Application of the Human Readiness Levels to AI Intensive Systems

Cansu Yalim, Holly Handley
Department of Engineering Management and Systems Engineering, Old Dominion University
Norfolk, Virginia
cyali001@odu.edu, hhandley@odu.edu

ABSTRACT

Human Readiness Levels (HRLs) assess how mature and robust the human-centered aspects are in a system's development process (ANSI/HFES, 2021). Designed as an adjunct to Technology Readiness Levels (TRLs), the HRLs are a measure of the readiness of the technology for use by human operators and maintainers. The U.S. Department of Defense (DoD) is currently in the process of adopting the ANSI/HFES-400 as a Tier 1 non-government standard, as human factors are critical for the safety and effectiveness of a broad range of systems**.** This includes systems that comprise Artificial Intelligence (AI) components, where failures like algorithmic bias (e.g., COMPAS recidivism tool) or automation overreliance (e.g., Tesla Autopilot) underscore the consequences of inadequate human integration. Yet developers often prioritize technical performance over socio-technical alignment by risking safety, equity, and user trust. This paper adapts the HRL framework to AI-intensive systems via a dual-exit criteria model which require concurrent technical validation (e.g., model accuracy, scalability) and social validation (e.g., interpretability, bias mitigation, user trust) at all nine HRL stages. We demonstrate its application through a maintenance prioritization case study, where an ML model reduces workflow subjectivity while preserving human oversight. The framework provides actionable checkpoints—from data collection to real-world deployment—to embed human-centricity throughout the lifecycle. HRLs have the potential to become a component within broader governance efforts for AI to transform "human-centered AI" from aspiration to measurable engineering practice. Challenges in dynamic AI contexts (e.g., continuous learning) and future extensions are discussed.

ABOUT THE AUTHORS

Cansu Yalim is a Ph.D. candidate, graduate research and teaching assistant in the Department of Engineering Management and Systems Engineering at Old Dominion University. She has professional work experience in thermotechnology and the automotive industries. She received her bachelor's degree in industrial engineering and her master's degree in engineering management. Her research interests encompass predictive maintenance, intelligent root cause diagnosis for industrial applications, and inferring causality from observational data.

Holly Handley is a professor in the Department of Engineering Management and Systems Engineering and the interim dean of Interdisciplinary Schools at Old Dominion University. She has a B.S. in Electrical Engineering from Clarkson College, M.S. in Electrical Engineering from the University of California Berkeley, M.B.A. in Management Information Systems from the University of Hawaii, and a Ph.D. in Information Technology and Engineering from George Mason University. Dr. Handley applies engineering principles and experience in computational modeling to support projects in human systems engineering research and development.

Application of the Human Readiness Levels to AI Intensive Systems

Cansu Yalim, Holly Handley
Department of Engineering Management and Systems Engineering, Old Dominion University
Norfolk, Virginia
cyali001@odu.edu, hhandley@odu.edu

INTRODUCTION

The HRL framework gauges how effectively human factors are integrated and matured during system development (ANSI/HFES, 2021). It is evident that HRLs must be extended to AI-specific scenarios as AI systems get more ingrained in important activities. Recent movement in both practice and policy focus attention on the consequential exigency of including human elements into the creation and execution of AI-based systems. Conspicuous errors of AI applications in high-risk contexts have pointed up how inadequate focalization on human readiness might compromise equity as well as safety. For instance, the COMPAS recidivism algorithm was infamous for showing racial bias in its risk assessments. Whereupon it was disproportionately affecting minority groups and generating general worry over algorithmic justice (Angwin et al., 2022). Alike, some scrutiny testifying to user overreliance on automation in the omission of appropriate human oversight and interface design (Hawkins, 2024) have linked Tesla's Autopilot system to many accidents. These cases are not isolated technical failures but indeed feature a more collective deficiency. AI systems are oftentimes developed and deployed without enough addressing of decisive human factors (Yalim & Handley, 2023).

Such failures might be socio-technical rather than only technical in character—that is, problems resulting from a lack of integration between human roles and automated capabilities. They expose the results of ignoring how AI tools interact with human cognition, behavior, and organizational setting. In fields like healthcare, defense operations, and infrastructure maintenance—where decision accuracy, responsibility, and safety are thoroughly imperative (Erturun et al., 2024). Thereby, effective socio-technical integration calls for making sure AI technologies are not only functionally precise but also understandable, controllable, and trustworthy to their human counterparts (Handley & Yalim, 2024; Yalim & Handley, 2025).

Regulatory frameworks all around now formally reflect this increasing consciousness. Evolving legislative environments mirror this more and more. The AI Act of the European Union places strict limitations on "high-risk" AI systems (European Parliament & Council, 2024). This necessitates clear human-centered assessments at every stage of design and implementation. Likewise, minimizing unintentional bias and maintaining human control and accountability are key components of the U.S. Department of Defense (2022). In this context, a lifecycle-oriented framework for bringing together human-system integration throughout AI development is provided by HRLs. HRLs can be used as a governance and engineering tool to make sure that human-centered considerations are not an afterthought but rather a fundamental component of AI system readiness by implementing both technical and social validation checkpoints at every stage, from data preparation and model design to real-world deployment.

As systems evolve to include virtual team members, autonomous components, and adaptive technologies, the HRLs must be able to assess socio-technical systems that include AI components (Yalim, 2024). Researchers have highlighted that the integration of algorithms into human workflow remains a substantial challenge; developers may not fully consider the impact that AI has on human processes (Asan & Choudhury, 2021). Including Machine Learning (ML) models into systems provides the ability to analyze data, automate tasks, and enhance decision-making capabilities; these models can process large volumes of data to make predictions and offer solutions.

Nevertheless, these technological developments need to be accompanied by systematic assessments of human compatibility, which HRLs are uniquely positioned to facilitate. Therefore, this paper provides an example that illustrates the use of the HRLs during the development of a ML algorithm, a system component designed to expedite a maintenance prioritization process. Incorporating an ML model into this process can reduce the subjectivity caused by different human interpretations while minimizing human workload. Applying the HRLs to this socio-technical system evaluation can assess both the degree of consideration of the human operator but also monitoring the ML model for a continuous fit as a system component.

BACKGROUND

Current maturity assessment frameworks for developing technologies promote technological preparedness, which sometimes comes at the expense of human-system integration. Commonly used frameworks, such as Technology Readiness Levels (TRLs) (NASA, 2017) and Manufacturing Readiness Levels (MRLs) (DoD, 2020), provide defined criteria for evaluating the feasibility, performance, and scalability of technologies at various stages of development. They do this, however, with little or no consideration given to the human factors driving the system. Similarly, emerging AI tools such as the AI Readiness Index (Oxford Insights, 2024), which assesses institutional readiness for AI adoption, and ML Model Cards (Mitchell et al., 2019), which help transparency in model reporting, deliver intuitions into governance and documentation. Yet, these tools do not have a way to figure out how people use, trust, or act on AI systems in complicated settings. What is still missing is a framework that puts human readiness on the same level as technical readiness, especially in socio-technical systems where both need to grow at the same time.

The Human Readiness Levels (HRLs) framework has the potential to fill in this important gap by putting the social and technical aspects of system maturity into a set of rules. The HRL framework was created to go along with TRLs. It has a nine-level scale that rates how well a system has met human needs, capabilities, limitations, and behavioral patterns during its development. This fits with the growing calls from academics and standards groups for AI development that is in line with ethics. For instance, Floridi et al. (2018) contend that AI cannot make real progress unless it finds a balance between technical performance and social values including fairness, openness, and responsibility. At the same time, real-world studies have shown that even the high performing algorithms can be rejected by users if their decision-making processes seem unclear or do not match up with human judgment. This is known as "algorithm aversion" by Dietvorst et al. (2015).

The process of integrating ethical risk assessments into the design of intelligent and autonomous systems has started with initiatives like IEEE Standards Association (2021). Nevertheless, these frameworks frequently fail to provide a step-by-step, practical methodology that can be incorporated into engineering practice. By providing phase-specific criteria that evaluate human-system integration directly, HRLs, on the other hand, offer a retrospective evaluation mechanism in addition to a prescriptive design roadmap.

Human Readiness Levels

The HRL framework complements Technology Readiness Levels (TRLs) by explicitly addressing human-system maturity. The nine-level scale provides a mechanism to assess the degree that human-focused requirements have been incorporated into design decisions (See, 2021). Each HRL has a series of supporting questions that determine if the necessary human-system requirements at that level have been addressed (Handley, See & Savage-Knepshield, 2023).

The HRLs support integration of the human user or operator into the systems engineering effort, which is critical to the design of successful systems. HRLs have the potential to minimize the cost of design changes through early identification of human issues and reduce human error in fielded systems by tracking the mitigation of identified issues through subsequent HRL assessments (Salazar et al., 2021). The HRLs are grouped into three phases and decomposed into three distinct HRLs. This mimics the Technology Readiness Levels (TRL) widely used in government and industry.

Human Readiness Level Scale (ANSI/HFES, 2021)

Basic Research and Development: Scientific research, analysis, and preliminary development are conducted. This phase culminates in a validated proof of concept that addresses human needs, capabilities, limitations, and characteristics.

- ***HRL 1:*** Basic principles for human characteristics, performance, and behavior observed and reported.
- ***HRL 2:*** Human-centered concepts, applications, and guidelines defined.

- ***HRL 3:*** Human-centered requirements to support human performance and human-technology interactions are established.

Technology Demonstrations: The technology is demonstrated at increasing levels of fidelity, first in the laboratory and later in relevant environments. This phase concludes with demonstration of a representative system in a high-fidelity simulation or actual environment, with evaluation of human systems designs provided by representative users.

- ***HRL 4:*** Modeling, part-task testing, and trade studies of human systems design concepts and applications completed.
- ***HRL 5:*** Human-centered evaluation of prototypes in mission-relevant part-task simulations completed to inform design.
- ***HRL 6:*** Human systems design fully matured and demonstrated in a relevant high-fidelity, simulated environment or actual environment.

Full-Scale Testing, Production, and Deployment: Final testing, verification, validation, and qualification occur, with human performance evaluations based on representative users. This phase concludes with operational use of the system and continued systematic monitoring of human-system performance.

- ***HRL 7:*** Human systems design fully tested and verified in operational environment with system hardware and software, and representative users.
- ***HRL 8:*** Human systems design fully tested, verified, and approved in mission operations, using completed system hardware and software and representative users.
- ***HRL 9:*** System successfully used in operations across the operational envelope with systematic monitoring of human system performance.

METHODOLOGY

Framework Adaptation: Dual-Exit Criteria for AI Systems

In order to apply the HRLs framework to AI systems, this paper presents a dual-exit criteria model. This adaptation recognizes that AI systems function within socio-technical ecosystems, in contrast to conventional system readiness assessments that place an emphasis on technical functionality alone. Thus, two concurrent and equally important validations—technical readiness and social readiness—are used to assess each HRL level. As it relates to its intended operational context, the technical validation component evaluates whether the AI model or system component satisfies key performance metrics. Simultaneously, the social validation component assesses the systematic treatment of human-centric considerations.

All three of the HRL phases—Basic Research and Development (HRLs 1–3), Technology Demonstration (HRLs 4–6), and Full-Scale Deployment (HRLs 7–9)—have different goals and integration challenges, and this dual-validation approach corresponds with the full development lifecycle of AI systems. Foundational tasks like data collection, feature engineering, model selection, and preliminary human-centered requirement elicitation are prioritized in the Basic R&D phase (HRLs 1–3). While social checkpoints concentrate on determining stakeholder needs, evaluating interpretability requirements, and assessing the risk of ingraining bias in system assumptions, technical checkpoints in this stage guarantee the relevance, completeness, and quality of input data. The system moves from conceptual modeling to iterative prototyping and early validation during the Technology Demonstration phase (HRLs 4–6). In order to reach desired performance metrics, this stage's technical checkpoints include model training, tuning, and verification in simulated or semi-controlled environments. At the same time, social validation evaluates user interaction with early-stage prototypes, focusing on human-in-the-loop compatibility. For expressive feedback and improvement, this provides confidence that AI outputs are accurate and understandable to end users. The system is exposed to real-world scenarios during the Full-Scale Deployment phase (HRLs 7–9). Tracking performance under operational loads, identifying degradation, and guaranteeing dependability in mission-critical situations are all part of technical validation. Social validation also includes long-term assessments of the field's user experience. At this point, ethical auditing and feedback systems are essential for ongoing system development while maintaining human oversight. This adapted HRL framework functions as a methodology that ensures every system maturity transition is based on both the accomplishment of functional goals and the confirmation that the AI system is still suitable and reliable for its human stakeholders.

Human Readiness Level Example: Machine Learning Models

The example illustrates the use of the HRLs during the development of a Machine Learning (ML) model included as a system component in a maintenance prioritization work process. The HRLs include specifications for intelligent technology where the human component is the authority rather than the user (See et al., 2018); this example expands on that by viewing the ML model through both a human and a technology perspective.

Embedding an ML model into a system is a multi-step process including assessing system compatibility, preparing data, training models, deploying them within the system, and continuously monitoring performance for optimization. Socio-technical system evaluation in this case involves not just evaluating the fit of the system with the human operator but also monitoring of the ML model to ensure a continuous fit as a system component. Additionally, ML models also require an assessment of societal aspects of its use of data, such as trust, transparency, and ethics. The HRLs can evaluate effective human use across the entire lifecycle of the AI-assisted system.

The maintenance prioritization example (Kovacic, 2024) illustrates the applicability of the HRLs to an AI-assisted system. In this example, different types of maintenance employees, including engineers, technical staff, and managers, have access to a large amount of system failure records. This makes the maintenance work order prioritization process complex and time-consuming, as it depends on the humans' interpretation of the maintenance records, which is influenced by varying levels of experience and knowledge. Incorporating an ML model into the process of maintenance prioritization can reduce the subjectivity caused by different human interpretations. The goal of automation is to reduce subjectivity while minimizing human workload.

The maintenance prioritization process example is used to describe how the HRLs can assist in the systematic incorporation and evaluation of ML models into traditional systems. The following paragraphs describe how each level of the HRLs can be extended to an AI-assisted system. Additionally, the exit criterion for each level is described in two parts –the requirement for the model or data (technical) and the human (social) considerations.

HRL 1 ensures that basic principles for human characteristics, performance, and behavior are observed and reported. For AI-assisted systems, the focus of HRL 1 is on the data and initial human-centric requirements definition. The exit criteria for HRL 1 include a thorough understanding and characterization of both the data needs for the ML model and the essential human behaviors, capabilities, and limitations that are relevant to the developing concept or proposed applications must be demonstrated. For maintenance prioritization example, the operational effectiveness of the model is tied to the qualitative and quantitative attributes of the data it processes; in instances where data information is not adequately robust, the model may not yield satisfactory outcomes. Thus, the identification of the types of maintenance records and failure narratives, which are the inputs to the ML model, must be completed (technical). Additionally, an initial assessment is conducted, focusing on the tasks and roles of maintenance professionals, such as engineers and managers, who will engage with the system. This preliminary exploration may involve deciphering their interpretative constraints related to maintenance records and their adeptness in manipulating AI-supported tools (social).

HRL 2 guarantees that human-centered concepts, applications, and guidelines are defined. For AI-assisted systems, the focus of HRL 2 is on data processing and human-centered design guideline*s.* The exit criteria for HRL 2 include, first, completion of data acquisition, labeling, cleaning, text processing, and feature extraction, and second, establishment of key human-centered design principles, standards, and guidance for human interaction with the technology. For maintenance prioritization example, the procurement and labeling of maintenance logs, followed by text manipulation and feature extraction, are executed to condition the data for model training (technical). The appropriate design guidelines, which will represent the interaction of the ML model with maintenance personnel, are identified. This might include a conceptual user interface that conforms to ergonomic standards and satisfactorily conveys ML outputs to the users (social).

HRL 3 requires that human-centered requirements to support human performance and human-technology interactions are established. In the case of AI-assisted systems, HRL 3 focuses on model selection, design, and human-centered requirements. The exit criteria for HRL 3 include, first, a suitable ML model must be determined and the accompanying design of experiment created, and second, the requisite human-centric analyses must be conducted, resulting in the identification of human-centric requirements and key performance parameters that are integrated into overarching system requirements. For maintenance prioritization example, multiple ML models are identified, and experiments are structured for these models to identify the optimal model along with its hyperparameters (technical). The requirements that govern how the outcomes will be communicated to maintenance personnel are also identified. This might involve defining the modalities of notifications or alerts that will be utilized to inform staff regarding the urgency of various maintenance tasks (social).

HRL 4 requires the completion of modeling, part-task testing, and trade studies of human systems design concepts and applications. In the case of AI-assisted systems, HRL 4 focuses on model training, tuning, and part-task human testing. The existing criteria for HRL 4 include, first, adequate training and calibration of the ML model must be completed to satisfy predetermined performance metrics, and second, the assessment and characterization of human interactions and performance utilizing analytical instruments, modeling methodologies, and partial task testing with swift prototypes must be evaluated. For the maintenance prioritization example, the ML model, applied to the processed maintenance records, undergoes training and fine-tuning to achieve optimal performance (technical). Then, partial task testing with maintenance personnel, utilizing a preliminary prototype of the model completed. This might encompass simulated scenarios where staff are tasked with prioritizing maintenance tasks based on these scores (social).

HRL 5 assesses the human-centered evaluation of prototypes in mission-relevant part-task simulations to inform design. In the case of AI-assisted systems, HRL 5's focus remains on the preliminary model validation/testing and mission-relevant prototype testing. The exit criteria for HRL 5 include, first, the initial validation and examination of the ML model within a simulated or regulated environment should be completed, and second, an assessment of the human interactions and performance should be completed while interfacing with the evolving system prototype. For the maintenance prioritization example, the ML model is validated within a simulated environment, utilizing a subset of maintenance records to ensure that the model's outputs are congruent with expert evaluations (technical). Additionally, mission-pertinent partial-task simulations, involving maintenance personnel and utilizing a high-fidelity prototype, are executed. This might encompass simulated scenarios where staff are tasked with prioritizing maintenance tasks based on outputs generated by the validated model (social).

HRL 6 evaluates that the human systems' design is fully matured and demonstrated in a relevant high-fidelity, simulated environment or actual environment. In the case of AI-assisted systems, the focus of HRL 6 is on system integration, verification, and human-centric maturity. The exit criteria for HRL 6 include, first, successful integration of the ML model as a component into the overall system and confirmation of its functionality must be completed, and second, the evaluation of human interactions within the context of high-fidelity simulated or actual environments, utilizing a functional and realistic prototype, representative users, and a comprehensive array of usage scenarios and tasks must be completed. For the maintenance prioritization example, the ML model is integrated into the existing maintenance work order system, followed by a verification of its functionality. This ensures that the outcomes derived from the ML model are accurately influencing the system's prioritization algorithms (technical). Then, execution of high-fidelity simulations or real-world tests with maintenance personnel, utilizing the integrated system, is conducted. This might span a spectrum of scenarios, from standard maintenance tasks to emergency situations, to evaluate the efficiency of the integrated model in aiding work order prioritization (social).
HRL 7 evaluates that the human systems design is fully tested and verified in an operational environment with system hardware and software including representative users. In the case of AI assisted systems, the focus of HRL 7 is on operational testing and human-centric validation. The exit criteria for HRL 7 include, first, extensive testing of the ML model within an actual operational environment must be completed to ascertain adherence to all performance and safety criteria, and second, the assessment and characterization of human interactions and performance within an operational environment must be completed, employing the final development system, representative users, and a comprehensive spectrum of usage scenarios and tasks. For the maintenance prioritization example, the ML model should be subjected to testing within a live operational environment, utilizing authentic maintenance records and work orders. Performance metrics, such as accuracy, precision, and recall, in addition to safety metrics like false positives and negatives, should be evaluated (technical). Additionally, operational tests should be conducted with maintenance personnel, utilizing the final integrated system within a live environment. This might span a variety of scenarios, from standard maintenance tasks to emergency situations, to evaluate the efficiency of the integrated model in assisting work order prioritization (social).

HRL 8 evaluates that the human systems' design is fully tested, verified, and approved in mission operations, using completed system hardware and software and representative users. In the case of AI assisted systems, the focus of HRL 8 is still on continuous monitoring and improvement. The exit criteria for HRL 8 include, first, ongoing monitoring and periodic updating of the ML model must be completed to ensure it continues to meet performance and safety criteria, and second, the evaluation and characterization of human interactions and performance with the actual system in mission operations during the full range of usage scenarios and tasks completed by representative users must be completed. For maintenance prioritization example*,* monitoring protocols should be established to regularly assess the performance of the ML model. This could include automated alerts for performance degradation and mechanisms for model retraining (technical). Additionally, extensive testing of the integrated system in mission operations should be completed, involving representative users and a full range of scenarios. This could include emergency situations to assess how effectively the integrated ML model aids in work order prioritization (social).

HRL 9 evaluates that the system is successfully used in operations across the operational envelope with systematic monitoring of human system performance. In the case of AI-assisted systems, this HRL is focused on ethical, societal, and operational monitoring and evaluation. The exit criterion for this HRL is twofold – the first focuses on the system itself and the ongoing monitoring, testing, and evaluation of the fielded system to ensure it supports the mission as intended. The second criterion is focused on the impact of the use of AI through the assessment and mitigation of ethical and societal implications, including data privacy, fairness, and transparency. For the maintenance prioritization example, the ML model would undergo ongoing human systems monitoring to ensure that the intended levels of operational performance are achieved. This could include periodic evaluations involving human systems experts to assess the effectiveness of system outcomes. Additionally, these evaluations would ensure that the ML model does not introduce bias in the maintenance work order prioritization process. This could involve periodic audits of the model's decision-making process and outcomes (technical). Additionally, the ethical considerations include a thorough review of data usage, ensuring that data privacy is maintained, and that the model's decisions do not unfairly favor any group. This could involve a detailed analysis of the model's predictions and decision-making processes to ensure transparency and fairness in all operations (social).

CHALLENGES AND SOLUTIONS FOR HRLs IN AI-INTENSIVE SYSTEMS

The most fundamental challenge is the evolving needs of the users themselves. Workflow, team composition, procedural practices, and operational conditions all evolve continually around maintenance management, say, as organizations, regulations, and technology develop. Any valid evaluation using HRL conducted previously may no longer accurately represent the manner users are employing the system today. To preserve fidelity between system functionalities and human needs, HRL analysis would have to be treated as live assessments—subject to revision with large system upgrades, role redefinition, or reconfiguring the workflow.

A second challenge concerns the operational reality that AI models are non-stationary and continuously retrained, which can unpredictably alter behavior, prediction logic, or user interpretability. These changes can reduce the consistency of interaction patterns between humans and AI and erode user trust. Another solution is to incorporate version-specific human factors documentation, such as Model Cards (Mitchell et al., 2019), directly into the HRL framework. This makes it possible to track explicitly the usability, explainability, and measures of trust and to reconfirm them on each iteration of the AI system. A third, increasingly pressing concern is the build-up of ethical debt—the gradual appearance of unintentional skews or biases not evident at the time of initial deployment. Skews may get created through evolving infrastructural data, for instance, to disproportionately deprioritize mature assets or certain operational settings within maintenance priority schemes. To handle such concerns, it is necessary to transition to proactive instead of reactive governance paradigms. Integrating mechanisms for compliance at the level of HRL 9 into DevOps lifecycles—for instance, automated fairness checks or bias detection routines built into CI/CD (continuous integration/continuous deployment) processes—can maintain ethical watchfulness across the system's operating lifespan.

Collectively, such adaptations—embedded ethical automation, model-aware documentation, and periodic reassessment procedures—empower the HRL framework to sustain its validity and fidelity to changing user contexts, ongoing learning, and more autonomous AI behaviors. By doing so, they evolve HRLs from checklists for development to dynamic governance instruments whose technical efficacy is supplemented with long-term human-system veracity.

FUTURE DIRECTIONS

Any future deployment enveloping HRLs should expand from static evaluations to a dynamic and context-aware roadmap, which involves human-centric evaluations throughout the lifecycle of the system. This requires a move away from one-time assessments towards more iterative, milestone-aligned evaluations, especially for systems that operate under conditions of lifelong learning, environmental variation, and evolving human-AI role entanglement.

For adaptive AI systems, that are getting periodically retrained or optimized based on the new data streams, an HRL checkpoints must be revisited with every significant update. These reassessments are needed to preserve human-system alignment—including trust calibration, interpretability, and task allocation—while the system's internal logic changes. Without conscious consideration, there is a risk to degrade user trust, exacerbate bias, or to mismatch automated behavior from a user's goals, particularly in safety-sensitive domains. Similarly important is their use for collaborative human-AI teams where the user-AI agent relationship is not static, but "co-evolves" over time. In such settings, AI could evolve from being a passive decision support tool to an active collaborator or team member. This shift requires different forms of socio-technical coordination such as real-time feedback loops, adaptive interface design, and emergent role clarity. When interpreted with respect to these dimensions, HRLs provide a systematic approach to both formalize, and monitor, interaction protocols ensuring that transparency, shared situational awareness and mutual intelligibility are maintained among human and machine actions.

Finally, it is through these forward-looking directions that the HRL framework will continue to substantiate its status, not only as a way to evaluate readiness, but to be a living blueprint for responsible design of AI as the ways in which technical excellence meets timeless human-centered and entrenched human values.

LIMITATIONS OF THE STUDY

The application of HRLs framework reveals several limitations that warrant consideration. The domain-specific nature of the illustrative maintenance prioritization case study may raise questions about the framework's generalizability across fundamentally different contexts such as real-time autonomous navigation, clinical diagnostics, or generative AI applications. Domain-specific variables—including risk criticality, decision urgency, and error consequences—may necessitate substantial adaptations of the dual-exit criteria, which have yet to be systematically explored. Furthermore, the framework faces inherent tensions when applied to dynamic AI systems characterized by continuous learning and adaptation. The evolving nature of these systems challenges static HRL validation assumptions, as performance drift, emergent biases, or altered interaction patterns may develop silently between formal reassessments. This dynamism complicates version control for socio-technical documentation (e.g., Model Cards) and introduces subjectivity in defining thresholds for "significant updates" that warrant HRL re-evaluation, particularly for online learning systems (Yalim & Handley, 2025).

The qualitative nature of many social validation criteria—such as "appropriate interpretability," "trust calibration," or "ethical compliance"—presents another limitation due to the absence of standardized, objective metrics. This subjectivity complicates consistent assessment, especially during early HRL phases where prototypes are immature, and hinders cross-organizational benchmarking. Implementing the dual-exit criteria also demands considerable resources, requiring parallel expertise in AI engineering and human factors science, extending development cycles through iterative human-in-the-loop testing, and necessitating sustained funding for longitudinal monitoring (particularly HRLs 8-9). This resource intensity may render comprehensive HRL adoption impractical for smaller organizations or agile development teams. Scalability poses an additional challenge for complex AI systems featuring multiple interdependent components (e.g., ensemble models, multi-agent architectures). Mapping socio-technical interfaces across subsystem boundaries and applying HRLs holistically becomes combinatorially complex without established methodological guidance.

The framework's focus on anticipated human-AI interactions creates a gap in addressing unpredictable emergent behaviors. Examples include operator complacency developing in mature systems (HRL 9), unforeseen patterns of system misuse, or novel failure modes arising from complex human-AI co-adaptation, which may evade detection by predefined checkpoint criteria. Cross-cultural validity represents another constraint, as the current social validation criteria largely assume universal human factors principles. Variations in cultural norms regarding automation acceptance, decision-making hierarchies, and transparency expectations (e.g., contrasting individualist versus collectivist contexts) are underexplored and could significantly impact framework effectiveness globally.

While HRLs support alignment with regulations like the European Parliament & Council (2024), the practical mapping of specific legal mandates (e.g., "fundamental rights impact assessments") to corresponding HRL checkpoints requires further articulation. Regulatory fragmentation across jurisdictions further complicates standardized implementation.

A critical limitation is the current reliance on case studies and theoretical alignment for validation. Robust empirical evidence quantifying the HRL framework's impact—such as reductions in human error rates, operational cost savings attributable to early human factors integration, or longitudinal trends in trust calibration metrics across the system lifecycle—is needed to substantiate claims about its return on investment. Finally, the dual-exit model encounters unique challenges when applied to non-deterministic generative AI systems (e.g., LLMs). The inherent unpredictability of outputs complicates traditional technical validation, risks like hallucinations defy conventional reliability metrics, and social criteria such as "transparency" often conflict with the proprietary opacity of foundation models, revealing a frontier where the framework may require fundamental rethinking.

CONCLUSION

Our exploration of HRLs for AI-intensive systems reveals the following insight: technical capability alone cannot guarantee operational success. The maintenance prioritization case study demonstrates concretely how HRLs bridge the socio-technical gap—not as an afterthought, but as a structured framework embedded throughout development. By defining dual exit criteria (technical validation of AI models + social validation of human integration) at each maturity level, we have shown how HRLs force explicit accountability for human factors that often get overshadowed by algorithmic performance.

This dual-path approach directly addresses recurring failures in AI deployment—from biased recidivism algorithms to autonomous vehicle accidents—where inadequate human integration compromised safety and equity. When ML models handle maintenance prioritization, for instance, HRLs ensure the system reduces subjectivity without eroding human oversight or introducing new risks. The framework’s value is further underscored by its real-world traction: the U.S. DoD’s adoption of ANSI/HFES-400 as a Tier 1 standard signals a watershed shift toward mandating human readiness as non-negotiable.

Yet our work is not a panacea. As discussed, dynamic AI systems demand adaptive refinements to the HRL framework. What remains undeniable is this: Until AI is evaluated as rigorously for human compatibility as for technical accuracy, deployments will keep stumbling over preventable socio-technical pitfalls. We argue HRLs offer the scaffolding to make “human-centered” more than a buzzword—transforming it into measurable, auditable engineering practice.

Future work must expand validation beyond maintenance domains and develop standardized metrics for social criteria. But as AI permeates high-stakes domains—healthcare, defense, infrastructure—this framework provides an actionable blueprint for building systems that are not just capable, but responsible.

REFERENCES

ANSI/HFES (2021). Human readiness level scale in the system development process. American National Standards Institute/Human Factors and Ergonomics Society. my.hfes.org/online-store/publications.

See, J. (2021). Human Readiness Levels Explained, *Ergonomics in Design: The Quarterly of Human Factors Applications,* 29(4) pp. 5-10. https://doi.org/10.1177/10648046211017410.

Salazar, G., See J., Handley H. & Craft, R. (2021). Understanding human readiness levels. *Proc Hum Factors Ergon Soc Annu Meet*; 2021; 64(1): 1765–1769. doi:10.1177/1071181320641427

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2022). Machine bias. In Ethics of data and analytics (pp. 254-264). Auerbach Publications.

Hawkins, A. J. (2024). Tesla’s autopilot and full self-driving linked to hundreds of crashes, dozens of deaths. The Verge, April, 24.

Yalim, C., Handley, H.A.H. : The effectiveness of visualization techniques for supporting decision-making. In: Proceedings of the Modeling, Simulation and Visualization Student Capstone Conference 2023. Suffolk, VA, (2023). https://doi.org/10.25776/Y65Q-QY42.
Erturun U, Yalim C, West JE. Energy harvesting face mask using a thermoelectric generator for powering wearable health monitoring sensors. Electron Lett. 2024 ; 60 : 12. https://doi.org/10.1049/ell2.13241.

H. A. H. Handley and C. Yalim, "An Information Model for Synergistic Decision Support," *2024 IEEE International Systems Conference (SysCon)*, Montreal, QC, Canada, 2024, pp. 1-5, doi: 10.1109/SysCon61195.2024.10553604.

C. Yalim and H. A. H. Handley, "Execution of the Information Model for Synergistic Decision Support Using IMPRINT Simulation," *2025 IEEE International systems Conference (SysCon)*, Montreal, QC, Canada, 2025, pp. 1-7, doi: 10.1109/SysCon64521.2025.11014786.

Yalim, C., & Handley, H. A. (2025). Integrating Data Management Plans Into the Unified Architecture Framework Standards Views. *Systems Engineering*, e70002. https://doi.org/10.1002/sys.70002.

European Parliament and Council. (2024, June 13). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (AI Act). Official Journal of the European Union, L 1689. Retrieved from https://eur-lex.europa.eu/eli/reg/2024/1689/oj

U.S. Department of Defense, Responsible AI Working Council. (2022, June). Responsible Artificial Intelligence Strategy and Implementation Pathway. Washington, DC: U.S. Department of Defense. Retrieved from https://media.defense.gov/2022/Jun/22/2003022604/-1/-1/0/Department-of-Defense-Responsible-Artificial-Intelligence-Strategy-and-Implementation-Pathway.PDF

Yalim, C. (2024). INTEGRATING GENERATIVE ARTIFICIAL INTELLIGENCE WITH SYSTEMS ARCHITECTING DIAGRAM CREATION: ADVANCEMENT, CHALLENGES, OPPORTUNITIES AND FUTURE PERSPECTIVES. In Proceedings of the International Annual Conference of the American Society for Engineering Management. (pp. 1-11). American Society for Engineering Management (ASEM).

Asan, O. & Choudhury, A. (2021). Research Trends in Artificial Intelligence Applications in Human Factors Health Care: Mapping Review. *JMIR Human Factors*, 8(2). https://doi.org/10.2196/28236.

NASA. (2017). Technology readiness level definitions. Retrieved from https://www.nasa.gov/directorates/somd/space-communications-navigation-program/technology-readiness-levels/

Department of Defense. (2020). Manufacturing Readiness Level (MRL) Deskbook. Retrieved from https://www.dodmrl.com/MRL%20Deskbook%20V2020.pdf

Oxford Insights. (2024). Government AI Readiness Index 2024. Retrieved July 2025, from https://oxfordinsights.com/ai-readiness/ai-readiness-index/

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., ... & Gebru, T. (2019, January). Model cards for model reporting. In Proceedings of the conference on fairness, accountability, and transparency (pp. 220-229). doi/10.1145/3287560.3287596

Floridi, L., Cowls, J., Beltrametti, M. et al. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds & Machines 28, 689–707 (2018). https://doi.org/10.1007/s11023-018-9482-5

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: people erroneously avoid algorithms after seeing them err. Journal of experimental psychology: General, 144(1), 114. https://doi.org/10.1037/xge0000033

IEEE Standards Association. (2021). IEEE standard model process for addressing ethical concerns during system design (IEEE Std 7000-2021). IEEE. Retrieved from https://standards.ieee.org/ieee/7000/6781/

Handley, H., See, J., & Savage-Knepshield, P. (2024). Human Readiness Levels and Human Views as Tools for User-Centered Design. Systems Engineering.2024;1-14. https://doi.org/10.1002/sys.21773

See, J., Morris, J. Craft, R. Moulton, M. & Trujillo, S. (2018). Incorporating Human Readiness Levels at Sandia National Laboratories. *Journal of Human Performance in Extreme Environments*, 14(1). DOI: 10.7771/2327-2937.1085

Kovacic, S. (2024). *Digital Transformation for Ship Maintenance.* Technical Report, Military Sealift Command. Norfolk, VA: Old Dominion University, pp. 254.