Paper

02

Teaming Humans and Artificial Intelligence–Assisted Machines

Embracing the Human Systems Integration Approach

Guy André Boy

FlexTech Chair Holder, CentraleSupélec,
Paris Saclay University, France

ABSTRACT

Lying at the intersection of human factors and ergonomics and human-computer interaction, human systems integration (HSI) is an emerging discipline within the field of systems engineering that promises to resolve the challenge of seamlessly teaming humans with artificial intelligence-assisted machines. Given the expanding role of automation and autonomous systems, delineating between both concepts is essential to understanding the various challenges and opportunities posed by the integration of AI into military tasks and decision-making. Trust, collaboration, and familiarity are critical factors in achieving optimal teaming results, especially in high-risk and dynamic environments where the necessity of preserving human autonomy to operate with freedom in unexpected situations is highlighted. In this regard, emerging "digital twins" and advanced human-in-the-loop simulations present powerful tools to leverage in the HSI context. Utilizing a model-based HSI approach, validation metrics, and building systemic flexibility, a framework for designing multi-agent systems with superior results can be made possible. Standardizing integration principles, ensuring interoperability, and improving policies for continuous participation in the development of AI-assisted systems, ensuring user feedback for creating more capable interfaces, will prove crucial in overcoming key challenges.

INTRODUCTION

Artificial intelligence (AI) refers to systems that demonstrate intelligent behavior by analyzing their environment and acting to achieve specific goals with some degree of autonomy. AI supports users by performing tasks that usually require human intelligence, such as perception, conversation, and decision-making (Kanaan, 2020). Current and future applications of AI extend across a growing range of potential uses, from knowledge-based systems, vision, speech, and natural language processing, to robotics, machine learning, and planning. As AI gains an increasingly pervasive role in the design and delivery of air and space power, there is a need to adopt human systems integration-led approaches to team humans with AI-assisted machines (Boy, 2020; 2023b). Human systems integration (HSI) is an emerging discipline in the field of systems engineering and lies at the intersection of human factors and ergonomics (HFE) and human-computer interaction (HCI).

HSI applies knowledge of human capabilities and limitations throughout the design, implementation, and operation of hardware and software, placing *the human* – referring to all people involved, such as users, operators, and maintainers – as a system on par with the hardware and software systems. This paper explores issues of autonomy, trust, familiarity, control, and responsibility relating to AI-assisted systems that are progressively being embedded into tasks executed by human operators and the machines they use. This paper discusses the need for developing model-based HSI and robust validation metrics while making use of 'digital twins,' which can advance the understanding of high-impact human factors in human-machine interaction leading to more useful and usable interfaces. Along this journey, achieving systemic flexibility by addressing design gaps is essential for multi-agent systems to dynamically deliver user and warfighter requirements in unexpected situations.

CROSS-FERTILIZING PRINCIPLES AND CONCEPTS

Machine automation is generally designed to accomplish a specific set of largely deterministic steps to achieve a limited set of predefined outcomes (Hancock, 2017). Both people and machines can be automated: Human cognitive functions can be automated through intensive training, standard operating procedures (SOPs), tactics, techniques, and procedures (TTPs), for example, in addition to the human experience gained over time. On the other hand, machine automation is achieved by developing software and algorithms that replace human cognitive functions with machine cognitive functions, as is increasingly the case with AI. On the other hand, machine autonomy is essential when a system must make a timely and critical decision that cannot wait for external support, such as when systems must operate remotely. Autonomous systems, usually equipped to use embedded data or receive it from external sources to make decisions accordingly, can provide functions to operators that humans do not naturally have and offer robustness to function without external intervention or supervision (Fong, 2018). Consider the example of NASA's Curiosity and Perseverance rovers, which can autonomously maneuver on Mars from one point to another using stereo vision and onboard path planning. Machine autonomy is a growing trend in defense, though it is important to note that while different levels of autonomy exist, the current generation of machines is not self-directing (Sheridan and Verplank, 1978; Fong, 2018).

This automation-autonomy distinction is vital to address the flexibility challenge in HSI. The idea of 'human autonomy' remains a significant concern as operators must be able to solve problems in unexpected situations using the appropriate combination of creative thought, knowledge, skills, organizational and infrastructural resources, and technology. Over the past century, systems have been developed with increasing levels of automation, often without sufficiently considering unexpected situations. These challenges were left to end-users to resolve as best they could. Usually, in the military operational environment, following SOPs and TTPs present dependable solutions for automation monitoring – but only in expected situations. In unexpected situations, on the other hand, this can be counterproductive and even dangerous. Unexpected situations often demand flexible solutions based on human autonomy, deeper knowledge, and problem-solving skills.

Generally, human and machine automation only works effectively when used in specific contexts – for example, in expected situations. Outside these contexts (for instance, in unexpected situations), their rigidity can cause unintended results such as incidents or accidents (Boy, 2013b; Endsley, 2018). Human operators also tend to become complacent when using automation, primarily because it is most frequently used in the contexts in which it has been validated and where it performs more or less perfectly. Consequently, in unexpected situations, human controllers must be able to solve unexpected problems autonomously, using any available physical and cognitive resources, whether human, machine, or a combination of both. In such situations, human controllers or operators require the flexibility to apply an appropriate mix of knowledge, skills, and coordination necessary to solve problems. Given this, automation that places the human controller or operator 'out of the loop' can be highly inefficient, decrease situational awareness, and result in performance degradation (Endsley, 2015b).

> *Human operators also tend to become complacent when using automation, primarily because it is most frequently used in the contexts in which it has been validated and where it performs more or less perfectly.*

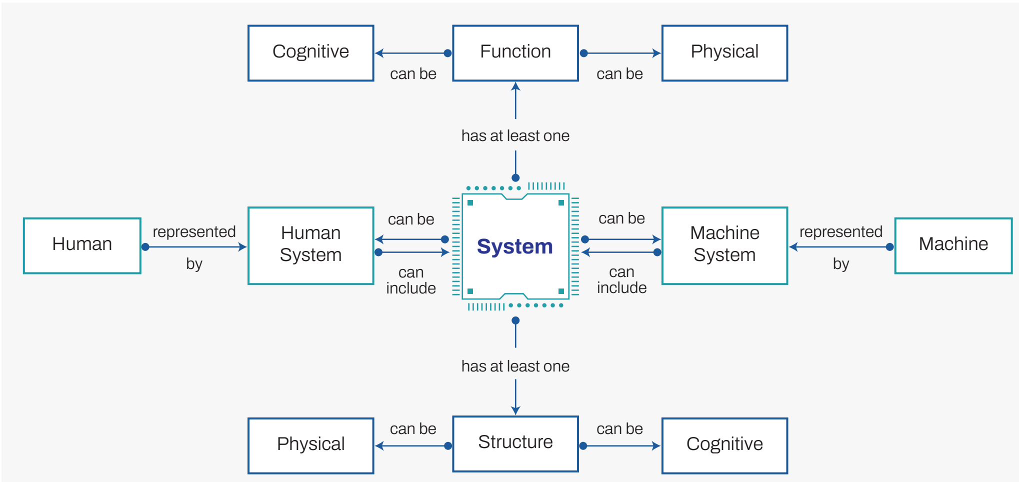

Figure 2.1: A System as a Representation of Human and Machine Agents

Maintaining business continuity during operations in unexpected, uncertain, unforeseen, and unpredictable situations demands a highly resilient system. To achieve this resilience, it is vital at the design level of systems engineering to ensure redundancy, decentralization, modularity, and interoperability, alongside providing distributed decision-making and shared situational awareness. But what do we mean by a 'system'? Adopting a consistent definition for a 'system' where humans and machines can be considered using a common representation is essential. Most people think of a system as a machine, and several definitions have been proposed. A system is "a combination of interacting systemic elements organized to achieve one or more stated objectives" (ISO/IEC 15288, 2015). Alternatively, a system is "something that does something (an activity, function) and has a structure, which evolves, in something (environment) for something (purpose)" (Le Moigne, 1990). As *Figure 2.1* illustrates, a system and its agents or sub-systems can be humans or machines capable of identifying a situation, deciding and planning the appropriate course of action. The modern air force perfectly depicts the example of a system of systems composed of human and machine agents or subsystems, where both are increasingly equipped with AI. AI-based systems can be purely software and act in the virtual world (e.g., voice assistants, image analysis software, search engines, speech, and face recognition systems) or be embedded within hardware such as robotics, semi-autonomous vehicles, or other Internet of Things applications (Smuha, 2019).

MODEL-BASED HSI AND VALIDATION METRICS

System models that reflect an abstraction of real-world phenomena can allow us to evaluate numerous metrics and outcomes to increase our understanding of those systems and their capabilities. Establishing a model-based approach with robust metrics is critical for verifying and validating a system from an HSI perspective (Damacharla et al., 2018). Against this backdrop, essential challenges relate to the design, analysis, and evaluation of an HSI-centric teaming model to resolve. The development of such models must begin with an integrative effort to generate an ontology based on the cross-fertilization of HSI and AI principles and concepts to capture the complex multi-agent representations associated with human-AI teaming (Boy, 2019). A robust conceptual model in this regard will be able to develop and validate the role of each agent or sub-system (human and machine) and how authority is shared between them to achieve optimal results in decision-making.

An essential focus of the metrics will be evaluating trust and collaboration factors (Boy et al., 2022; Boy, 2023a). Trust is critical in shaping HSI and reflects a multifaceted concept influenced by competence, predictability, transparency, and reliability factors. Trust is often associated with perceiving, understanding, and projecting a situation to decide and act, and it relates to knowing who or what controls the system or situation (Boy, 2016; Boy, 2021b). In high-risk and dynamic environments, trust is a crucial element in enabling independent and interdependent decision-making across multi-agent systems (Schaefer et al., 2019). Atkinson (2012) has deconstructed the concept of “trust” into three areas: trustworthiness, which requires possessing the necessary and sufficient qualities needed for a person to trust (e.g., competence); usability, which requires the qualities of reliability to be manifested so that they can be observed or inferred, either directly or indirectly (e.g., behaviors, signals, communications, reference, reputation, etc.); and trusting, which is the process of becoming dependent (i.e., dependent on another agent for something of value).

When there are few conflicts between agents, stable and consistent interactions between agents, both human and machine, foster trust-building. A sound reputation with consistent signals and behaviors increases trust, which is enhanced as interactions and outcomes become highly predictable and understandable. This way, trust is strongly associated with a system’s maturity level in consistently delivering satisfactory results with minimal inconvenience and transparency, making the decision-making process more understandable to human controllers and operators (Boy, 2021b). Since trust in complex systems is generally based on familiarity, when adopting new human-AI systems, it becomes essential to determine the length of time necessary for users to become familiar with the system, its functions, limitations, etc. The more familiar that users are with a complex system, the better they will understand how to work with it as a partner and, by extension, trust or distrust its behavior and properties (Salas et al., 2008).

Human-in-the-loop simulations can be highly effective in consolidating and accelerating the process of such familiarization for users of autonomous systems. Human-in-the-loop modeling and simulation tools have led to the development and growing use of ‘digital twins,’ which allow virtual representations of a system spanning its life cycle from inception to obsolescence. Digital twins are being developed across several domains, from applications relating to the operational maintenance of helicopter engines to remote operations using robotics (Camara Dit Pinto et al., 2021; Lorente et al., 2021).

> *Human-in-the-loop modeling and simulation tools have led to the development and growing use of 'digital twins,' which allow virtual representations of a system spanning its life cycle from inception to obsolescence.*

Even considering the potential advantages of digital twins and simulated training, achieving proficient familiarity and a deep understanding of complex systems can take a long time for users. Despite decades of research on interpersonal trust in human teams, human-animal teams, and human-machine interaction, critical gaps remain in our understanding of trust, perception, and manipulation, among other areas, that are highly relevant to human-machine interaction. This challenge is compounded in the context of AI, which is giving rise to more autonomous machines that have not reached maturity from three critical perspectives: Technology Readiness Levels (TRL); Human Readiness Levels (HRL); and Organization Readiness Levels (ORL), the latter being expandable further into Societal Readiness Levels (SRL) (Endsley, 2015; Boy, 2021b; NASA, 2022).

Alongside trust, cooperation, and collaboration represent two concepts closely related to human-AI teaming in emerging multi-agent systems in air power, such as Europe's Future Combat Air System (FCAS), which envisions digital avatars and loyal wingmen to operate collaboratively alongside manned fighter aircraft. During the cooperation, each agent may have individual interests but works toward a common goal through aligning actions and efforts to achieve a shared objective, despite potentially differing interests. This concept is often encountered when multiple entities, such as team-based games or multi-agent systems in the military environment, must work together. In contrast, collaboration involves agents coming together with a common interest and a shared goal. Pooling their resources, knowledge, and skills to achieve that shared goal, cooperation is typically characterized by strong communication, mutual understanding, and joint decision-making. In an orchestra, for example, musicians collaborate to perform a symphony, ensuring that their actions are precisely coordinated through the conductor's direction and the music scores (Boy, 2013a).

Training for collaboration is essential as effective teamwork is imperative for any successful teaming model. Still, special attention needs to be focused on cultivating adaptive and agile behavior in agents through HSI-based training, simulation, and resilience testing. Further understanding is required concerning AI and the autonomy it can enable, such as its ability to make choices when unexpected situations occur. Teaching agents how to communicate efficiently, understand each other's roles, and coordinate their actions to achieve desired outcomes is critical and can help build a better understanding of the autonomy that AI can enable human and machine fault tolerance and the ability of agents to make choices when unexpected situations occur. Progress here will necessarily demand solutions for shared mental or cognition models between agents, science-based authority allocation, ethics education and training, design and management for sustainability to emphasize stability, implementation, and governance, explainability of agent behavior, calibrating trust incrementally, human neutralization capability, communication protocols, and periodic evaluations to support continuous improvement.

CONCLUSION: THE FUTURE DIRECTION

The increasing importance of interconnectivity in current and emerging operational constructs for air and space power relies on a system of systems approach. For example, command and control (C2) are increasingly decentralized through computers: The digital cockpits of pilots in modern combat aircraft provide a human-centered integration of C2 systems to support distributed high-tempo operations (Boy, 2023b). However, fundamental challenges remain in human-AI teaming; model-based HSI offers tremendous potential. Measuring and leveraging the performance of human-AI teaming is essential for facilitating continuous improvement and ensuring that a system of systems comprising humans and machines increasingly assisted by AI can perform at optimal levels. To address design gaps, it will be necessary to evolve from rigid automation to flexible autonomy, where autonomy becomes a human-machine concept (as illustrated in *Figure 2.2*). Systemically built-in flexibility, such as at the technological, human, or organizational levels, will allow operators to easily preserve business continuity during operations in unexpected situations by finding support from one or more partners (Boy, 2021a).

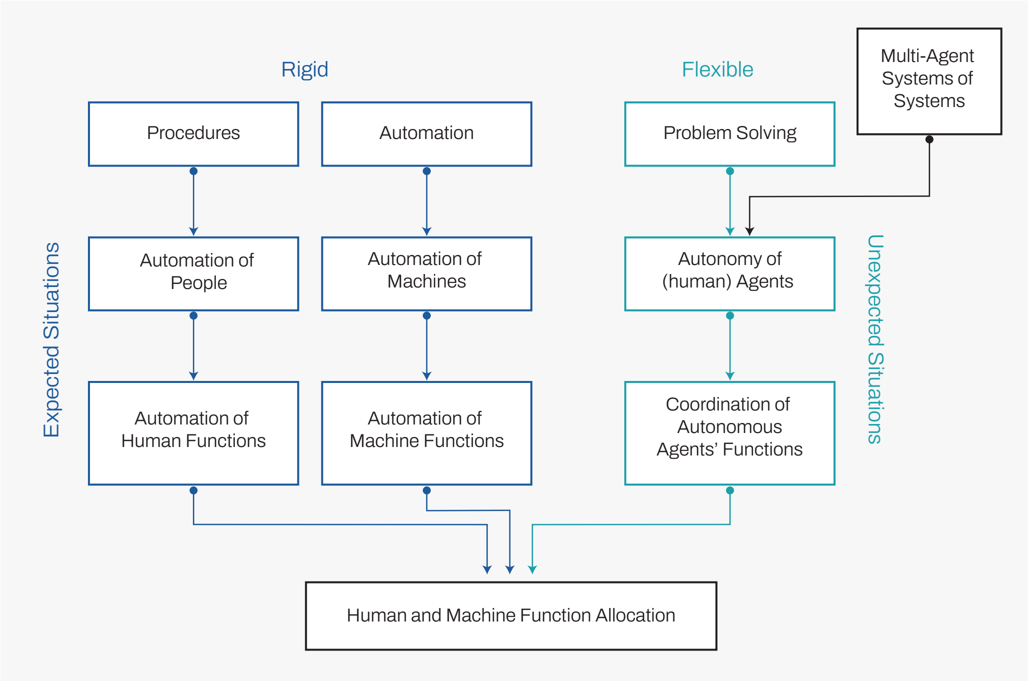

Figure 2.2: Associating Rigid Automation and Flexible Autonomy

To make this possible, building a shared understanding of HSI among stakeholders is essential for standardizing integration principles, which result in system specifications emphasizing interoperability. Interoperability has traditionally been challenging to achieve due to systems being developed by different companies using their respective methodologies, standards, and intellectual property. Beyond this, certifying the design of models and systems for human-AI partnering will depend on the ability to evaluate their safety, reliability, and effectiveness. This can be achieved by establishing robust metrics to collect and analyze meaningful data relating to performance. Developing scenarios and dynamic agent-based human-in-the-loop simulations can enable a step-by-step discovery of emerging functions and structures of systems through activity analyses that inform future iterations and evolution (Boy et al., 2022). They will also help identify critical human factors and analyze their impact. On the other hand, improving policies for continuous participation is essential to deliver a circular economy leading to more useful and usable human-machine interfaces. The technical certification and validation of systems, regulatory compliance, and managing input and feedback from user experience are crucial, presenting a future direction for research and development.

REFERENCES

Atkinson, D. (2012) *Human-Machine Trust*. Invited speech at IHMC (*Florida Institute for Human and Machine Cognition*), 4 November.

Boy, G.A. (2013) *Orchestrating Human-Cantered Design*. Springer, U.K.

Boy, G.A. (2013) 'Dealing with the Unexpected in our Complex Socio-technical World,' *Proceedings of the 12th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems*. Las Vegas, Nevada, USA. Also, Chapter in *Risk Management in Life-Critical Systems*, Millot P., and Boy, G.A., Wiley.

Boy, G.A. (2016) Tangible Interactive Systems. Springer, U.K.

Boy, G.A. (2019) 'Cross-Fertilization of Human-Systems Integration and Artificial Intelligence: Looking for Systemic Flexibility,' *Proceedings of AI4SE, First Workshop on the Application of Artificial Intelligence for Systems Engineering*. Madrid, Spain.

Boy, G.A. (2020) Human Systems Integration: From Virtual to Tangible. CRC Taylor & Francis Press, Miami, FL, USA.

Boy, G.A. (2021) Design for Flexibility - A Human Systems Integration Approach. Springer Nature, Switzerland. ISBN: 978-3-030-76391-6.

Boy, G.A. (2021) 'Socioergonomics: A few clarifications on the Technology-Organizations-People Tryptic,' Proceedings of *INCOSE HSI2021 International Conference*, INCOSE, San Diego, CA, USA.

Boy, G.A. (2023), 'Model-Based Human Systems Integration.' In the *Handbook of Model-Based Systems Engineering*, A.M. Madni & N. Augustine (Eds.). Springer Nature, Switzerland.

Boy, G.A. (2023), 'An epistemological approach to human systems integration,' *Technology in Society*, 102298. Available at: https://doi.org/10.1016/j.techsoc.2023.102298.

Boy, G.A. Masson, D, Durnerin, E. and Morel, C. (2022) 'Human Systems Integration of Increasingly Autonomous Systems using PRODEC Methodology,' FlexTech work-in-progress Technical Report. Available upon request, guy-andre.boy@centralesupelec.fr.

Pinto, C. D. Masson, D, Villeneuve, E. Boy, G.A. and Urfels, L. (2021) 'From requirements to prototyping: Application of human systems integration methodology to digital twin,' In the *Proceedings of the International Conference on Engineering Design, ICED* 2021, 16-20 August, Gothenburg.

Damacharla, P. Javaid, A.Y. Gallimore, J.J. and Devabhaktuni, V.K. (2018) 'Common Metrics to Benchmark Human-Machine Teams (HMT): A Review.' *IEEE Access* 6, 38637-38655. DOI: 10.1109/ACCESS.2018.2853560.

Endsley, M.R. (2015) 'Human Readiness Levels: Linking S&T to Acquisition [Plenary Address], *National Defense Industrial Association Human Systems Conference*, Alexandria, VA, USA.

Endsley, M.R. (2015), *Autonomous Horizon: System Autonomy in the Air Force – A Path to the Future,' Volume 1: Human-Autonomy Teaming*. Technical Report AF/ST TR 15-01, June 2015. DOI 10.13140/RG.2.1.1164.2003.

Endsley, M.R. (2018) 'Situation awareness in future autonomous vehicles: Beware the unexpected,' *Proceedings of the 20th Congress of the International Ergonomics Association*. Florence, Italy. Springer Nature, Switzerland.

Fong, T. (2018) *Autonomous Systems - NASA Capability Overview*, 24 August 24. Availale at: https://www.nasa.gov/sites/default/files/atoms/files/nac_tie_aug2018_tfong_tagged.pdf

Hancock, P.A. (2017) 'Imposing limits on autonomous systems.' *Ergonomics,* 60/2, 284-291.

Holbrook, J.B. Prinzel, L.J., Chancey, E.T., Shively, R.J., Feary, M.S., Dao, Q.V., Ballin, M.G. and Teubert, C. (2020) 'Enabling Urban Air Mobility: Human-Autonomy Teaming Research Challenges and Recommendations,' *AIAA, Aviation Forum* (virtual event), 15-19 June.

INCOSE (2022) '*Systems engineering definition*.' Available at: https://www.incose.org/about-systems-engineering/system-and-se-definition/systems-engineering-definition

INCOSE (2022) '*Systems Engineering Vision 2035*.' Available at: https://www.incose.org/docs/default-source/se-vision/incose-se-vision-2035.pdf?sfvrsn=e32063c7_2

ISO/IEC 15288 (2015) 'Systems Engineering – system life cycle processes,' IEEE Standard, International Organization for Standardization, JTC1/SC7.

Kanaan, M. (2020), *T-Minus AI – Humanity's Countdown to Artificial Intelligence and the New Pursuit of Global Power*. BenBella Books, Inc., Dallas, Texas, USA. ISBN 978-1-948836-94-4.

Le Moigne, J.L. (1990) *La modélisation des systèmes complexes*. Dunod, Paris.

Lorente, Q., Villeneuve, E., Merlo, C., Boy, G.A. and Thermy, F. (2021) 'Development of a digital twin for collaborative decision-making, based on a multi-agent system: application to prescriptive maintenance.' *Proceedings of INCOSE HSI2021 International Conference.*

Lyons, J.B., Sycara, K., Lewis, M. and Capiola, A. (2021) 'Human–Autonomy Teaming: Definitions, Debates, and Directions.' *Frontiers in Psychology – Organizational Psychology*. https://doi.org/10.3389/fpsyg.2021.589585

NASA Technology Readiness Levels (TRL) (2022) Available at: https://www.nasa.gov/directorates/heo/scan/engineering/technology/technology_readiness_level

NASEM (2021) Human-AI Teaming: State of the Art and Research Needs. National Academy of Sciences, Engineering, and Medicine, Washington D.C. The National Academy Press. Available at: https://doi.org/10.17226/26355

de Rosnay, J 1977, Le Macroscope, Vers une vision globale. Seuil - Points Essais, Paris.

Salas, E., Cooke, N.J. and Rosen, M.A. (2008) 'On Teams, Teamwork, and Team Performance: Discoveries and Developments.' *Human Factors*, Vol. 50, No. 3, June, 540–547. DOI 10.1518/001872008X288457.

Schaefer, K.E., Hill, S.G. and Jentsch, F.G. (2019) Trust in Human-Autonomy Teaming: A Review of Trust Research from the US Army Research Laboratory Robotics Collaborative Technology Alliance. Springer International Publishing AG, part of Springer Nature (outside the USA), J. Chen (Ed.): AHFE 2018, AISC 784, 102–114.
Available at: https://doi.org/10.1007/978-3-319-94346-6_10

Sheridan, T.B. and Verplank, W.L. (1978) 'Human and computer control of undersea teleoperators. Technical Report,' Man-Machine Systems Laboratory, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge. MA, USA.

Smuha, N. (2019) 'A definition of AI: Main capabilities and disciplines. High-Level Expert Group on Artificial Intelligence. European Commission, Brussels, Belgium,
Available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Woods, D.D. and Alderson, D.L. (2021) 'Progress toward Resilient Infrastructures: Are we falling behind the pace of events and changing threats?' *Journal of Critical Infrastructure Policy*, Volume 2, Number 2, Fall / Winter. Strategic Perspectives.