifying for subsidies, such as targeted parameterization of emission-measuring software to give different results under laboratory conditions as opposed to actual driving conditions.

- **Recommendation 3.1:** R&D Activity should proactively address (i) cognitive security (i.e. human factors and ergonomics problems related to digital and human agent perception of Digital Twins and related interfaces), (ii) cyber- and network-security, and (iii) threats to public safety, liberty, privacy, and general welfare.

- **Recommendation 3.2:** R&D Activity should prioritize exploration of dual-use research of concern, safety standards, and methods of data-sharing and process verification.

- **Recommendation 3.3:** R&D Activity should address mechanism, market, and institutional design factors related to standards and requirements incentives for Digital Twin design and use in industry, commerce, and related regulatory functions.

[3] The Chernobyl Disaster was caused and exacerbated by a wide variety of issues, among which were lack of clarity in technical specification (expectations of reactor dynamics were based on incomplete information) and a misrepresentation by the on-site supervisory control and data acquisition (SCADA).

[4] Both crashes were related to control, modeling, and sensor errors.

[5] Goodhart's Law can be stated as either (i) *When a measure becomes a target, it ceases to be a good measure* or (ii) *Any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes.*

[6] Campbell's Law can be stated as: *Quantitative measures used in decision making processes subject those processes to pressures which corrupt and distort the system factors intended to be quantified*.

Summary Recommendations and Overview

Our summary recommendations, based on the background provided in the preceding section and categorized by the areas of interest listed in the request for comment, are as follows:

- **Data Management and Uncertainty Quantification:**
    - Common data and reference management should be addressed as critical infrastructure for Digital Twin ecosystems.
    - R&D Activity should include investigation into new approaches to reference management and data sharing that address instability in data location, schema, ontology, and underlying data; and should prioritize work which does so without requiring centralization, single-sources of truth, or total agreement from all parties in order to interact.
    - R&D Activity should prioritize approaches which treat data values, system state, and system specification as fuzzy or intrinsically uncertain.
- **Workforce, Professionalization, Standardization, and Responsible Use:**
    - Research portfolios should be complemented by convenings, forums, and other forms of multi-sector, interdisciplinary community engagements to facilitate the formation of relevant professional and trade associations.
    - R&D Activity should be complemented by model-based systems engineering education and professionalization activities for researchers and engineers, and should investigate best practices related to education on the topic.
    - R&D Activity should prioritize common vulnerabilities and exploits (CVE) and other community pattern-finding and data-basing initiatives in order to help researchers avoid pitching or performing redundant work.
- **Stability, Sustainability, and Security:**
    - Cognitive security (i.e. human factors and ergonomics) related to perception and use of Digital Twin and related interfaces should be treated as equally important to cyber- and network-security within a broader security and assurance research portfolio.
    - Risks related to dual-use research of concern, public safety, liberty, privacy, and general welfare should be addressed proactively in the research agenda, and potential for perverse incentives in business use-cases related to design and quantification of Digital Twins can be explored through mechanism, market, and institution design.
    - Research related to implementation should require consideration of initial provisioning and assured provisioning related to maintenance requirements.