College of Engineering
UNIVERSITY OF WISCONSIN–MADISON

The HRL Playbook

A guide to winning the AI and Technology implementation game

The HEART Initiative



Tony McDonald, PhD
Co-Lead
Associate Professor, Industrial and Systems Engineering
Co-Captain, Embodied AI Moonshot



Ranjana Mehta, PhD
Co-Lead
Professor, Industrial and Systems Engineering
Interim Director, Grainger Institute for Engineering (GIE)

The HEART Initiative





UW–Madison
College of Engineering

Grainger Institute
for Engineering

Academic
Departments

HEART
Initiative

Research
Collaborators

Industrial & Systems
Engineering

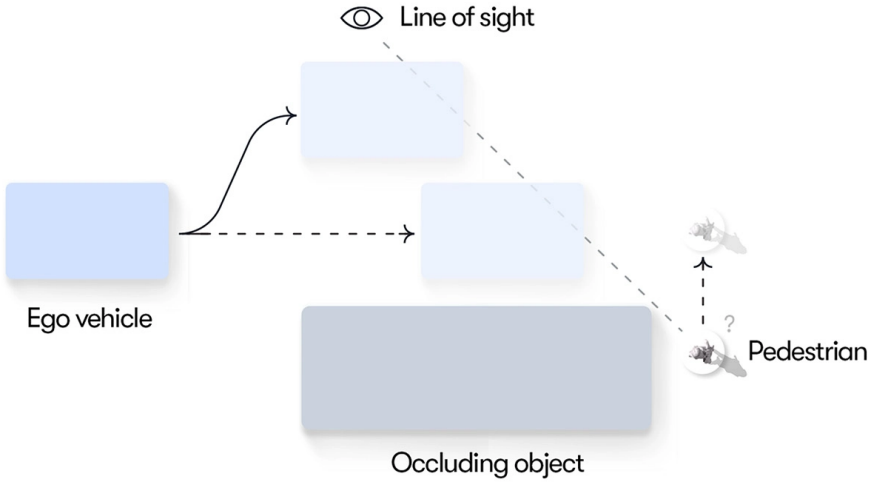
Mechanical
Engineering

Civil
Engineering

Engineering Systems for People

We develop methods, metrics, and frameworks that capture how humans interact with emerging technologies in intelligent systems, then translate this knowledge into human-ready technologies that empower people to thrive.







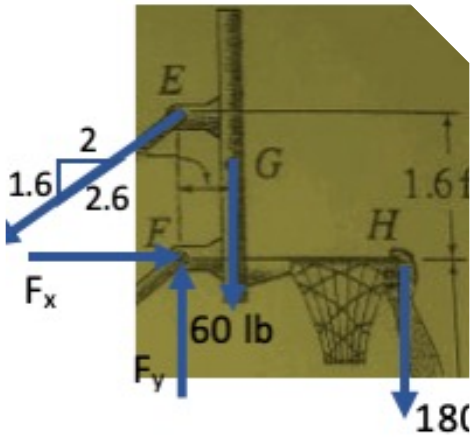
How most technology is developed today:

Design request:
We need a system that a user can throw a leather ball through.

Specifications:
- The system needs to fit a ball that is 29.5” in circumference and weighs 22 ounces.
- The system needs to be above the ground on one side and connected to the ground on the other.
- The system needs to bounce back throws that would otherwise go past it.
- The system should support the weight of a human up to the 90th percentile in human weight for brief periods

How most technology is developed today

Basic principle definitions



Bench prototypes and material selection



Full scale prototyping and testing



Design for X (assembly, mfg) and Release



What happens when this technology is used by humans?



User issues → Rework



What happens when this technology is used by humans?



User issues → Training needs



What happens when this technology is used by humans?

Will we be able to get the ball through at 99% effectiveness?

Technical Sales:

Yes, our testing showed over 99.99% is achievable

User issues → Limited ROI

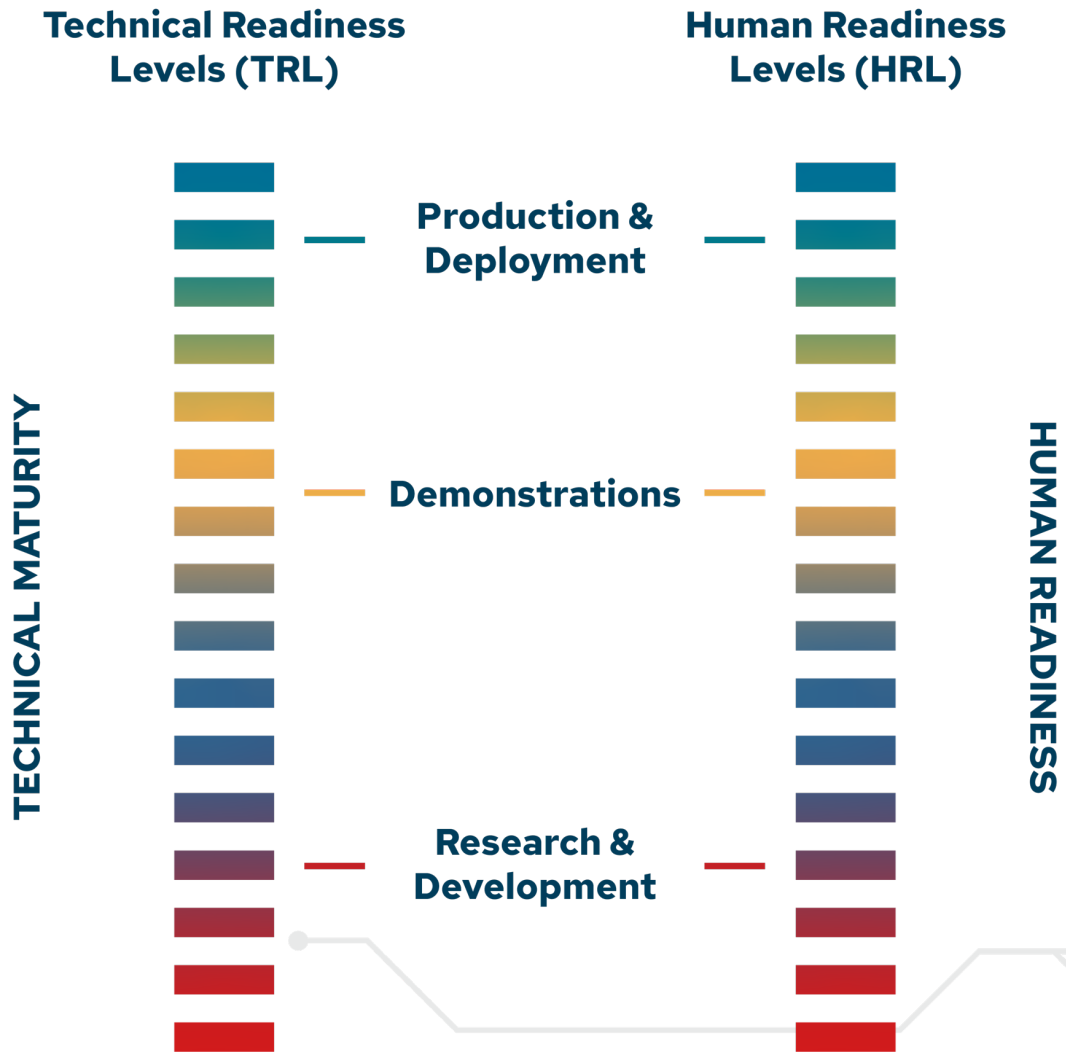
31	Winter, Nolan	31	30	.569
02	Boyd, Nick	35	35	.480
32	Bieliauskas, Aleksas	35	28	.431
25	Blackwell, John	34	34	.430
15	Gard, Isaac	14	0	.429
22	Rapp, Austin	30	12	.419

Even a well-trained team shoots less than 60%!

Has this happened to your organization?



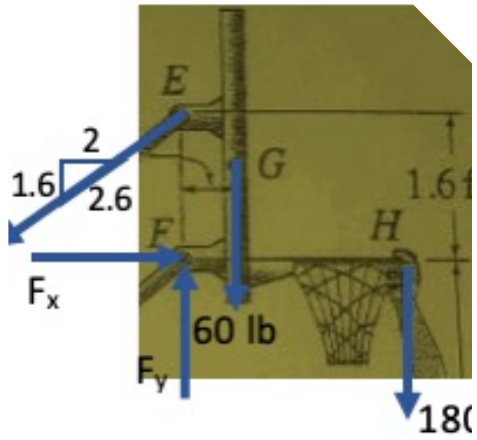
These issues can be prevented with Human Readiness Levels (HRLs)



Developing technology for humans: Levels 1-2

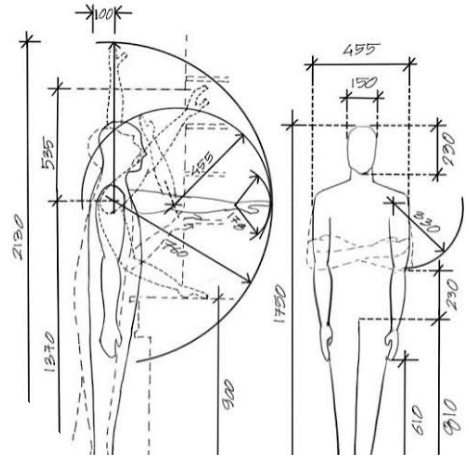
Technology Readiness

Basic principle definitions



Human Readiness

Basic human principle definitions

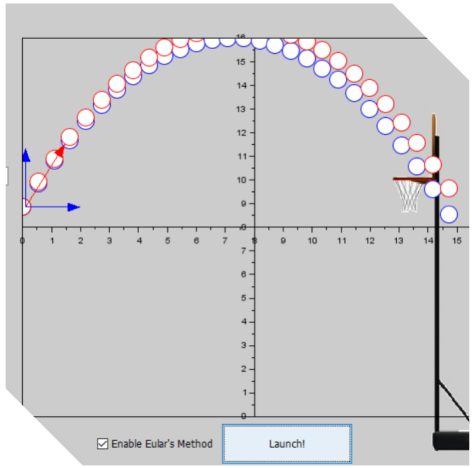




Developing technology for humans: Level 3

Technology Readiness

Physical simulation



Human Readiness

Human simulation across the user spectrum



Developing technology for humans: Levels 4-5

Technology Readiness

Physical testing



Human Readiness

Human testing with representative users





Developing technology for humans: Levels 6-7

Technology Readiness

Full-scale testing



Human Readiness

Interface testing with representative users



Developing technology for humans: Levels 8–9

Technology Readiness

Field deployment

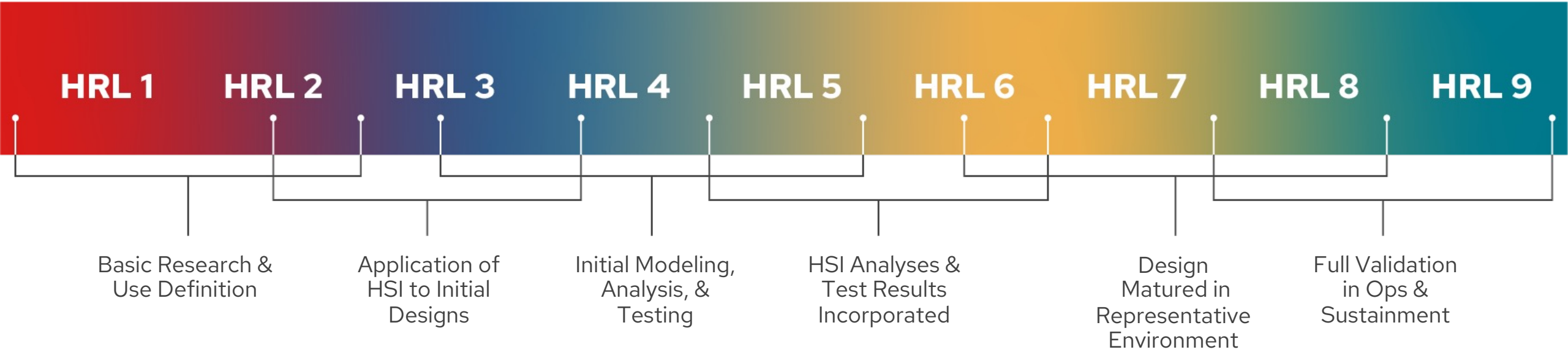


Human Readiness

System deployment & performance monitoring



Human Readiness Levels Scale



Group Discussion

1. In your past experience, what HRL level are most technologies at when they enter your company?

2. Think about a new technology that you are developing, implementing, or considering purchasing.

   Have you considered the human readiness of that technology? Why? Why not?



Case study: The Integrated Visual Augmentation System (IVAS)





Case study: The Integrated Visual Augmentation System (IVAS)

Army Scraps $1.8 Billion IVAS Investment

By Legis1 Editorial · Edited by Joanne Levine

Tuesday, August 18, 2026

Design need

- A tool to enhance soldier situation awareness (SA) on the battlefield
- Higher SA leads to better decision making, improved lethality, and fewer casualties



Technology: The Microsoft Hololens



HRL Levels 1 and 2: What needed to happen

HRL 1–2 are about doing the homework before building anything: identifying who will use the system and their capabilities, then translating that into concrete design guidelines and measurable human-performance metrics.

HRL Levels 1 and 2: What actually happened

- No metrics for user acceptance and use were identified
- Reports suggest that the Army leveraged IVAS technical readiness to assume human readiness
- Assumptions were made that research on consumer use cases (mostly stationary users indoors at desks) would translate directly to the field
- The initial operational test with real soldiers (conducted after a $1.8 Billion spend) found significant issues with cybersickness and reduced lethality

What you can ask to avoid this or ensure positive outcome

- What concepts might be relevant to investigate for this technology?
- Who are the users for this technology? (Hint: Think broad here)
- What types of studies or data would be relevant here?
- What guidelines would ensure that this system is safe and effective?
- What human performance data do we need?
- What objective and subjective metrics will best measure our success?

HRL Level 3: What needed to happen

HRL 3 focuses on completing a task analysis and safety review for each user population. These are typically aided by simulations that use known human limitations (e.g., field of view, fatigue) to predict expected outcomes.

HRL Level 3: What actually happened

- No evidence of task analysis, performance analysis, or simulation analysis being performed
- Assumptions were made that research on consumer use cases (mostly stationary users indoors at desks) would translate directly to the field
- The initial operational test with real soldiers (conducted after a $1.8 Billion spend) found significant issues with the goggles obscuring soldiers' peripheral vison

What you can ask to avoid this or ensure positive outcome

- Has a task/function analysis been completed for every type of system user?
- Has the target user population been characterized (physical size, capabilities, ages, accessibility needs)?
- Have the environmental conditions where the system will be used been analyzed?
- Have interactions between users and this system been simulated for different user groups and conditions?
- What are the human-centered requirements, how are they measured, and what are the necessary targets for success?
- Do we have a draft human-systems integration plan?

HRL Levels 4-5: What needed to happen

HRLs 4-5 use human studies experiments starting from basic lab-based studies using real tasks and expanding to real users in real environments performing task simulations needed to be conducted. Key metrics identified in HRLs 1-3 needed to be captured and design updates to address shortcomings needed to be completed.

HRL Levels 4–5: What actually happened

- The initial operational test with real soldiers (conducted after a $1.8 Billion spend) was the first human subjects study with real tasks, users, and environments
- No user benchmarks or metrics were available for assessment (the test was “bad” but the problems to solve were not clear)

What you can ask to avoid this or ensure positive outcome

- What human studies have been conducted with this device?
- What users, tasks, and environments has this system been used with/in?
- What are the human factors metrics around the use of this system (e.g., performance, workload, fatigue)?
- Can you share reports on the user testing that has been performed using this technology? Do these results include both objective (task performance) and subjective (e.g., usability) metrics?
- Has this technology been evaluated by Human Factors and Ergonomics experts?

HRL Levels 6-7: What needed to happen

HRL 6-7 focus on continued human studies focused on the specific planned user groups and high-fidelity scenarios or real operational environments

HRL Levels 6-7: What actually happened

- The Army did conduct these studies and identified the cyber sickness issues along with equipment and vision challenges
- The project was halted before soldiers were placed at risk on the battlefield

What you can ask to ensure a positive outcome

- Have my use cases been evaluated with high fidelity simulation or real-world pilots?
- What issues have been identified in testing or pilots within operational environments?
- How were these issues addressed? (note: beware of attempts to “train” errors out of a system)
- What verification exists that these issues have been mitigated?

HRL Levels 8-9: What needed to happen

At HRLs 8-9 the system should be pilot tested then fully deployed with a monitoring plan in place. The monitoring plan should capture both performance and human factors metrics identified in the lower levels.

HRL Levels 8–9: What actually happened

- The IVAS program never reached deployment

What you can ask to avoid this or ensure positive outcome

- What is the process for collecting Human Factors data when we deploy this system?
- Who is responsible for monitoring and evaluating this system?
- What is the continuous improvement plan for similar future products/integrations?
- Are we collecting anonymous near-miss data and investigating it?

Why this case is interesting

- The Army (and US Military at-large) is the gold standard for technology development and integration
- This technology was ready and it still failed
- The Army has ergonomists and Human Factors staff and resources

Group Exercise

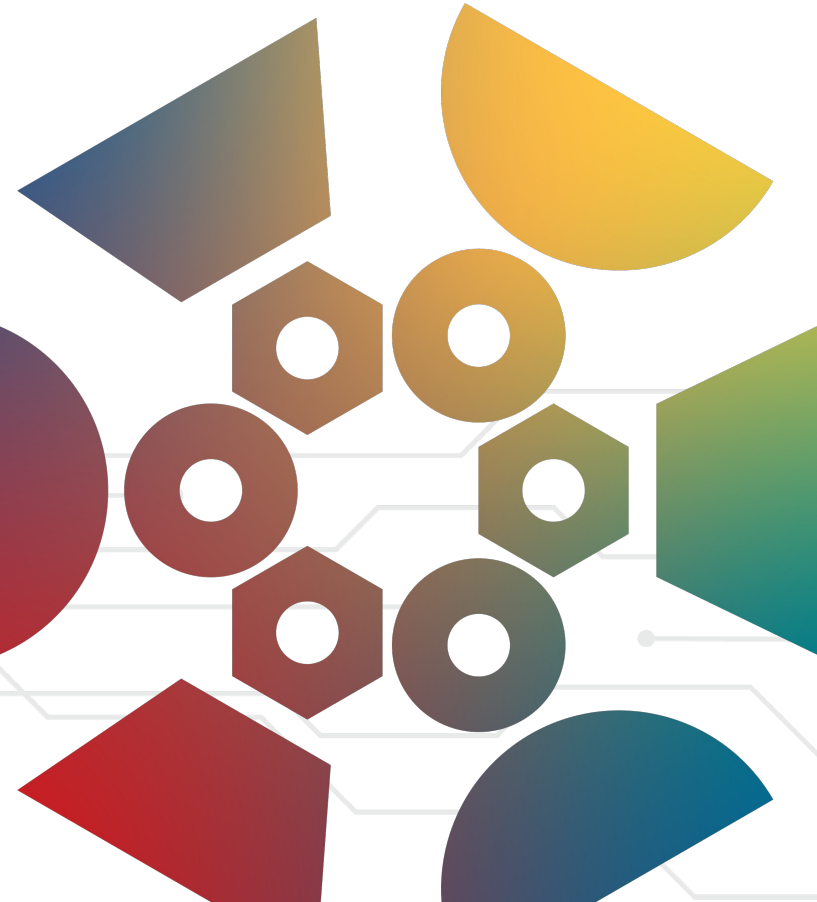
1. Identify one technology that your group is implementing.

2. Talk to the team about your plan for purchasing/implementation.

3. Where are you in the purchasing process/implementation process.

4. Go back to the questions slides for each HRL levels and evaluate what questions you have asked so far, some you need to ask, and who in your organization will ask them.

5. Report back



Where to go from here?

To do this on your own:
ANSI/HFES 400-2021

To do this with support from HEART:

- Meet with you to implement the playbook
- Conduct HRL evaluations with you
- Engage in research components of HRL (experiments, simulation, field testing)

heartinitiative.engr.wisc.edu/contact



HEART Initiative

Human-systems Engineering for Adaptive, Real-World, & Translational Applications