HSOAC | HOMELAND SECURITY OPERATIONAL ANALYSIS CENTER
AN FFRDC OPERATED BY RAND UNDER CONTRACT WITH DHS

RESEARCH REPORT

DANIEL M. GERSTEIN

EMERGING TECHNOLOGY AND RISK ANALYSIS

Digital Personhood

This report describes the implications of digital personhood in all its potential forms—from the digital proof of the existence of and representation for a physical human, to the use of digital twins that provide a virtual model of a physical object (in this case, a human), to a virtual (or synthetic) person. The report also describes the emerging technologies our team examined that make up these various forms of personhood and the risks that this technology could present for U.S. Department of Homeland Security (DHS) missions. The material in this report builds on "Western philosophical thought regarding the existential concept of 'the self'" and, for the purposes of this report, digital personhood.[1]

In the previous report, we concluded that artificial intelligence (AI) and other digital technologies have fueled AI research and development (R&D) and brought the possibilities—and the potential dangers—of digital personhood to the forefront. These technologies include increasingly available data sources, high-performance computing assets, semiconductor development and manufacturing, robotics, machine learning (ML), natural language processing, and generative AI. These technologies are enabled by converging technologies such as cyber, big data solutions, and the Internet of Things (IoT), to name a few. As the AI systems and enabling technologies have continued to mature and become more powerful, the uses of AI have expanded, the technology has become more readily available, the number of applications for the technology has expanded, and (in some cases) the technology has already surpassed human performance—which is to say it has achieved artificial general intelligence (AGI), albeit in a very limited sense.[2]

The concept of personhood has been debated for centuries by Western philosophers, yet the definition remains open to debate. The World Economic Forum (WEF) alludes to the importance of the concept of personhood in discussing the Fourth Industrial Revolution (FIR), which it calls "a new chapter in human development" that is "merging the physical, digital and biological worlds in ways that create both huge promise and potential peril," all while ushering in "a fundamental change in the way we live, work and relate to one another."[3] The WEF further highlights that "even what it means to be human" should be reconsidered, which would involve examining the legal, ethical, and societal meaning of the term *personhood*.

In his 2019 book, *A Theory of Legal Personhood*, author Visa A. J. Kurki says that his purpose in writing the book was "to provide a new, less

KEY FINDINGS

- The concept of personhood has been debated for centuries by Western philosophers, yet the definition remains open to debate. In the United States, the term *personhood* has been used inconsistently in the legal field and has often been used as a method of control to give and take rights.
- The maturing over the next decade of identity, credential, and access management (ICAM) technologies and of artificial intelligence—along with support of Fourth Industrial Revolution (FIR) capabilities (e.g., cyber, big data, and the Internet of Things) and tools—provides opportunities and challenges regarding the development of digital personages that could be created for licit or illicit purposes. The technologies will improve, but so will the countermeasures.
- Progress toward achieving and conferring digital personhood in nonhuman entities enabled by FIR technologies will lead to challenging questions about managing the risks of creating what some refer to as *digital persons*.
- Although FIR technologies are still developing, legal mechanisms to address them remain in the early stages of implementation, and no comprehensive legal frameworks are in place. Additionally, we assess that resources—funding, computing capabilities and facilities, or human factors—do not present significant barriers to entry for developing and proliferating digital personhood technology.
- Securing key data and algorithms that undergird these systems will likely remain a challenge and become a vulnerability. A future digital personhood environment with digital twins and nonbiological intelligences will only exacerbate the already challenging issues associated with registering, issuing, using, and managing ICAM systems. Additionally, we assess that digital personhood control measures are likely to move slowly because balancing protections, risk, stakeholder prerogatives, and innovation will take time.

confused, and less confusing way of understanding the conceptual scheme of legal personhood," which highlights the ambiguity embedded in the terminology.[4] As yet another example of the debate about personhood, Freya Blackmore, writing for the London School of Economics, concludes that the question on what "constitutes the self has proved to be an unending task rife with disagreement" while later identifying four "theoretical frameworks" for defining *personhood*: (1) a genealogical one that posits the sole criterion as "having human DNA"; (2) one positing five requirements that define personhood: "consciousness, reasoning, self-motivated activity, capacity to communicate, and self-awareness"; (3) a social criterion that requires that "others consider them to be a person"; and (4) a gradient theory that "considers personhood as a continuous rather than binary category," implying that "some individuals have more personhood than others."[5]

U.S. federal law defines a *person* in 18 U.S.C. § 2510(6) "to mean any individual person as well as natural and legal entities. It specifically includes United States and state agents. According to the legislative history, '(o)nly the governmental units themselves are excluded.'"[6] To further elaborate (and perhaps muddy the waters), *Black's Law Dictionary* defines a person or someone having legal personhood as "any being whom the law regards as capable of rights and duties" but further offers that a person is "a human being (i.e., natural person), though by statute the term may include a firm, labor organizations, partnerships, associations, corporations, legal representatives, trustees, trustees in bankruptcy, or receivers."[7]

Identity management technologies (IMTs) constitute important tools and capabilities to be considered regarding the establishment of proof of existence, identity, and personhood. In delineating U.S. IMT, we considered three periods. The first consists of the use of compulsory birth certificates (1853), driver's licenses in Missouri and Massachusetts (and later throughout the United States) (1903), and Social Security numbers (1935). The second period covers the use of early digital IMT, with the first digital identities and passwords (1960), the commercial internet (1990), development of an identity and access management (IAM) stack (2000), managed services for IAM (2006), identity as a service cloud (2010), centralized identity disruption (2014), decentralized "bring your own" IAM (2016), and ledger-based IAM and self-sovereign identity (2020). Finally, the third emerging period involves full digital government (2028) and quantum computing IAM (2030). With the maturing of these IMT systems, an important addition has been the use of credentialing, which led to the new phrase *identity, credential, and*

access management (ICAM).[8] This continuum of identity management and personhood is useful in considering how these early mechanisms—which still form the basis for establishing personhood in the digital age—were not intentionally designed for this purpose and might need to be changed in the future.[9]

Advances in FIR ICAM technologies coupled with evolving interpretations of personhood (and digital personhood) have the potential to challenge societies, governments, the private sector, and individuals. Establishing digital personhood could present significant risks with the incorporation of mature FIR technologies that demonstrate AGI—that theoretical point (often called the *singularity*)[10] at which AI would have a "human level of cognitive function, including the ability to self-teach."[11] This report focuses on the legacy and emerging ICAM tools and capabilities that will be in use over the next decade.

For this analysis, we consider four attributes, divided into two categories, in assessing the technology: technology availability (T_{AV}), which is the first attribute and category, and risks and scenarios (R_S), which we divided into threats, vulnerabilities, and consequences. The risks and scenarios considered in this analysis pertain to emerging ICAM technologies that support establishing digital personhood. The R_S have been provided by the study sponsors from the DHS Science and Technology Directorate and the Office of Policy. We compared these four attributes across short, medium, and long terms (Figure 1).

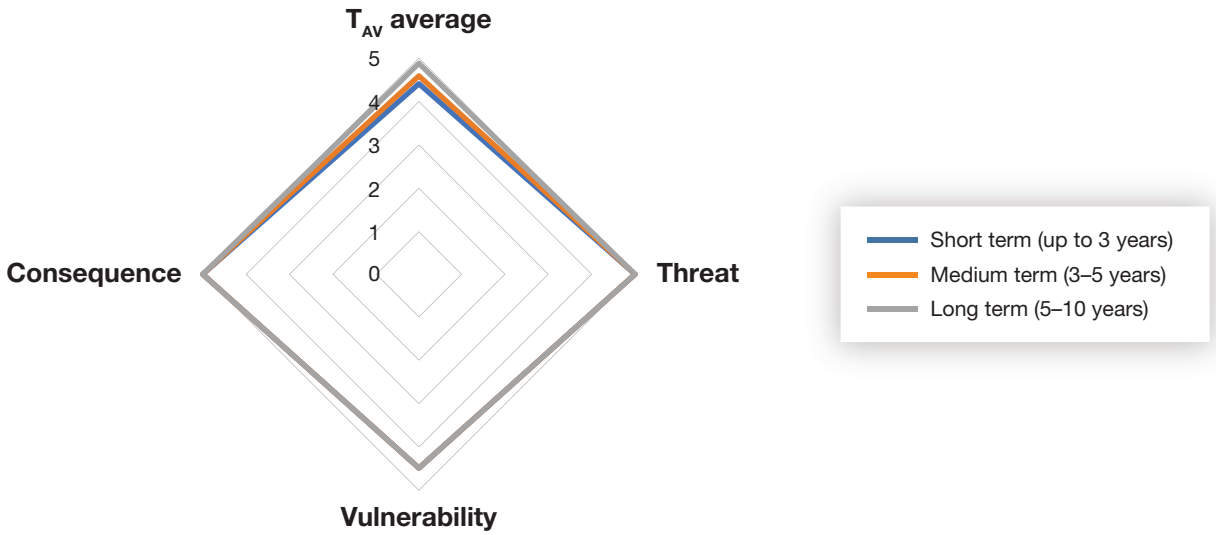
TECHNOLOGY DESCRIPTION AND SCENARIOS FOR CONSIDERATION

This report, written during the early stages of the FIR, focuses on emerging technologies for establishing digital personhood. We expect that, as humankind continues to learn more about the potential uses and misuses of AI, more applications will be discovered, including ICAM for establishing personhood. The material in this report focuses on a ten-year time horizon, likely long before AI would achieve singularity, and the point at which, in addition to AI achieving human-level cognition, the "technological growth becomes uncontrollable and irreversible, resulting in unforeseeable changes to human civilization."[12] Furthermore, this report is not intended to provide a detailed history, suggest a single authoritative (or legal) definition, or capture the many use cases likely to be discovered for digital personhood applications, either for the benefit or to the detriment of humankind.

Given the centrality of AI to the FIR and specifically to digital personhood, we begin with a discussion of this scientific and technology field. AI is "a branch of computer science devoted to developing data processing systems that perform functions normally associated with human intelligence, such as reasoning, learning, and self-improvement."[13] No single agreed-on AI taxonomy exists because the key technologies, which often encompass several different fields, have each developed

FIGURE 1

RISK ASSESSMENT FOR DIGITAL PERSONHOOD



NOTE: The emerging technology risk assessment scale: 0 to <2 = low impact or not likely feasible, 2 to <4 = moderate impact or possible, and 4 to 5 = high impact or likely feasible.

specific taxonomies for AI technologies. At a more granular level, supporting tools—some might say *supporting AI technologies*—include large language models, neural networks, supervised and unsupervised training, natural language processing, and computational linguistics. Other enabling and converging FIR technologies—such as cyber, big data, and the IoT, to name a few—have also become inextricably linked to AI R&D. These lists and the components that directly relate to and support AI are likely to expand as use cases continue to be identified, AI technologies continue to evolve, and operational AI-systems are fielded.

It is useful to delineate three AI categories when assessing AI maturity. The global Millennium Project describes the three categories as follows: "Artificial narrow intelligence is often better and faster than humans in, for example, driving trucks, playing games, and medical diagnostics. Artificial general intelligence is hypothetical ability of an intelligent agent to understand or learn any intellectual task that a human being can. Artificial super intelligence sets its own goals independent of human awareness and understanding."[14] Today, only artificial narrow intelligence has been seen, with some experts arguing that several individual applications have demonstrated early AGI-like capabilities.[15] The R_S evaluation focused on the three areas discussed previously: threats, vulnerabilities, and consequences.

METHODOLOGY

The Homeland Security Operational Analysis Center (HSOAC) team developed a framework for assessing the risks of emerging technologies. The assessment consisted of an evaluation of the T_{AV} and potential R_S for which a technology could be used. The T_{AV} evaluation focused on five areas: science and technology maturity; use cases, demand, and market forces; resources committed; policy, legal, ethical, and regulatory impediments; and technology accessibility.[16]

The ratings for the T_{AV} and R_S categories ranged from 1 to 5, where 1 corresponds to many challenges and 5 to very few, if any, challenges. The five technology availability areas were averaged and used in the emerging technology risk assessment. To allow for comparisons between different emerging technologies in assessing consequences, we rated impacts according to likely affected level (national, regional, or local), potential mortality and morbidity, and likely economic and societal disruption. By averaging the threats, vulnerabilities, consequences, and T_{AV} average calculated previously, an emerging technology risk could be assessed as low (0 to <2), moderate (2 to <4), or high (4–5).

These assessments were repeated for three periods: short term (up to three years), medium term (three to five years), and long term (five to ten years). This allowed the study team to assess individually and collectively how the T_{AV} and R_S would be affected over time. The assessment considered how the threats, vulnerabilities, and consequences evolved and whether preparedness, mitigations, and response activities had been undertaken that could reduce the risk.

THE TECHNOLOGY AVAILABILITY ASSESSMENT

The T_{AV} assessment is conducted without regard to specific risks or scenarios. Those factors will be considered in the R_S assessment section of this analysis. This is done to isolate the effects of the changes in technology over the ten-year time frame.

For this T_{AV} assessment, we use the following definition for *digital personhood* but note that many of the concepts are continuing to evolve in this emerging field of study:

> Digital persona is a part of the individual identity that has been extended into the online sphere to which corresponds a digital unconscious structuring a digitally divided self. It has personal, social, institutional, legal, scientific and technological aspects that have to be reconsidered to allow for new ways of understanding and managing identity. However, the fragmentation of scientific analysis fails to explain what happens to the digital personae in an interdisciplinary way. This is reflected by the current lack of comprehensive framework, the tendency to develop fragmentary management tools and gaps in legal frameworks.[17]

For our purposes, we identified three categories for analysis: the digital ICAM for a physical human, the use of digital twins that provide a virtual model of a physical object (in this case a human), and a virtual (or synthetic) person.

Several considerations provide an interesting point of departure in this report's discussion of digital personhood. First, the many technologies that make up AI and ICAM capabilities and tools mature along different time scales, making it challenging to identify an exact technology readiness level for ICAM, AI, or FIR tools and capabilities. As a result, a more productive approach in examining these technologies' availability is to associate

the readiness levels of the individual technologies with the use cases that are likely to be developed. Second, the potential of FIR technologies (particularly AI) to become "uncontrollable and irreversible, resulting in unforeseeable changes to human civilization,"[18] implies that consideration of the development of the technologies and of the associated guardrails should be done simultaneously to ensure that this runaway condition does not occur. Third, developers, users, and even society should expect that ICAM, AI, and supporting FIR (e.g., data science, the cyber domain, and the IoT) capabilities and tools will have unforeseen vulnerabilities. We assess that these effects will be cumulative and will require continual scrutiny to mitigate potential vulnerabilities associated with digital personhood uses. Fourth, the proliferation of open-source software, publicly usable closed-source models, ease of use of the technology, and the ability of nontechnical people to modify and even generate new capabilities will result in the proliferation of digital personhood applications for an increasingly wide variety of licit and illicit purposes.

SCIENCE AND TECHNOLOGY MATURITY

This section focuses on the scientific and technological maturity of the digital personhood–related ICAM, AI, and supporting FIR capabilities and tools over the next decade. In doing so, the temptation could be to provide lists of current and next-generation authentication methods and consider their effects on digital personhood trends; however, that would fail to identify the dramatic changes in context that accompany these two lists. These changes will entail the changing policies, permissions, and protocols that undergird managed access and will allow for digital personhood to proliferate. (More on the changing context will be discussed in the "Use Case, Demand, and Market Forces" section.)

ICAM technologies involve use of documents (e.g., birth certificates and Social Security numbers) and passwords with augmentation from biometrics, hardware tokens, and multi-factor authentication. As highlighted previously, the very early personhood documents were "not intentionally designed" for digital purposes and therefore present challenges for use in digital personhood systems.[19] Despite this concern, the next-generation list would likely continue to rely on those early documents and build on technologies that are in use while adding behavioral biometrics, mobile authentication, and zero-trust architecture in the short term.[20] According to one industry source that specializes in identity verification, within the ten-year horizon, emerging ICAM technology would likely incorporate full digital government (2028) and quantum computing IAM (2030) solutions.[21] We assess that digital personhood applications would seek to incorporate increasing digital identity capabilities, privacy-enhancing technologies, and trust and safety concepts.[22]

AI and FIR technologies will be foundational for developing digital personhood capabilities. Over the past decade, the science and technology of AI have rapidly improved, providing the "capabilities and accessibility of generative AI that can use training data to generate content in the forms of text, images, audio, and videos."[23] These are many of the same tools that would be useful in creating digital persons. Applications include the use of generative AI in art, journalism, education, acting, politics, and even science, to name just a few. As an example, OpenAI's Sora "can create realistic and imaginative scenes from text instructions."[24] Although the visual output Sora displays often has anomalies, the high-quality images it creates are a harbinger of what is to come and how quickly the technology is maturing.

The FIR enabling and converging technologies (e.g., cyber, big data, and the IoT) are also maturing rapidly and contribute to expanding potential for digital personhood applications. These technologies often contribute to the development and proliferation of "realistic AI-generated fake content . . . by facilitating the dissemination of disinformation to a targeted audience and at scale by mali-

ICAM technologies involve use of documents (e.g., birth certificates and Social Security numbers) and passwords with augmentation from biometrics, hardware tokens, and multi-factor authentication.

cious stakeholders" through use of the IoT and web applications.[25] Such fake content can include digital persons, either portraying synthetic persons or actual persons' content that has been modified.

To demonstrate the rapid maturing of AI and FIR technologies and their potential use for digital personhood, ContextualAI assesses that handwriting analysis, speech recognition, image recognition, reading comprehension, and language understanding have all surpassed human performance while other tasks—such as common-sense completion, grade-school math, and code generation—have achieved approximately 85–98 percent of human performance.[26] Those that have already exceeded human performance could be said to have achieved an AGI-like capability in a single activity or field. In highlighting the progress that has been made and concerns about this proliferation, CNET's "AI Misinformation: How It Works and Ways to Spot It" begins with an ominous assessment: "Determining what's real online is getting more difficult as AI and deepfakes spread across social media platforms."[27]

Supporting digital technologies and the continued transformation of the World Wide Web to Web 3.0 will contribute to advances in and proliferation of digital personhood applications and concerns.[28] Recalling the transformation of the web is instructive in looking to the future. Web 1.0 provided "static, read-only" access for interfacing with content through "web browsers, HTML, HTTP and URL technology."[29] Web 2.0 allowed users to have "read-write interface . . . driven by mobile, social networks and cloud technology," on which data were highly centralized, with a "small group of big tech companies like Meta (previously Facebook), YouTube and Twitter [now X]" controlling the data.[30] In this version of the web, the interactions were optimized for human users. Web 3.0's innovation comes from the "digitization of assets via tokenization" and such key features as the semantic web, more-powerful AI, enhanced 3D graphics, new levels of connectivity, ubiquity through the proliferation of IoT devices, blockchain to protect and encrypt data, decentralization of data with peer-to-peer connectivity, and edge computing for data and apps across a variety of digital platforms.[31] The semantic web capability, in particular, sets the conditions for Web 3.0 that will be optimized for machine-to-machine communication by "[improving] the abilities of web technologies to generate, share and connect content through search and analysis by understanding the meaning of words rather than by keywords or numbers."[32] These supporting technologies would also have benefits for digital personhood applications. For example, Web 3.0 and the supporting technologies could create "a more open, secure, and user-centric internet experience" and facilitate "blockchain technology, decentralized protocols, and the principles of decentralization."[33]

Earlier, the concern was introduced about the how the risks associated with digital personhood technology would be cumulative. These ICAM, AI, and FIR capabilities and tools have risks associated with their use both individually and collectively. Although these risks will be discussed more comprehensively in the "Risks and Scenarios" section, the reader should bear in mind that the ICAM, AI, and FIR technologies should be subjected to validation and verification (V&V) to ensure the fidelity of the systems, models, and products. For example, the data sets, code, and algorithms used to develop and train AI applications should be subjected to V&V. Checks should be conducted prior to deployment of digital personhood systems—for AI systems, the U.S. Department of Defense identified that the systems adhere to five ethical principles: They must be responsible, equitable, traceable, reliable, and governable.[34] These types of V&V assessments are essential through the life cycle of the deployed digital personhood tools beginning with the initial deployment.

The maturing of ICAM, AI, and supporting FIR capabilities and tools over the next decade provide opportunities and challenges for the development of digital personages that could be created for licit or illicit purposes. It also signifies progress toward achieving and conferring digital personhood in nonhuman entities, which leads to challenging questions about managing the risks of creating digital persons. These will be addressed in the T_{AV} discussion in the "Policy, Legal, Ethical, and Regulatory Impediments" section.

USE CASE, DEMAND, AND MARKET FORCES

The use cases, demand for technologies, and market forces will likely contribute to the growing proliferation of digital personhood applications. We assess that these demand forces—which will be embedded in the policies, permissions, and protocols that undergird future managed access—will allow digital personhood applications to proliferate. Today such applications include video game avatars, but future uses could encompass "a new era of digital twins to design, simulate, operate, and optimize their products and production facilities."[35] The synergy among these use cases; demand for the technologies and market forces; and the ICAM, AI, and supporting FIR capabilities and tools will contribute to rapid and wide adoption of digital personhood technology. As new technologies mature, additional use cases and greater demand will likely result.

In looking to this digital personhood future, one account describes digital access as needing to be all-encompassing and dynamic to allow for managing access to the "entitlement level," "fueled by AI/ML," with "access granted on an as-needed or 'just in time' basis."[36] In this new environment, permissions should be "driven by policy—not roles—to determine if and when access is granted, to what degree, and within what time frame. . . . [E]nterprises need the ability to create a dynamic trust model that is context aware, with policy as the blueprint" and the ability to "manage and secure their identities at any speed, at any scale."[37]

One industry expert summarizes the transformation as follows:

> True "next-gen" identity security represents a seismic shift in the way organizations think about identities. Employee identities are no longer front and center, flanked instead by third-party users, smart devices, cloud applications, automated software, and dozens of other human and nonhuman identities.[38]

This new approach to identity management and ultimately to establishment of digital personhood should also be robust enough to handle nonhuman identities such as bots, databases, and applications that could also achieve digital personhood status, and the approach will leverage AI and ML. Some challenges associated with implementing systems that could handle supporting activities for digital personhood would involve integrating with existing or legacy systems, developing new architectures for managing data silos and fragmentation, and establishing new user adoption requirements.[39]

Some of the capabilities and tools already exist but would likely require updates to handle digital personhood issues more effectively. For example, the Cybersecurity and Infrastructure Security Agency is the proponent for the "Continuous Diagnostics and Mitigation (CDM) Program [which] provides a dynamic approach to fortifying the cybersecurity of government networks and systems" and "[delivers] cybersecurity tools, integration services, and dashboards that help participating agencies improve their security posture."[40] CDM works by "[confirming] potential risks, [alerting] affected agencies, and actively [tracking] mitigation."[41] The "CDM Program Overview" advertises that it provides "Identity and Access Management," which "[m]anages account/access/managed privileges (PRIV), trust determination for people granted access (TRUST), credentials and authentication (CRED), and security related training (BEHAVE)."[42]

As a result of these growing use cases, we assess that *proof of personhood* will take on greater importance and receive increased scrutiny as weightier issues allow and employ digital personhood identities for ICAM. Answering such questions as (1) "Who are you?" (for digital identity verification); (2) "Are you who you say you are?" (for digital authentication); and (3) "Are you human and unique?" (for establishing proof of personhood) will be essential in a digital personhood zero-trust system.[43] These questions will need to be addressed by continually employing ICAM, AI, and other FIR tools and capabilities in real time, especially given the likely changes to digital personhood concepts and use cases.

Some future benefits—if proof of personhood by employing one's digital identity could be established—might be preventing identity theft, improving digital security, and improving internet interactions. Some potential future use cases could involve voting, finance, and government—again, assuming that proof of personhood could be established. Despite the potential benefits, challenges could be faced in establishing proof of personhood on a global scale and in ensuring the security and privacy of personhood systems.[44]

This new approach to identity management and ultimately to establishment of digital personhood should also be robust enough to handle nonhuman identities such as bots, databases, and applications that could also achieve digital personhood status.

RESOURCES COMMITTED

AI presents an interesting dichotomy between the cost to develop and the cost to use the technology; this same dichotomy pertains to digital personhood. We assess that generative AI (which is a core technology in developing digital personhood capabilities and tools) has become ubiquitous and will continue to become more powerful as the FIR technologies mature.[45] We further assess that these trends are likely to continue. As a result, resources—funding, computing capabilities, facilities, and human factors—do not present significant barriers to entry for the developing and proliferating digital personhood technology.

In assessing the potential for the addressable market, *Forbes* offers, "Globally, businesses across every sector and industry imaginable are striving to digitize their business; that is where companies in the digital identity space stand to benefit from the strong demand requiring digital transformation in operations."[46] The article highlights the drivers of the digital identity market as saving costs, removing friction, and "meeting ever-increasing data privacy regulations and security requirements."[47] The *Forbes* article further identifies key market subsectors that will be important in understanding market valuation: "biometrics, encryption technologies, payment wallets and self-sovereign identity."[48]

Specific ICAM technologies and supporting FIR technologies also represent a large market, but attempting to parse it into the specific funding associated with digital personhood is not possible given the ubiquity and lack of precision that would come from trying to allocate these digital capabilities into a single area, such as digital personhood. Even attempting to identify the myriad use cases to assess the total addressable market for digital personhood is extremely complex and fraught with uncertainty. Obvious uses would be in customer service, health care, education, and entertainment, but this short list does not address how the technology could be used in a far wider variety of applications—such as voting, finance, and government—that rely on surety in establishing personhood. Such a list also does not account for the wide variety and potential numbers of use cases from purely digital persons (i.e., nonhuman intelligences or synthetic persons), depending on changes in legal status or growth in opportunities through regulatory changes. These possible changes are addressed in the next section.

POLICY, LEGAL, ETHICAL, AND REGULATORY IMPEDIMENTS

In considering policy, legal, ethical, and regulatory impediments (or laws, guidance, directives, etc.), an important point of departure is recognition that little direct discussion of digital personhood exists in many of the government and academic writings. An important reason for this omission is that digital personhood capabilities and tools are made up of numerous converging ICAM, AI, and other FIR technologies. The result is that recent executive orders and policies often relate to digital personhood but do not directly address the topic.

For example, recent policy documents have described the growing global and national concerns about AI and recommendations for managing the technology. Executive Order (EO) 13859 in February 2019 directed federal agencies to ensure that the nation maintains its leadership position in AI; [49] the previously cited U.S. Department of Defense document from February 2020 lists five ethical principles to be considered in AI development and use;[50] EO 13960 in December 2020 directed federal agencies to inventory their AI use cases and share their inventories with other government agencies and the public;[51] and the January 2023 "voluntary use" National Institute of Standards and Technology (NIST) AI Risk Management Framework (AI RMF) issued a call "to improve the ability to incorporate trustworthiness considerations into the design, development, use, and evaluation of AI products, services, and systems."[52] In October 2023, President Joe Biden issued EO 14110 titled "Safe, Secure, and Trustworthy Artificial Intelligence."[53] A draft NIST report from April 2024 "examines the existing standards, tools, methods, and practices, as well as the potential development of further science-backed standards and techniques," for authenticating synthetic content, such as "images, text, audio, videos, as well as multimodal content," which have applicability to digital personhood applications and concerns.[54]

Likewise, the international community has expressed concerns about policy, legal, ethical, and regulatory issues regarding a variety of FIR technologies but has not directly addressed digital personhood with the same comprehensiveness. For example, the United Nations has been examining FIR technologies, most notably on AI and data privacy.[55] The European Union is arguably the most progressive governing body in developing policy, legal, ethical, and regulatory measures regarding AI, given its 2018 general data protection regulation and recent proposed European law on AI.[56]

Even existing legislation pertains to only a part of the digital personhood issue. For example, the Senate Committee on Homeland Security and Governmental Affairs drafted the Improving Digital Identity Act of 2023 "to recommend secure methods and coordinate efforts for digital identity verification" and to call for establishing "a task force within the Executive Office of the President to coordinate a government-wide effort for promoting digital identity credentials (e.g., electronic driver's licenses and birth certificates) for use in the public and private sectors."[57] Specifically, the legislation addresses digital verification of humans but does not address the status of digital twins or nonhuman persons—an issue that will be important to address in the future.

The legal status of nonhuman persons is an issue of such importance that it will encompass the remainder of the discussion in this section. Worldcoin, a for-profit cryptocurrency company, identified two potential concerns that relate to digital personhood: (1) "[p]rotecting against sybil attacks, or online attacks from multiple pseudonymous identities generated by a single attacker" and (2) "[p]reventing the dissemination of realistic, AI-generated content intended to deceive or spread disinformation at scale."[58] A third set of concerns relates to legal rights that could be afforded to nonhumans. This concern will become more pressing should the point of sentience be reached through achieving AGI in cases for which nonbiological intelligences (NBIs) have "human mental capacities."[59] As the author of a *Minnesota Journal of Law, Science & Technology* article argues, "Protection of NBIs, equivalent to protection of human research subjects, should be preemptively implemented to prevent injustice and potential grave harm to them."[60]

The history of personhood in the United States can be traced to the Constitution and Bill of Rights. The 1787 "'3/5ths compromise' extended personhood to 3/5ths of the enslaved population."[61] Such court decisions as the 1886 one in *Santa Clara County v. Southern Pacific Railroad* expanded personhood further, "[declaring] that the protections of the 14th Amendment covered both corporations and 'natural persons,' giving birth to the notion of 'corporate personhood.'"[62] Over time, the allowances have been expanded: "As early as 1906, and then again in 1978, the Supreme Court held that corporations were entitled to assert Fourth Amendment rights against the warrantless search of commercial premises. And in 2010, the Court also made it clear that corporations even have rights under the First Amendment. . . . In 2014, the Supreme Court held that corporations also have the right to the free exercise of religion," essentially providing a form of digital personhood to nonhumans.[63]

A 2024 *Yale Law Journal* article highlights that the U.S. history of this term *personhood* has been "mutable" and "a powerful tool that humans have used as a lever to control the giving and taking of legal rights. But even within the definition of who or what constitutes a 'person,' there are gradations." The article further warns that "[w]hen human society is confronted with sentient AI, we will need to decide whether it has any legal status at all."[64]

Taking this to a logical conclusion, one legal expert warns that such issues as labor rights, ethics, and involuntary servitude or slavery will all need to be considered as sentient AI becomes a reality.[65] Another issue to be confronted is "when certain AI no longer resides in a single place [and] it will be 'distributed,'"[66] which will lead to various use cases across the globe (and perhaps even within different U.S. states), likely employing different definitions of personhood, operating simultaneously within and across different nations (and in different U.S. states).

Understanding such corporate loopholes is instructive for demonstrating both how establishing corporate personhood has been done and how readily such personhood could be done in the future. One legal scholar describes the process:

> Giving AIs rights similar to humans involves a technical lawyerly maneuver. It starts with one person setting up two limited liability companies

The international community has expressed concerns about policy, legal, ethical, and regulatory issues regarding a variety of FIR technologies but has not directly addressed digital personhood.

> [LLCs] and turning over control of each company to a separate autonomous or artificially intelligent system. Then the person would add each company as a member of the other LLC. In the last step, the person would withdraw from both LLCs, leaving each LLC—a corporate entity with legal personhood—governed only by the other's AI system.[67]

As Hon. Katherine B. Forrest (fmr.) mentions in a *Yale Law Journal* article, "The evolution of corporate legal personhood has taught us that when humans find it useful to bestow rights, a lack of human-like sentience or human-type awareness is not a precondition."[68]

In framing the issue, several scenarios come to mind. First, there is the question of what happens when an AI-enabled entity—"acting independent from a human—itself causes harm."[69] Where should the challenge (and the responsibility) lie? Asked another way, "[W]ho is the agent with regard to a particular action: AI or its designer, coder, licensor, or licensee?"[70] These are no longer hypothetical questions; "there are already cases in which courts are asked to allocate responsibility for harm caused by AI tools—for instance, tools that harm humans, facilitate forms of discrimination, violate due process, or are alleged to be instrumentalities of price fixing."[71] Model drift could also occur if the AI is used outside its original purpose and a harm occurs.[72]

These questions would become more complex should someone (either a human or digital AI-enabled entity) modify the code (or the underlying logic based on new data inputs), which could easily occur if the AI were permitted to conducted unsupervised learning. In such a case, "[i]f the AI tool is considered a legal agent of the entity, this entity would typically bear responsibility under agency principles."[73] However, given the myriad systems that interact in digital personhood capabilities and tools, establishing causality with legal certainty might not be possible. All this is to say that each of these questions will depend on the facts of the case—which might not be fully understood if they involve ICAM, AI, and other FIR capabilities and tools. As Forrest offers, "Humans are, in effect, creating tools that have a toxic-tort-like potential to enter the world and do damage in ways that we cannot yet imagine or understand, but for which our act of creation confers personal responsibility."[74]

This highlights the breakneck speed with which governments, the private sector, society, and individuals are embracing these new digital personhood capabilities and tools, many before they have had proper validation and verification. The early ChatGPT 4 rollout in March 2023 provides clear warning of the challenges of rolling out untested (or at least not properly tested) large language models. The initial rollout highlighted that there were accuracy and consistency problems, forcing developers to quickly make changes to respond to the issues. We assess that this cycle—development, deployment, identification of shortcomings and other areas of potential use, and rapid updating of AI systems—will likely be a feature of many FIR technologies, including ICAM and AI.

The American Bar Association summarizes the state of policy, legal, ethical, and regulatory measures regarding FIR technology in the United States as "still in [the] early stages" and having "no comprehensive federal legislation dedicated solely to AI regulation." The organization notes, however, that "there are existing laws and regulations that touch upon certain aspects of AI, such as privacy, security and anti-discrimination."[75] Congress has also begun to deliberate on AI issues; however, these efforts remain in the early stages with concrete measures likely to come well in the future. Given the state of FIR technologies (such as AI and cyber legislation and social media deliberations), we assess that digital personhood control measures are likely to move slowly because balancing protections, risk, stakeholder prerogatives, and innovation will take time.

Industry has also talked about the need for FIR technology governance in several key areas (including AI, cybersecurity, big data, and the IoT), highlighting the need for assistance in development of laws, policies, and regulations. Despite these calls for moderation and governance, R&D continues, and new capabilities are unveiled at a rapid pace as new technologies continue to mature. It is noteworthy that other industries have established policies, regulations, standards, norms, and ethics, either nationally or internationally.[76] For example, export controls serve to restrict the proliferation of key technologies. But many ICAM, AI, and other FIR technologies—which would be essential for digital personhood applications—are already available for public use, so export controls might have little practical effect. As a result, we assess that, despite all the attempts to manage and control the proliferation of digital personhood FIR technology, there are still relatively few impediments that would hinder further development of these systems.

In concluding this section, it is worth highlighting that although the policy, ethical, and regulatory challenges are of great importance, the remedies for breaches in these areas generally come from legal enforcement mechanisms. More on these topics will be addressed in the "Risk Assessment" section.

TECHNOLOGY ACCESSIBILITY

Regarding technology accessibility, we assess that FIR technologies that would be useful for licit or illicit digital personhood applications have become increasingly available and accessible to the public. With the unveiling of ChatGPT 4 in March 2023, awareness of generative AI models has increased dramatically, and a competition between platforms has arisen. Many companies and developers are offering free availability for trials or for limited uses.[77] Despite sometimes generating results that incorporate bias, stereotypes, hallucinations, and false information into generated outputs, a steady increase in generative AI capabilities is occurring, and developers are continuing to refine models, reduce these occurrences, and release updated models.[78]

Arati Prabhakar's 2023 speech at the Carnegie Endowment illustrates the meteoric increase in availability and decreasing costs associated with generative AI. In less than a year, the state of play went from technology costing companies more than $100 million to develop, to companies being able to develop models that were even more powerful for under $1 million, to narrow models being customizable based on open-source code for the cost of "a few hours of effort on commonly available hardware that is off the shelf that's $100—not $100 million, but $100."[79]

We assess that it is unlikely that average citizens would be able to develop the next generation of advanced AI technology, given the needs for sophisticated semiconductors and high-performance computing, as well as the volumes of data that would be needed in conducting supervised and unsupervised training.[80] However, access to these technologies is no longer required given the open-source software capabilities that are already available.

Finally, we assess that digital personhood capabilities and tools in ICAM, AI, and other FIR technologies (including the skills to operate commercial applications) are proliferating rapidly. The technology for developing digital personhood capabilities continues to mature and be available to the public in open-source voice and video generators, which have already been used in making convincing deepfakes and are continuing to become both more efficient and effective at making deepfakes that are harder to identify.[81]

OVERALL TECHNOLOGY AVAILABILITY

Digital personhood capabilities and tools in ICAM, AI, and other FIR technologies will continue to become integrated across new markets and use cases, leading to increased T_{AV}, especially for publicly available applications such as those used in generating false information. We assess that technologies will be converging and self-reinforcing, increasing the speed of development of digital personhood technologies with even greater capabilities, more sophistication, and increased accuracy. Once these AI systems are in use, problematic changes to their inputs—through supervised training with erroneous data or unsupervised training, which also makes traceability more challenging—are also likely to continue increasing.

In evaluating the individual T_{AV} categories, we assess that **science and technology maturity** as measured by the deployment of these ICAM, AI, and other FIR technologies are already high and should be expected to increase. Despite the relative maturity of the technologies, we expect that, as new capabilities are made available, the development, deployment, and update cycle discussed previously will continue to speed up in the rush to move new systems and capabili-

It is unlikely that average citizens would be able to develop the next generation of advanced AI technology, given the needs for sophisticated semiconductors and high-performance computing, as well as the volumes of data that would be needed in conducting supervised and unsupervised training.

ties to market. Despite the convergence and synergy of the FIR technologies, we assess that AI will be the pacing technology for digital personhood technology maturity. As digital personhood applications continue to be employed, additional **use cases, demand, and market forces** will be identified, and the technology will become more readily available over the ten-year time frame under study. New digital personhood applications are likely to be identified for use in more sectors and applications. ICAM technology protections will likely become more prevalent to better secure supporting hardware, software, algorithms, and data. In terms of both **science and technology maturity** and **use cases, demand, and market forces**, we assess that the ChatGPT 4 experience will be repeated with other breakthroughs in the FIR technologies, including those important for digital personhood applications. This will create a constant need to update not only tools and capabilities but also countermeasures to improve system security, performance, and functionality.

We assess that the **resource** dichotomy will remain. Although developing the initial complex technologies will be outside the capability of most users who might seek to use many of the FIR technologies, the low cost for use of the technology will make it available for a variety of licit and illicit purposes. Furthermore, modifying open-source software generative AI code or producing false information from publicly available models will be well within the reach of nefarious actors, requiring relatively few resources in terms of dollars or time.

Policy, legal, ethical, and regulatory impediments will not present significant barriers to the employment of digital personhood technologies. Much of the work to date has been to establish guardrails using **policy, legal, ethical, and regulatory** approaches that have focused on the individual FIR technologies. These guardrails are important, but they will not be sufficient to address the likely future digital personhood issues. While we assess that the **policy, ethical, and regulatory** challenges are of great importance, the remedies for breaches in these areas generally come from **legal** enforcement mechanisms. Thus, the legal advances in digital personhood will be imperative and might need to take precedence. Given the ratings for the other four elements of T_{AV}, we assess the final category of **technology accessibility** is high and will remain so through the end of the ten-year period under consideration.

Table 1 provides an assessment for the potential for AI technology availability in the short term, medium term, and long term. To reiterate, this T_{AV} assessment does not necessarily indicate that the technology will be used for illicit purposes—this will be considered in the R_S section. However, it does indicate that T_{AV} barriers have been considerably lowered, and the potential for employing ICAM, AI, and FIR capabilities and tools for illicit purposes is growing.

THE RISK ASSESSMENT

Our risk assessment focuses on the threats, vulnerabilities, and consequences of illicit uses of digital personhood technologies. As highlighted earlier, AI is one of the drivers of this technology, particularly the use of generative AI for creating and disseminating false information (including digital persons).[82] Our analysis builds on the previous T_{AV} section, which concluded with the observation that ICAM, AI, and other FIR technologies have reached a level of maturity at which they are ubiquitous and used frequently for a variety of licit and illicit purposes but are also far from reaching full maturity—which, for AI, could be considered achieving AGI. We expect that this

TABLE 1

TECHNOLOGY AVAILABILITY FOR DIGITAL PERSONHOOD

Scenario	Science and Technology Maturity	Use Cases, Demand, and Market Forces	Resources Committed	Policy, Legal, Ethical, and Regulatory Impediments	Technology Accessibility	T_{AV} Average
Short term (up to three years)	4.0	4.0	4.0	5.0	5.0	4.4
Medium term (three to five years)	4.5	4.5	4.5	4.5	5.0	4.6
Long term (five to ten years)	5.0	5.0	5.0	4.5	5.0	4.9

NOTE: The emerging technology risk assessment scale is 0 to <2 = low impact or not likely feasible, 2 to <4 = moderate impact or possible, and 4 to 5 = high impact or likely feasible.

level of maturity would not occur during the study window under consideration.

Technological advances coupled with evolving interpretations of personhood (and digital personhood) have the potential to challenge societies, international and national governments, the private sector, and individuals. With the incorporation of mature FIR technologies that demonstrate AGI—that theoretical point (often called the *singularity*)[83] at which the AI would have a "human level of cognitive function, including the ability to self-teach"[84]—digital personhood issues could present significant risks. That level of technological maturity is very unlikely to occur during the ten-year period of the study.

In assessing the risks, it should be stressed that this assessment does not represent a comprehensive discussion of all the potential risks associated with digital personhood. Rather, for each of the three areas—threats, vulnerabilities, and consequences—key issues and potential mitigations are highlighted.

THREAT

Digital personhood applications—from the digital identity for a physical human, to the use of digital twins that provide a virtual model of a physical object (in this case, a human), to a virtual (or synthetic) person—will present threats to ICAM applications. The threats could come from a variety of state and nonstate actors using this technology for a variety of purposes. For example, criminals, terrorists, and hackers could use fraud or identity theft to gain access to networks to seek financial gain or perpetrate terrorist acts. This could result in a loss of confidence in those institutions that suffer the attack. Digital twins acting autonomously on behalf of their human representative or "the 'person' closest in the chain of causation" could exceed their authorities,[85] causing a dangerous event to occur, or they could commit to or perpetrate an ill-advised or illegal activity. Through generative AI unsupervised learning, a virtual (or synthetic) person could act on erroneous logic and cause loss of human life.

The threat vectors for targeting digital personhood technologies include the ICAM, AI, or supporting FIR technologies employed in establishing personhood, identity management, credentialing, and access management. The targets could involve the hardware, software, algorithms, protocols, and data used to establish personhood of a human, create a digital avatar or twin of a human to allow that entity to act on a human's behalf, or create a digital nonhuman entity. Each of these personhood cases has different policy, legal, ethical, and regulatory implications and potentially grant different permissions to the users, be they human or nonhuman entities.

We assess that, in the future, there will be far more reliance on digital means for establishing personhood, which will increase the scope, scale, and complexity of the interactions. *Scope* refers to the number of applications into which digital personhood systems are incorporated, *scale* refers to the size of the population affected and the speed of the interactions, and *complexity* refers to humans' ability to understand and evaluate the outcomes of digitally enabled procedures for establishing personhood. These increases in scope, scale, and complexity are also likely to increase the potential attack surface for threat actors.

The contextual changes associated with widespread use of digital personhood also have threat vector implications. These changes require digital access to be all-encompassing and dynamic to allow for managing access to the "entitlement level . . . [with] access granted on an as-needed or 'just in time' basis" and "fueled by AI/ML."[86] These threats will be increasingly driven by the speed of machine-to-machine interactions inherent in Web 3.0, and these interactions will be more frequent. A report by the Europol Innovation Lab spurred an observation at an international convention that, "We are fast approaching a 'tipping point' when non-human generated content will vastly outnumber the amount of human-generated content."[87] The Europol Innovation Lab has assessed that "90 percent of Online Content will be AI-generated by 2026."[88] This nonhuman-generated content will undoubtedly include data on digital personhood for use in ICAM applications.

In this new environment, permissions will need to be "driven by [constantly adapting] policy—not roles—to determine if and when access is granted, to what degree, and within what timeframe."[89] This means "enterprises need the ability to create a dynamic trust model that is context aware, with policy as the blueprint," all while having the ability to "manage and secure their identities at any speed, at any scale."[90] This new approach to identity management and, ultimately, to establishing digital personhood should also be robust enough to handle nonhuman identities—such as bots, databases, and applications that could also achieve digital personhood status—and must necessarily leverage AI and ML. In short, the very fabric of identity security is undergoing a profound transformation and will change the way organizations think about identities.[91]

This does not mean that the previous threats have been eliminated. Other threats that will need to be managed include insiders that could create insecurities and

backdoors in the ICAM systems that could be targeted for exploitation. Data from supervised or unsupervised learning could be adulterated in the development phase, resulting in incorrect training of the frontier models. Early releases of digital personhood tools without the appropriate guardrails could result in insecurities in the ICAM systems. Phishing attacks, malware, and social engineering will also remain threats, although digital ICAM capabilities and tools could be employed to potentially mitigate them.

All of this means that constant vigilance will be required to address potential threats that target digital personhood ICAM systems.

VULNERABILITY

Continuing with the theme outlined in the "Threat" section, the previous vulnerabilities that have confounded ICAM, AI, and other FIR technologies—such as cyber insecurities and other digital attempts to unlawfully or illegitimately establish one's digital identity—remain. The recent National Public Data (NPD) cyberattack provides an example of a massive data breach that might have compromised the personally identifiable information (PII) of up to 2.7 billion people.[92] As a DeepAI publication states, "[Digital] identity solutions—ID checking, biometrics, self-sovereign identity, and trust networks—all present flaws, leaving users vulnerable to exclusion, identity loss or theft, and coercion. These flaws may be insurmountable because digital identity is a cart pulling the horse."[93] It follows that such a large breach as the NPD could threaten personhood protections and create dangerous vulnerabilities.

Inherent in this concern is the human element, which has proven to be the largest and most intractable source of personhood vulnerabilities and remains essential to address for mitigating vulnerabilities associated with the proliferation of digital insecurities. To this point, the WEF found that 95 percent of cybersecurity incidents were caused by human error.[94]

Although AI can be used to identify vulnerabilities for the purpose of patching, the same technology could be used by illicit actors or adversaries to identify vulnerabilities with the intent of penetrating a network. As a result, FIR technologies, including generative AI used in targeting digital personhood applications, should be considered dual-use technology that can be used for licit and illicit purposes and has the potential to both decrease and increase vulnerabilities. One example is generative adversarial networks (GANs), which consist of two neural networks: the generator and discriminator. Generators "produce realistic fake data from a random seed," and discriminators learn to "distinguish the fake data from realistic data."[95] This technology has been employed to automate anomaly detection in the financial sector for identifying and mitigating fraud, to identify vulnerabilities in software to allow for faster patching of the insecurities, and as "AI-driven verification tools capable [of] identifying and flagging artificially generated content."[96] Research has also demonstrated how GANs have been effective at targeting individuals based on their social media presence for social engineering attacks capable of penetrating networks: The findings demonstrated that "threat actors can enhance the effectiveness of their phishing attacks by using AI as a malicious tool."[97] These examples highlight the tension—some might say the race—between the identification of false information to eliminate or mitigate it and the dissemination and amplification of false information for nefarious purposes.

Other emerging technologies have been identified as being potentially able to mitigate vulnerabilities. Blockchain technology might offer an approach for ensuring trust in one's data by ensuring that "only authorized parties can create or modify data, while anyone with the public key can verify its authenticity."[98] This issue of data trust has become central to protecting information and could generate new positions called "digital intermediaries or 'trust brokers' who specialize in verifying the authenticity of content."[99] Mitigation techniques should also include more-granular approaches, such as considering new methods of content moderation for AI-generated content. GANs are being used more frequently to identify deepfakes and other adulterations of the input data.[100] Some have also called for such measures as licensing AI models,[101] scrutinizing algorithmic decisionmaking to avoid biases and technical errors,[102] and pursuing content warnings and watermarking to highlight AI-generated content.[103]

Despite several previous congressional bills and executive branch policy documents relating to digital security and privacy (including cybersecurity, big data, and AI), these efforts remain works in progress. Implementation has been piecemeal and often voluntary (e.g., the NIST AI RMF). EO 14110 from October 2023 seems to be having more impact, directing a variety of outcomes that set the tone for AI development and fielding that seek to limit the vulnerabilities and the potential for adverse consequences.[104] For example, the document "directs generative AI developers to submit safety test results and federal agencies to establish guidelines on how they use AI."[105] Despite the promise of the EO, it does not "impose red lines against applications of the technology that seem to be prone to abuse, such as facial

recognition," which potentially have a role in ICAM.[106] The article's author makes a compelling case for the necessity of legislation that grants "federal authorities new powers and funding to regulate the tech industry."[107]

In a recent article, "a group of prominent academics and researchers are advocating for a new approach to digital ID and online identity in an age when artificial intelligence has effectively passed the Turing Test."[108] The approach argues for development of "personhood credentials," or "digital credentials that empower users to demonstrate that they are real people—not AIs—to online services, without disclosing any personal information."[109] Such an approach could provide benefits for the physical humans or real people by creating a *one person, one credential* environment but might not be as useful for granting authorized access for corporate personhood applications, digital twins, or machine-to-machine communications.

In critiquing the personhood credential concept, a September 2024 *Washington Post* article offers alternative views. One independent researcher is quoted as pointing out that "[i]t's worth asking why the onus should be on individuals to prove their humanity rather than on the AI companies to prevent their bots from impersonating humans, as some experts have suggested,"[110] and that "[a] lot of these schemes are based on the idea that society and individuals will have to change their behaviors based on the problems introduced by companies stuffing chatbots and large language models into everything rather than the companies doing more to release products that are safe."[111] Other concerns mentioned in the article are the "possibility of massive data-gathering" and that "[i]f artificial intelligence systems can convincingly impersonate humans[,] . . . presumably they could also hire humans to do their bidding."[112]

Finally, regarding the implications of digital personhood, a *Lawfare* article identifies three risks: "(1) the autonomy risk, which has its origin in stand-alone 'decisions' taken by the software agents, (2) the association risk, which is due to the close cooperation between people and software agents, and (3) the network risk that occurs when computer systems operate in close integration with other computer systems."[113] These concerns potentially pose the greatest challenges for private law, and the article cautions that "a new legal status for autonomous digital information systems" should be defined.[114]

In defining *digital personhood*, we identified three categories for analysis: the digital ICAM for a physical human, the use of digital twins that provide a virtual model of a physical object (in this case, a human), and a virtual (or synthetic) person. For the first category, the vulnerabilities reflect traditional approaches to ICAM and to legal standing. The challenges are well known and continue to present concerns and lead to insecurities. For the other two categories, however, the digital legal status has yet to be adequately defined, and more analysis would be useful to better understand the full measure of these potential vulnerabilities.[115]

CONSEQUENCE

The consequences of the proliferation of digital personhood applications—such as using false information, employing data and human resemblances to gain access to protected areas and processes, creating digital likenesses that portend to represent trusted people and messages, and establishing false realities—could undermine the foundations of society.

The consequences could result from the blurring of the line between fact and fiction, the inability to assess the veracity of the digital information that is being generated, and the increase in content being generated by nonhuman entities (perhaps by NBIs, should AGI become a reality). We further assess that the consequences associated with digital personhood use cases will be affected by the scope, scale, and complexity of the applications that are employed. We also assess that the Web 3.0 changes could mean orders of magnitude more attacks, with greater precision in targeting ICAM and digital personhood systems. For example, the Europol Innovation Lab assessment that "90 percent of online content will be AI-generated by 2026" will likely increase the potential for attacks against digital personhood applications, including ICAM.[116]

The WEF assesses these effects as AI potentially "amplifying manipulated and distorted information that could destabilize societies,"[117] including for use in target-

The consequences of the proliferation of digital personhood applications could undermine the foundations of society.

Industry also warns of the increasing potential for intrusions into digital personhood applications.

ing digital personhood-related activities. In its 2024 *Global Risks Report*, the WEF warns about the dangers of the "widespread adoption of generative AI to produce . . . synthetic content" that includes "deepfake videos, voice cloning and the production of counterfeit websites."[118] Similarly, the United Nations Development Programme warns of information pollution that is "affecting the citizens' capacity to make informed decisions,"[119] which could also affect digital personhood by employing false ICAM information, introducing biases, and causing a loss of trust.

Industry also warns of the increasing potential for intrusions into digital personhood applications. A Deloitte report highlights that these growing adverse consequences are because "access to sophisticated tools that employ machine learning, automated software bots, and natural language generation makes it easier for people with limited technical skills to create and spread manipulated information at scale."[120] This democratization (global spread) and deskilling (less skills required) of tools for generating false information and digital personhood deepfakes portend the increased frequency of penetrations by malicious actors, such as "nation-states, organized crime groups, companies, and even disgruntled customers," according to the Deloitte report.[121] In capturing potential consequences, the report lists "brand and reputational damage[,] . . . loss of public trust in organizations[,] . . . financial losses[,] . . . [and] difficulty identifying the individual or groups responsible."[122]

The consequences pertain differently to the three forms of digital personhood: (1) the digital proof of the existence of and representation for a physical human, (2) the use of digital twins that provide a virtual model of a physical object (in this case, a human), and (3) a virtual (or synthetic) person. It is useful to look specifically at each of the forms in considering the consequences. In conducting this assessment, we note that it is likely that the ICAM, AI, and supporting FIR technologies will likely increase significantly over the next decade; however, we do not anticipate that full AGI will be achieved during this time frame.

For targeting against the digital proof of the existence of and representation for a physical human, the scope, scale, and complexity of the FIR technologies will likely lead to a larger number of successful attacks in which penetrations will occur, human identities will be stolen or hijacked, and adverse consequences from illicit uses of a human's personhood will likely occur. The speed of Web 3.0 will also likely lead to more-effective attacks at a greatly reduced cost to the attacker. Social engineering attacks (such as phishing, spear phishing, baiting, and tailgating) are also likely to become more sophisticated, and the seam between the digitally enabled attacks and the human targets will become more challenging to defend. In turn, we assess that ICAM technologies will become more layered, making them more challenging to penetrate. However, we also expect continued persistence and innovation in the nature of the attacks, leading to a real-time competition to defend digital identities—today's efforts to prevent cyberattacks provides a useful analogy. Minimizing the cost of the intrusion in terms of loss of data, systems, dollars, property, and identity, for example, will be a necessary strategy to mitigate the effects of these attacks and penetrations.

Regarding the use of digital twins that provide a virtual model of a human, similarities exist with the physical human case with some important exceptions. Digital twins that act as virtual representatives authorized by their human twin would still be subjected to the same types of attacks intended to steal identities or gain unauthorized entry. However, limiting the legal authorities of the digital twin and ensuring that guardrails exist to prevent the digital twin from changing or exceeding those human-directed authorities would be essential in ensuring human agency in the use of a person's identity. To address this issue, mitigations, such as limiting the scope and duration of the activities in which the digital twin could represent the human twin, could be employed. Another concern would be the creation of a digital twin that does not represent the human twin—that is, it was created without the knowledge of the human twin. In this case, the digital twin could be acting on behalf of a third party for illicit or nefarious purposes. Such a rogue digital twin—one that has either exceeded its human-specified authority or been created by a third party (either a human or AI deepfake)—could create legal, ethical, or regulatory issues and lead to such questions as who is legally responsible for adverse outcomes by a digital twin. Regaining one's digital identity would likely prove challenging in such a scenario. Again, borrowing from a cybersecurity analogy,

identities might need to be untangled and rebuilt in the same way this is done by companies seeking to recover from a ransomware attack.

The third case of digital personhood, in which a virtual (or synthetic) person has been created using ICAM, AI, and FIR technologies, poses potentially catastrophic outcomes depending on the scope, scale, and complexity of the scenarios. As the technologies mature throughout the ten-year study horizon, we assess that the capabilities and tools will become more readily available and more advanced, and they will be deployed more frequently for a variety of licit and illicit purposes. They will present a risk to ICAM systems that seek to protect digital systems.

One of the major areas of concern are legal remedies for addressing legal, ethical, or regulatory concerns regarding digital personhood. The courts will need to establish precedent for which principles or rules are established based on cases involving identical or similar facts and issues.[123] In the meantime, the ICAM, AI, and FIR technologies continue to be developed, and the debates continue about policies, laws, ethics, and regulations surrounding digital personhood issues. These are very complex issues—as one expert questioned, "Who is the agent with regard to a particular action: AI or its designer, coder, licensor, or licensee?"[124] And what if either a human or digital AI-enabled entity modifies the code—who would be responsible? The answers to these questions become even less certain in the case of corporate personhood.[125]

Judging from experiences with cybersecurity and early generative AI, we assess that the digital personhood capabilities (ICAM, AI, and FIR technologies) are likely to conform to the cycle of development, deployment, identification of shortcomings and other areas of potential uses, and rapid updating of AI systems. Although the assessments are often mixed, we assess that attackers are likely to retain an advantage over defenders, at least in the short term. As BlackBerry assesses,

> While AI has the potential to help both attackers and defenders in the short term, threat actors may have an advantage, as they can quickly adopt new techniques without worrying about production readiness. However, over the long run, AI will likely equalize capabilities as defenders gain more context and build robust detection systems.[126]

The Cyentia Institute also assesses that the attacker-defender balance of power changes over time, with the attacker having the initial momentum, the defender gaining the advantage as vulnerabilities are discovered and addressed, and, finally, the defenders losing momentum over time.[127] Sophos News summarizes the competition as follows: "Slowed by multiple headwinds, defenders are falling behind while adversaries continue to accelerate. Organizations need to speed up the defender flywheel to enable them to pull ahead."[128]

Although there is hope on the horizon that ICAM, AI, and FIR technologies might be able to reverse these trends, *Forbes* ominously summarizes the landscape, stating, "The year 2023 saw a notable increase in cyberattacks, resulting in more than 343 million victims. Between 2021 and 2023, data breaches rose by 72 percent, surpassing the previous record."[129] With such a dire record related to cyberattacks, it is hard to project a rosy picture regarding digital personhood protections.

Finally, estimating consequences remains challenging because many insecurities are underreported.[130] ICAM capabilities and tools are moving from the current technologies that rely on documents and passwords with augmentation from biometrics, hardware tokens, and multi-factor authentication through next-generation technologies such as behavioral biometrics, mobile authentication, and zero-trust architecture.[131] When the emerging ICAM technology that is projected to bring "full digital government" in 2028 and "quantum computing and the encryption apocalypse" in 2030 actually arrives,[132] it is likely that ICAM insecurities and the overall consequences of attacks against ICAM systems can be reduced.

EMERGING TECHNOLOGY RISK ASSESSMENT

Digital personhood capabilities and tools for ICAM, AI, and other FIR technologies will continue to become more readily available; have greater capabilities with a reduced cost to obtain and use the technology; and have few policy, legal, ethical, and regulatory impediments. As a result, we assess that the $\mathbf{T}_{\mathbf{AV}}$, which is already high, will present increasing challenges throughout the ten-year period under consideration.

Regarding the **threat** landscape, we assess that the continued democratization and deskilling of the technology will mean significant proliferation to a wider array of actors seeking to use the personhood-related technologies for illicit or nefarious purposes. The greater reliance on Web 3.0 capabilities means that NBIs approaching AGI or having achieved AGI in a narrow subfield (e.g., ContextualAI, which we identified previously as having demonstrated AGI-like capacity in handwriting analysis, speech recognition, image recognition, reading comprehension, and language understanding) would likely increase the scope,

scale, and complexity of any illicit or nefarious actors seeking to employ personhood credentials.

The likely increase in threats is, in turn, likely to contribute to a corresponding increase in **vulnerability and consequences**. It is noteworthy that *optimal* use of false information is not required to be successful. Less-than-perfect deepfakes could still achieve the intended effect of allowing penetration of ICAM systems. As FIR technologies continue to mature and are allowed to operate with greater autonomy in the cyber and physical domains, human control will diminish in the ICAM systems, and digital transactions will occur even faster, far exceeding human capacities to prevent, mitigate, and respond to ICAM issues (e.g., penetrations, insecurities, and attacks) that might arise. As a result, we assess that FIR technologies will continue to be incorporated for preventing illicit access, identifying system **vulnerabilities**, and reducing the **consequences** of activity that could jeopardize personhood-controlled data and systems.

The goal would be to assist in preventing, mitigating, and responding to these insecurities in real time. In making this assessment, we note that this will be a competition in which FIR technologies will be employed by threat actors (including NBIs acting autonomously) to target ICAM systems while these ICAM systems (which include the data, networks, processes, organizations, and workforce) will be seeking to halt penetrations of their systems. The use of GANs that have two neural networks—the generator and discriminator, which produce realistic fake data and learn to identify fake data, respectively—provides an example of one way in which this competition will be waged.

Despite these approaches, we find it highly unlikely that the current and future threats, vulnerabilities, and adverse consequences can be addressed to prevent, mitigate, and respond to *all* envisioned R_S. Even once the next-generation ICAM technologies are augmented with behavioral biometrics, mobile authentication, and zero-trust architecture[133]—or when full digital government and quantum computing IAM have been integrated[134]—achieving such a definitive outcome is likely not realistic.

As a result, we assess that the overall risk regarding personhood protection and the associated FIR-related technologies will continue to be challenging because indicators in the four areas—**T_{AV}**, **threats**, **vulnerabilities**, and **consequences**—are trending toward creating greater risks. Generally, we assess that preparedness, response, and mitigations should continue to be developed to lower the R_S. Additionally, the activities should consider the ICAM, AI, and FIR technologies because they would affect ICAM outcomes. As an example, cyber and big data insecurities have the potential to put ICAM security at risk.

Table 2 provides our risk assessment of digital personhood applications. But regardless of the measures taken, a constant battle to prevent, respond to, and mitigate concerns to ensure ICAM integrity will need to be waged.

U.S. DEPARTMENT OF HOMELAND SECURITY EQUITIES

DHS relies on ICAM applications for satisfying most of its core functions. The technologies support credentialling within the largest law enforcement department in the United States. ICAM technologies support granting access to key information for DHS stakeholders across the department; in the state, local, tribal, and territorial (SLTT) Public Safety Community; and in critical infrastructure sectors. ICAM also serves as a vetting mechanism for DHS when interfacing with the private sector and helps establish the identities of members of the traveling

TABLE 2

TECHNOLOGY RISK ASSESSMENT FOR DIGITAL PERSONHOOD

	Emerging Technology Assessment				
Scenario	T_{AV} Average	Threat	Vulnerability	Consequence	Average
Short term (up to three years)	4.4	5	4.5	5	4.7
Medium term (three to five years)	4.6	5	4.5	5	4.8
Long term (five to ten years)	4.9	5	4.5	5	4.9

NOTE: The emerging technology risk assessment scale is 0 to <2 = low impact or not likely feasible, 2 to <4 = moderate impact or possible, and 4 to 5 = high impact or likely feasible.

public (i.e., when DHS employees at ports of entry examine travelers' credentials to establish their identities as a precondition for entry).

Vast troves of PII are routinely accessed by the department for various ICAM functions as well. These functions all depend on an information-sharing system that involves identity proofing, multi-factor authentications, attributes, and access management.[135] Breaching of these systems has the potential to threaten the core functions of the department and potentially to result in catastrophic outcomes. For example, such a breach can halt airline flights, as could a misidentification of a person. Similarly, these situations can halt or delay recovery payments following a disaster.

Since the creation of DHS with the Homeland Security Act of 2002, the use of digital ICAM, and now digital personhood, has undergone a transformation. Today, most of the department's missions and functions can be done digitally; in the future, nearly all missions and functions will rely on digital personhood for identity, credentialing, and access management. We anticipate that the use of these systems will increase, given the move to the Web 3.0 architecture that will incorporate digital ICAM systems that allow for machine-to-machine authentication.

At the same time, there has been a corresponding increase globally and across the United States regarding insecurities associated with the digital age. Cybersecurity attacks have increased, and data breaches are increasingly compromising the PII of large numbers of Americans. The IoT creates security and privacy concerns, and deepfakes blur the lines between fact and fiction. The increasing pace of these changes and growing use of ICAM, AI, and FIR technologies show no sign of abating. Finally, we assess that these trends will continue and lead to increasing challenges for DHS. The implementation of policy, legal, ethical, and regulatory guardrails has been slow.

DHS has an important role to play—as do a wide array of interagency, international, and SLTT partners and the private sector—in addressing the types of risks and scenarios that digital personhood poses across the ICAM space. Early consideration of the issues, prior to the fielding of ICAM, AI, and FIR capabilities and tools would be a better strategy than waiting to do so until after the technologies are deployed at scale and the ills are being discovered. An important step would be to incorporate "security by design" into ICAM and digital personhood systems.[136] Another issue that must be addressed is determining the legal standing of the various digital personhood cases—particularly for the use of digital twins that provide a virtual model of a physical object (in this case, a human) and as a virtual (or synthetic) person.

Issues related to policies, regulations, standards, norms, and ethics that are associated with digital personhood also require addressing. A thorny example that some are already discussing is voting. As mentioned earlier, humans could use digital personhood ICAM technology to represent themselves for voting, but what about giving voting rights to synthetic persons? Although the issue has been raised, critics have expressed little enthusiasm for the idea, with one cautioning that doing so "would render humans' votes meaningless."[137] Still, these are issues that relate to and affect DHS missions and equities and therefore require consideration.

CONCLUSIONS

Digital personhood applications continue to proliferate as the technologies that facilitate their development mature and new uses are discovered. It should be expected that both these trends will continue and that the technology will be used for both licit and illicit purposes. Key technologies fueling the digital personhood trends include generative AI, high-performance computing, advanced semiconductor development and manufacturing, robotics, machine learning, natural language processing, and the growth of the internet and the IoT. Although none of these technologies have likely reached their full maturity, they are already widely in use and provide users with capabilities and tools to challenge digital personhood ICAM systems using, for example, deepfakes, false information, and AI-based social engineering.

Furthermore, it is noteworthy that each of these technologies has potential risks that need to be considered, and their cumulative risks will also be inherent in digital personhood applications. The use of the technology for generating false information and deepfakes can be expected to expand, and addressing these malicious uses will require constant vigilance.

Preventing, mitigating, and responding to these risks thus require considering the threats, vulnerabilities, and consequences associated with the individual technologies and within the integrated digital personhood applications. Although today's state-of-the art technologies will continue to mature, evolve, and even transform, they already have reached a level of maturity at which the results—false information and deepfakes—confound efforts to identify and prevent ICAM system intrusions. Today, many of these intrusions are based on human-generated social engineering hacks to gain access; in the future, a growing

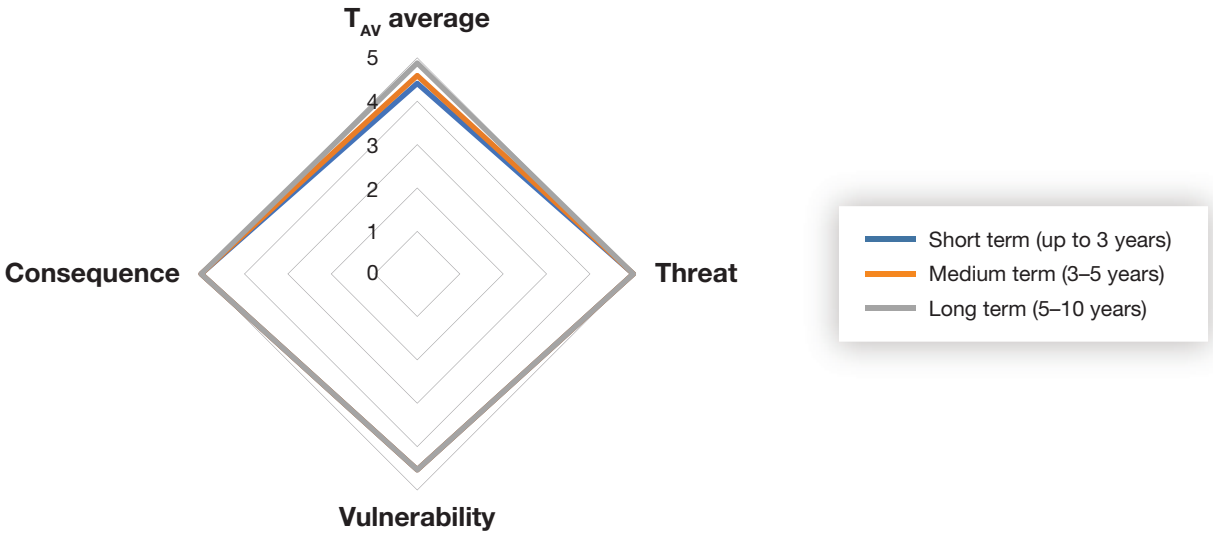
percentage of the intrusions will likely come via machines or NBIs.

All of this is occurring in a competitive environment in which humankind seeks to define appropriate behaviors and set up guardrails for digital personhood applications and for NBIs. Recent government documents have been important in this regard but are not sufficient. Follow-through on implementation will be key to ensuring that adequate guardrails and protections are in place. These measures include defining the legal authorities and liabilities for the various uses of digital personhood.

Figure 2 provides a final digital personhood risk assessment.

FIGURE 2

RISK ASSESSMENT FOR DIGITAL PERSONHOOD



NOTE: The emerging technology risk assessment scale: 0 to <2 = low impact or not likely feasible, 2 to <4 = moderate impact or possible, and 4 to 5 = high impact or likely feasible.

NOTES

[1] Blackmore, "The Making and Breaking of People."

[2] Henshall, "4 Charts That Show Why AI Progress Is Unlikely to Slow Down."

[3] WEF, "Fourth Industrial Revolution."

[4] Kurki, *A Theory of Legal Personhood*.

[5] Blackmore, "The Making and Breaking of People."

[6] U.S. Code, Title 18, Section 2510; Criminal Resource Manual, "1048. Definition—'Person.'"

[7] Garner, *Black's Law Dictionary*. The term *personhood* continues to receive scrutiny and further elaboration because of ongoing national and state discussions regarding reproductive law and related rights. These discussions are beyond the scope of this report. However, personhood has been prominently featured in the United States going back to when the U.S. Constitution was amended to "[establish] the infamous '3/5ths Compromise,' which extended personhood to 3/5ths of the enslaved population" (Blackmore, "The Making and Breaking of People"). The U.S. Supreme Court rulings on corporations having personhood are discussed later in this report; they are important to the understanding of digital personhood.

[8] The technology timelines and descriptions come from IdRamp, "History of Identity Management Infographic."

[9] This finding comes from an unpublished DHS Science and Technology briefing: Arun Vemury and Jonathan Prisby, "T3: Tech Centers Talk Tech: The Evolving World of Digital Identity," February 21, 2024.

[10] Hayeri, "Are We Ready to Face Down the Risk of AI Singularity?"

[11] Bell, "Artificial General Intelligence (AGI)."

[12] Kuusi and Heinonen, "Scenarios from Artificial Narrow Intelligence to Artificial General Intelligence—Reviewing the Results of the International Work/Technology 2050 Study."

[13] National Institute of Standards and Technology (NIST), Computer Security Resource Center, "Artificial Intelligence."

[14] Kuusi and Heinonen, "Scenarios from Artificial Narrow Intelligence to Artificial General Intelligence . . ."

[15] Roser, "The Brief History of Artificial Intelligence."

[16] Gerstein, *The Story of Technology:*, Ch. 6 and Appendix B.

[17] de Kerckhove and de Almeida, "What Is a Digital Persona?"

[18] Kuusi and Heinonen, "Scenarios from Artificial Narrow Intelligence to Artificial General Intelligence . . ."

[19] Unpublished briefing by Arun Vemury and Jonathan Prisby, "T3: Tech Centers Talk Tech: The Evolving World of Digital Identity," presented February 21, 2024.

[20] Gupta, "Next-Gen Authentication."

[21] IdRamp, "History of Identity Management Infographic."

[22] Unpublished briefing by Arun Vemury and Jonathan Prisby, "T3: Tech Centers Talk Tech: The Evolving World of Digital Identity," presented February 21, 2024.

[23] Zhou et al., "Synthetic Lies."

[24] OpenAI, "Sora."

[25] Bontridder and Poullet, "The Role of Artificial Intelligence in Disinformation."

[26] Henshall, "4 Charts That Show Why AI Progress Is Unlikely to Slow Down."

[27] Gonzalez, "AI Misinformation."

[28] Markoff, "Entrepreneurs See a Web Guided by Common Sense."

[29] Expert.AI, "The 8 Defining Features of Web 3.0."

[30] Expert.AI, "The 8 Defining Features of Web 3.0."

[31] Expert.AI, "The 8 Defining Features of Web 3.0."

[32] Expert.AI, "The 8 Defining Features of Web 3.0."

[33] Calibraint, "Here Is How Web 3 Is Transforming Everyday Applications."

[34] Lopez, "DOD Adopts 5 Principles of Artificial Intelligence Ethics."

[35] Nvidia, "Digital Twins."

[36] Mills, "What Defines 'Next Gen' Identity Security?"

[37] Mills, "What Defines 'Next Gen' Identity Security?"

[38] Murphy, "Matt Mills is spot on!"

[39] Nalegave, "Understanding CIAM."

[40] Cybersecurity & Infrastructure Security Agency, "Continuous Diagnostics and Mitigation (CDM) Program."

[41] Duffy, "Evolving CDM to Transform Government Cybersecurity Operations and Enable CISA's Approach to Interactive Cyber Defense."

[42] Cybersecurity & Infrastructure Security Agency, "CDM Program Overview."

[43] World.org, "Proof of Personhood."

[44] FutureEngage, "Proof of Personhood."

[45] Henshall, "4 Charts That Show Why AI Progress Is Unlikely to Slow Down."

[46] Shoback, "Four Key Drivers of the Digital Identity Market's Growth."

[47] Shoback, "Four Key Drivers of the Digital Identity Market's Growth."

[48] Shoback, "Four Key Drivers of the Digital Identity Market's Growth."

[49] Trump, "Maintaining American Leadership in Artificial Intelligence"; NIST, *U.S. Leadership in AI*.

[50] Lopez, "DOD Adopts 5 Principles of Artificial Intelligence Ethics."

[51] AI.gov, "The Government Is Using AI to Better Serve the Public."

[52] NIST, "AI RMF Development."

[53] Biden, "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence."

[54] NIST, *Reducing Risks Posed by Synthetic Content*.

[55] United Nations, "Secretary-General Urges Security Council to Ensure Transparency, Accountability, Oversight, in First Debate on Artificial Intelligence."

[56] Talagala, "The AI Act."

[57] U.S. Senate, Improving Digital Identity Act of 2023.

[58] World Foundation, "Proof of Personhood." Because much of the technology in this space is emergent and being led by the private sector, there is a dearth of rigorous research on the subject and much of it comes from corporate materials. Given the financial interest of these parties, we recognize that the claims in these materials may contain inaccuracies and not be complete.

[59] Dowell, "Fundamental Protections for Non-Biological Intelligences or: How We Learn to Stop Worrying and Love Our Robot Brethren."

[60] Dowell, "Fundamental Protections for Non-Biological Intelligences or: How We Learn to Stop Worrying and Love Our Robot Brethren."

[61] Blackmore, "The Making and Breaking of People."

[62] Blackmore, "The Making and Breaking of People."

[63] Forrest, "The Ethics and Challenges of Legal Personhood for AI."

[64] Forrest, "The Ethics and Challenges of Legal Personhood for AI."

[65] NPR, "What If an Artificial Intelligence Program Actually Becomes Sentient?"

[66] Forrest, "The Ethics and Challenges of Legal Personhood for AI."

[67] Yampolskiy, "Could an Artificial Intelligence Be Considered a Person Under the Law?"

[68] Forrest, "The Ethics and Challenges of Legal Personhood for AI."

[69] Forrest, "The Ethics and Challenges of Legal Personhood for AI."

[70] Forrest, "The Ethics and Challenges of Legal Personhood for AI."

[71] Forrest, "The Ethics and Challenges of Legal Personhood for AI."

[72] Forrest, "The Ethics and Challenges of Legal Personhood for AI"; also see Striveworks, "What Is AI Model Drift?"

[73] Forrest, "The Ethics and Challenges of Legal Personhood for AI."

[74] Forrest, "The Ethics and Challenges of Legal Personhood for AI."

[75] Li, "What Could AI Regulation in the US Look Like?"

[76] Gerstein, *Tech Wars*, p. 183.

[77] Forrest, "The Ethics and Challenges of Legal Personhood for AI."

[78] MIT Management, *When AI Gets It Wrong*.

[79] Carnegie Endowment, "The Future of AI Regulation."

[80] Smith, "What Large Models Cost You—There Is No Free AI Lunch."

[81] Bond, "It Takes a Few Dollars and 8 Minutes to Create a Deepfake. And That's Only the Start."

[82] For this risk assessment, the threat is focused on the capabilities and intentions of the actor; the vulnerabilities are weaknesses to be exploited; and the consequences are the outcomes of the even occurrence. Cox, "Some Limitations of 'Risk = Threat × Vulnerability × Consequence' for Risk Analysis of Terrorist Attacks."

[83] Hayeri, "Are We Ready to Face Down the Risk of AI Singularity?"

[84] Bell, "Artificial General Intelligence (AGI)."

[85] Forrest, "The Ethics and Challenges of Legal Personhood for AI."

[86] Mills, "What Defines 'Next Gen' Identity Security?"

[87] Europol, *Facing Reality?* Pereira, "By 2026, Online Content Generated by Non-Humans Will Vastly Outnumber Human Generated Content."

[88] Europol, *Facing Reality?*

[89] Mills, "What Defines 'Next Gen' Identity Security?"

[90] Mills, "What Defines 'Next Gen' Identity Security?"

[91] Mehta, "Addressing Security Gaps in Managing Non-Human Identities (NHI)."

[92] Santana, "National Public Data Cyber Attack."

[93] Ford, "Identity and Personhood in Digital Democracy."

[94] Zhadan, "World Economic Forum Finds That 95% of Cybersecurity Incidents Occur Due to Human Error."

[95] Wood, "Generative Adversarial Network."

[96] Marr, "How to Stop Generative AI from Destroying the Internet."

[97] Bahnsen et al., *DeepPhish*.

[98] Pehar, "How Blockchain Revolutionizes Data Integrity and Cybersecurity."

[99] Marr, "How to Stop Generative AI from Destroying the Internet."

[100] Wood, "Generative Adversarial Network."

[101] Smith, "Licensing Frontier AI Development: Legal Considerations and Best Practices."

[102] Toh, "The Algorithms Too Few People Are Talking About."

[103] Craig, "AI Watermarking."

[104] Biden, "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence."

[105] Toh, "The Algorithms Too Few People Are Talking About."

[106] Toh, "The Algorithms Too Few People Are Talking About."

[107] Toh, "The Algorithms Too Few People Are Talking About."

[108] Kennedy and Perala, "Researchers Spanning Harvard, OpenAI Propose 'Personhood Credentials' to Counter AI Deception Online."

[109] Adler et al., "Personhood Credentials."

[110] Oremus, "AI Researchers Call for 'Personhood Credentials' as Bots Get Smarter."

[111] Oremus, "AI Researchers Call for 'Personhood Credentials' as Bots Get Smarter."

[112] Oremus, "AI Researchers Call for 'Personhood Credentials' as Bots Get Smarter."

[113] Teubner, "Digital Personhood?"

[114] Teubner, "Digital Personhood?" We have chosen to identify the information in this paragraph as vulnerabilities, despite the author calling them risks. We assess that these questions of legal standing are central to the use of digital personhood for nonhuman entities and lead to vulnerabilities in their potential use.

[115] Teubner, "Digital Personhood?"

[116] Pereira, "By 2026, Online Content Generated by Non-Humans Will Vastly Outnumber Human Generated Content."

[117] Torkington, "These Are the 3 Biggest Emerging Risks the World Is Facing."

[118] Torkington, "These Are the 3 Biggest Emerging Risks the World Is Facing."

[119] United Nations Development Programme, Istanbul Regional Hub, *Mapping and Analysis of Efforts to Counter Information Pollution in Europe and Central Asia Region*.

[120] Deloitte & Touche LLP, "Combating Weaponized Misinformation."

[121] Deloitte & Touche LLP, "Combating Weaponized Misinformation."

[122] Deloitte & Touche LLP, "Combating Weaponized Misinformation."

[123] Cornell University Legal Information Institute, "Welcome to LII."

[124] Forrest, "The Ethics and Challenges of Legal Personhood for AI."

[125] Forrest, "The Ethics and Challenges of Legal Personhood for AI."

[126] Valenzuela, "The AI Standoff."

[127] Baker, "Who Has the Advantage—Attackers or Defenders?"

[128] Patel, "Defenders vs. Adversaries."

[129] St. John, "Cybersecurity Stats."

[130] Insurance Information Institute, "Facts + Statistics."

[131] Gupta, "Next-Gen Authentication."

[132] IdRamp, "History of Identity Management Infographic."

[133] Gupta, "Next-Gen Authentication."

[134] IdRamp, "History of Identity Management Infographic."

[135] Owen et al., *Identity, Credential, and Access Management (ICAM) Executive Primer*.

[136] Lostri and Sherman, "'Security by Design' in Practice."

[137] Yampolskiy, "Could an Artificial Intelligence Be Considered a Person Under the Law?"

REFERENCES

Adler, Steven, Zoë Hitzig, Shrey Jain, Catherine Brewer, Wayne Chang, Renée DiResta, Eddy Lazzarin, Sean McGregor, Wendy Seltzer, Divya Siddarth, et al., "Personhood Credentials: Artificial Intelligence and the Value of Privacy-Preserving Tools to Distinguish Who Is Real Online," arXiv, arXiv:2408.07892, last revised August 26, 2024.

AI.gov, "The Government Is Using AI to Better Serve the Public," webpage, undated. As of December 18, 2024: https://ai.gov/ai-use-cases/

Bahnsen, Alejandro, Correa Bahnsen, Ivan Torroledo, Luis David Camacho, and Sergio Villegas, *DeepPhish: Simulating Malicious AI*, Cyber Threat Analytics, Cyxtera Technologies, 2018.

Baker, Wade, "Who Has the Advantage—Attackers or Defenders?" Cyentia, November 19, 2020.

Bell, Elysse, "Artificial General Intelligence (AGI): Definition, How It Works, and Examples," *Investopedia*, updated September 24, 2024.

Biden, Joe, "Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence," Executive Order 14110, Executive Office of the President, October 30, 2023.

Blackmore, Freya, "The Making and Breaking of People: A History of Personhood in United States Constitutional Law," London School of Economics Undergraduate Political Review, blog post, March 26, 2024. As of December 17, 2024: https://blogs.lse.ac.uk/lseupr/2024/03/26/the-making-and-breaking-of-people-a-history-of-personhood-in-united-states-constitutional-law/

Bond, Shannon, "It Takes a Few Dollars and 8 Minutes to Create a Deepfake. And That's Only the Start," *Morning Edition*, NPR, March 23, 2023.

Bontridder, Noémi, and Yves Poullet, "The Role of Artificial Intelligence in Disinformation," *Data & Policy*, Vol. 3, November 25, 2021.

Calibraint, "Here Is How Web 3 Is Transforming Everyday Applications: A Look at Use Cases!" blog post, August 6, 2024. As of December 27, 2024: https://www.calibraint.com/blog/real-world-web-3-0-use-cases

Carnegie Endowment, "The Future of AI Regulation: A Conversation with Arati Prabhakar," video, November 14, 2023. As of December 18, 2024: https://www.youtube.com/watch?v=3uovOOUL4zg

Cornell University Legal Information Institute, "Welcome to LII," webpage, undated. As of December 19, 2024: https://www.law.cornell.edu/

Cox, Louis Anthony (Tony), Jr., "Some Limitations of 'Risk = Threat × Vulnerability × Consequence' for Risk Analysis of Terrorist Attacks," *Risk Analysis*, Vol. 28, No. 6, December 2008.

Craig, Lev, "AI Watermarking," TechTarget, October 2023.

Criminal Resource Manual, "1048. Definition—'Person,'" U.S. Department of Justice Archive, webpage, undated. As of December 17, 2024: https://www.justice.gov/archives/jm/criminal-resource-manual-1048-definition-person

Cybersecurity & Infrastructure Security Agency, "CDM Program Overview," fact sheet, undated.

Cybersecurity & Infrastructure Security Agency, "Continuous Diagnostics and Mitigation (CDM) Program," webpage, undated. As of December 17, 2024: https://www.cisa.gov/resources-tools/programs/continuous-diagnostics-and-mitigation-cdm-program

de Kerckhove, Derrick, and Cristina Miranda de Almeida, "What Is a Digital Persona?" *Technoetic Arts*, Vol. 11, No. 3, 2013.

Deloitte & Touche LLP, "Combating Weaponized Misinformation: Future of Risk in the Digital Era," 2019.

Dowell, Ryan, "Fundamental Protections for Non-Biological Intelligences or: How We Learn to Stop Worrying and Love Our Robot Brethren," *Minnesota Journal of Law, Science & Technology*, Vol. 19, No. 1, 2018.

Duffy, Michael, "Evolving CDM to Transform Government Cybersecurity Operations and Enable CISA's Approach to Interactive Cyber Defense," Cybersecurity & Infrastructure Security Agency, blog post, July 21, 2023. As of December 17, 2024: https://www.cisa.gov/news-events/news/evolving-cdm-transform-government-cybersecurity-operations-and-enable-cisas-approach-interactive

Europol, *Facing Reality? Law Enforcement and the Challenge of Deepfakes*, European Union Agency for Law Enforcement Cooperation, 2022.

Expert.AI, "The 8 Defining Features of Web 3.0," blog post, April 8, 2022.

Ford, Bryan, "Identity and Personhood in Digital Democracy: Evaluating Inclusion, Equality, Security, and Privacy in Pseudonym Parties and Other Proofs of Personhood," arXiv, arXiv:2011.02412v1, November 5, 2020.

Forrest, Katherine B., "The Ethics and Challenges of Legal Personhood for AI," *Yale Law Journal*, Vol. 133, April 2024.

FutureEngage, "Proof of Personhood: The Future of Digital Identity," LinkedIn, blog post, August 27, 2023. As of December 17, 2024: https://www.linkedin.com/pulse/proof-personhood-future-digital-identity-futur-engage/

Garner, Bryan A., ed., *Black's Law Dictionary*, Thomson Reuters, 2024.

Gerstein, Daniel M., *The Story of Technology: How We Got Here and What the Future Holds*, Prometheus Books, 2019.

Gerstein, Daniel M., *Tech Wars: Transforming U.S. Technology Development*, Praeger, 2022.

Gonzalez, Oscar, "AI Misinformation: How It Works and Ways to Spot It," CNET, November 8, 2023.

Gupta, Brij, "Next-Gen Authentication: Moving Beyond Passwords in Cybersecurity," Medium, March 16, 2024.

Hayeri, Amir, "Are We Ready to Face Down the Risk of AI Singularity?" *Forbes*, November 10, 2023.

Henshall, Will, "4 Charts That Show Why AI Progress Is Unlikely to Slow Down," *Time*, updated November 6, 2023.

IdRamp, "History of Identity Management Infographic," webpage, undated. As of December 17, 2024: https://idramp.com/history-of-identity-management-infographic/

Insurance Information Institute, "Facts + Statistics: Identity Theft and Cybercrime," webpage, undated. As of December 19, 2024: https://www.iii.org/fact-statistic/facts-statistics-identity-theft-and-cybercrime

Kennedy, Cass, and Alex Perala, "Researchers Spanning Harvard, OpenAI Propose 'Personhood Credentials' to Counter AI Deception Online," ID Tech Wire, August 20, 2024.

Kurki, Visa A. J., *A Theory of Legal Personhood*, Oxford Legal Philosophy 2019.

Kuusi, Osmo, and Sirkka Heinonen, "Scenarios from Artificial Narrow Intelligence to Artificial General Intelligence—Reviewing the Results of the International Work/Technology 2050 Study," *World Futures Review*, Vol. 14, No. 1, 2022.

Li, Victor, "What Could AI Regulation in the US Look Like?" *ABA Journal*, podcast, June 14, 2023. As of December 18, 2024: https://www.americanbar.org/groups/journal/podcast/what-could-ai-regulation-in-the-us-look-like/

Lopez, C. Todd, "DOD Adopts 5 Principles of Artificial Intelligence Ethics," U.S. Department of Defense, February 25, 2020.

Lostri, Eugenia, and Justin Sherman, "'Security by Design' in Practice: Assessing Concepts, Definitions, and Approaches," *Lawfare*, August 19, 2024.

Markoff, John, "Entrepreneurs See a Web Guided by Common Sense," *New York Times*, November 12, 2006.

Marr, Bernard, "How to Stop Generative AI from Destroying the Internet," *Forbes*, August 14, 2023.

Mehta, Kunal, "Addressing Security Gaps in Managing Non-Human Identities (NHI)," CredenceIA, December 3, 2018.

Mills, Matt, "What Defines 'Next Gen' Identity Security?" SailPoint, blog post, October 4, 2023. As of December 17, 2024: https://www.sailpoint.com/blog/what-defines-next-gen-identity-security

MIT Management, *When AI Gets It Wrong: Addressing AI Hallucinations and Bias*, MIT Sloan Teaching & Learning Technologies, undated.

Murphy, Stephen "Noel," "Matt Mills is spot on!" LinkedIn post, May 2024. As of December 26, 2024: https://www.linkedin.com/posts/stephennoelmurphy_what-next-gen-identity-security-actually-activity-7191251761477611520-K-3G/

Nalegave, Sid, "Understanding CIAM: The Next Generation of Identity and Access Management," ActiveCyber, July 13, 2023.

National Institute of Standards and Technology, *U.S. Leadership in AI: A Plan for Federal Engagement in Developing Technical Standards and Related Tools*, 2019.

National Institute of Standards and Technology, "AI RMF Development," webpage, updated January 2, 2024. As of December 18, 2024: https://www.nist.gov/itl/ai-risk-management-framework/ai-rmf-development

National Institute of Standards and Technology, *Reducing Risks Posed by Synthetic Content: An Overview of Technical Approaches to Digital Content Transparency*, NIST AI 100-4, April 2024.

National Institute of Standards and Technology, Computer Security Resource Center, "Artificial Intelligence," webpage, undated. As of December 17, 2024: https://csrc.nist.gov/Topics/Technologies/artificial-intelligence

NIST—*See* National Institute of Standards and Technology.

NPR, "What If an Artificial Intelligence Program Actually Becomes Sentient?" *Morning Edition*, June 21, 2022.

Nvidia, "Digital Twins," webpage, undated. As of December 17, 2024: https://www.nvidia.com/en-us/omniverse/solutions/digital-twins/?ncid=pa-srch-goog-444539

OpenAI, "Sora," webpage, undated. As of December 17, 2024: https://openai.com/sora/?ref=aihub.cn

Oremus, Will, "AI Researchers Call for 'Personhood Credentials' as Bots Get Smarter," *Washington Post*, August 21, 2024.

Owen, Christine, Larry Kroll, Chris Price, and Ryan Page, *Identity, Credential, and Access Management (ICAM) Executive Primer*, U.S. Department of Homeland Security, Science and Technology Directorate, 2019.

Patel, Raja, "Defenders vs. Adversaries: The Two-Speed Cybersecurity 2023 Race," Sophos News, April 4, 2023.

Pehar, Danny, "How Blockchain Revolutionizes Data Integrity and Cybersecurity," *Forbes*, January 17, 2024.

Pereira, Dan, "By 2026, Online Content Generated by Non-Humans Will Vastly Outnumber Human Generated Content," OODA Loop, March 6, 2024.

Roser, Max, "The Brief History of Artificial Intelligence: The World Has Changed Fast—What Might Be Next?" Our World in Data, December 6, 2022.

Santana, Danni, "National Public Data Cyber Attack: Massive Data Breach Exposed Countless Social Security Numbers and Personal Info," CNET, updated August 21, 2024.

Shoback, Jackie, "Four Key Drivers of the Digital Identity Market's Growth," *Forbes*, November 30, 2022.

Smith, Craig, "What Large Models Cost You—There Is No Free AI Lunch," *Forbes*, updated January 1, 2024.

Smith, Gregory, "Licensing Frontier AI Development: Legal Considerations and Best Practices," *Lawfare*, January 3, 2024.

St. John, Mariah, "Cybersecurity Stats: Facts and Figures You Should Know," *Forbes*, updated August 28, 2024.

Striveworks, "What Is AI Model Drift?" undated.

Talagala, Nisha, "The AI Act: Three Things To Know About AI Regulation Worldwide," *Forbes*, June 29, 2022.

Teubner, Gunther, "Digital Personhood? The Status of Autonomous Software Agents in Private Law," Ancilla Iuris, 2018.

Toh, Amos, "The Algorithms Too Few People Are Talking About," *Lawfare*, January 4, 2024.

Torkington, Simon, "These Are the 3 Biggest Emerging Risks the World Is Facing," World Economic Forum, January 13, 2024.

Trump, Donald J., "Maintaining American Leadership in Artificial Intelligence," Executive Order 13859, Executive Office of the President, February 11, 2019.

United Nations, "Secretary-General Urges Security Council to Ensure Transparency, Accountability, Oversight, in First Debate on Artificial Intelligence," press release, July 18, 2023.

United Nations Development Programme, Istanbul Regional Hub, *Mapping and Analysis of Efforts to Counter Information Pollution in Europe and Central Asia Region*, 2022.

U.S. Code, Title 18, Crimes and Criminal Procedure; Part I, Crimes; Chapter 119, Wire and Electronic Communications Interception and Interception of Oral Communications; Section 2510, Definitions.

U.S. Senate, Improving Digital Identity Act of 2023, Bill 884, March 21, 2023.

Valenzuela, Ismael, "The AI Standoff: Attackers vs. Defenders," BlackBerry, March 7, 2024.

WEF—*See* World Economic Forum.

Wood, Thomas, "Generative Adversarial Network," DeepAI, undated.

World Economic Forum, "Fourth Industrial Revolution," webpage, undated. As of December 17, 2024: https://www.weforum.org/focus/fourth-industrial-revolution

World Foundation, "Proof of Personhood: What It Is and Why It's Needed," blog post, February 21, 2024. As of December 17, 2024: https://world.org/blog/worldcoin/proof-of-personhood-what-it-is-why-its-needed

Yampolskiy, Roman V., "Could an Artificial Intelligence Be Considered a Person Under the Law?" PBS, October 7, 2018.

Zhadan, Anna, "World Economic Forum Finds That 95% of Cybersecurity Incidents Occur Due to Human Error," Cybernews, January 18, 2022.

Zhou, Jiawei, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury, "Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions," in *CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023.

ABBREVIATIONS

AGI	artificial general intelligence
AI	artificial intelligence
DHS	U.S. Department of Homeland Security
EO	Executive Order
FIR	Fourth Industrial Revolution
GAN	generative adversarial network
IAM	identity and access management
ICAM	identity, credential, and access management
IMT	identity management technology
IoT	Internet of Things
ML	machine learning
NBI	nonbiological intelligence
NIST	National Institute of Standards and Technology
PII	personally identifiable information
R&D	research and development
RMF	Risk Management Framework
R_S	risks and scenarios
T_{AV}	technology availability
WEF	World Economic Forum

About the Author

Daniel M. Gerstein is a senior policy researcher at RAND and a professor of policy analysis at the Pardee RAND Graduate School. He formerly served as the Acting Under Secretary and Deputy Under Secretary of the DHS Science and Technology Directorate from 2011 to 2014. He has a doctorate in biodefense.

ABOUT THIS REPORT

This report is part of a series of analyses on the effects of emerging technologies on U.S. Department of Homeland Security (DHS) missions and capabilities. As part of this series, the Homeland Security Operational Analysis Center (HSOAC) was charged with developing a technology and risk assessment methodology for evaluating emerging technologies and understanding their implications within a homeland security context. The methodology and analyses provide a basis for DHS to better understand the emerging technologies and the risks they present.

This research was sponsored by the DHS Science and Technology Directorate and conducted by the Management, Technology, and Capabilities Program of the Homeland Security Research Division (HSRD), which operates HSOAC.

HSOAC
HOMELAND SECURITY
OPERATIONAL ANALYSIS CENTER

An FFRDC operated by RAND
under contract with DHS

The Homeland Security Act of 2002 authorizes the Secretary of the Department of Homeland Security (DHS), acting through the Under Secretary for Science and Technology, to establish one or more federally funded research and development centers (FFRDCs) to provide independent analysis of homeland security issues. RAND operates the Homeland Security Operational Analysis Center (HSOAC) as an FFRDC for DHS under contract HSHQDC-16-D-00007.

The HSOAC FFRDC provides the government with independent and objective analyses and advice in core areas important to the Department in support of policy development, decisionmaking, alternative approaches, and new ideas on issues of significance. The HSOAC FFRDC also works with and supports other federal, state, local, tribal, and public- and private-sector organizations that make up the homeland security enterprise. The HSOAC FFRDC's research is undertaken by mutual consent with DHS and is organized as a set of discrete tasks.

The information presented in this publication does not necessarily reflect official DHS opinion or policy.

For more information on this publication, visit www.rand.org/t/RRA2876-1.

This research was published in 2025.