

# Unification Standard Level for Physical AI Oncology Trials

Standardizing and Evaluating Robot Unification Readiness

for Multi-Site Physical AI Oncology Clinical Trials

Kevin Kawchak\*

DOI: [10.5281/zenodo.18778220](https://doi.org/10.5281/zenodo.18778220)

February 2026

## Abstract

As physical AI systems advance toward clinical deployment in oncology, no standardized framework exists to evaluate how ready a robotic platform is for unified, multi-site clinical trials. Current technology readiness assessments (e.g., NASA TRL, MLTRL) do not capture the unique demands of cross-platform simulation switching, AI integration, inter-organizational robot progress sharing, and federated regulatory compliance required for multi-site oncology trials. This paper introduces the **Unification Standard Level (USL)**, a 1.0–10.0 scoring framework that evaluates physical AI robots across four equally weighted dimensions: (A) Simulation Framework Switching, (B) Generative/Agentic AI Integration, (C) Cross-Robot Progress Sharing, and (D) Multi-Site Clinical Trial Collaboration. We apply USL to nine robots across three categories—collaborative robots (cobots), surgical robots, and humanoid robots—finding per-dimension scores ranging from 1.5 to 8.5 and final composite scores from 3.4 to 7.4. The Franka Emika Panda (USL 7.4) and da Vinci dVRK (USL 7.1) lead their respective categories, both driven by large open-source ecosystems. Clinical trial readiness (Dimension D) remains the weakest dimension for seven of nine robots evaluated, revealing a field-wide gap between research maturity and clinical deployment infrastructure. All scoring code, robot evaluation modules, and documentation are open-source at <https://github.com/kevinkawchak/physical-ai-oncology-trials> under MIT license.

**Keywords:** Unification Standard Level, Physical AI, Oncology Clinical Trials, Robot Readiness, Technology Readiness Level, Collaborative Robots, Surgical Robots, Humanoid Robots

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	The Need for Standardized Robot Readiness . . . . .	2
1.2	Repository Overview . . . . .	2
1.3	The Unification Directory and Path to USL . . . . .	2
<b>2</b>	<b>Methods</b>	<b>3</b>
2.1	Development Process and AI Tools . . . . .	3
2.2	USL Scoring Methodology . . . . .	3
2.2.1	Dimension Definitions and Weights . . . . .	3
2.2.2	Score Computation . . . . .	3
2.3	Category-Specific Scoring Engines . . . . .	4
2.4	Individual Robot Evaluation Modules . . . . .	5
<b>3</b>	<b>Results</b>	<b>5</b>
3.1	Complete USL Scores . . . . .	5
3.2	Cobot Results and Score Rationale . . . . .	5
3.3	Surgical Robot Results and Score Rationale . . . . .	5
3.4	Humanoid Robot Results and Score Rationale . . . . .	6
3.5	Cross-Category Visual Comparison . . . . .	6
3.6	USL Impact on Future Trials . . . . .	6
<b>4</b>	<b>Discussion</b>	<b>6</b>
4.1	Open-Source Ecosystem Size Predicts USL Score . . . . .	6
4.2	Hardware Excellence Does Not Equal Readiness . . . . .	7
4.3	Clinical Trial Readiness Lags Field-Wide . . . . .	7
4.4	Why Scoring Differs Across Categories . . . . .	7
4.5	Why Individual Robot Code Differs . . . . .	7

\*Corresponding author. GitHub: <https://github.com/kevinkawchak/physical-ai-oncology-trials>

<b>5</b>	<b>Limitations and Future Work</b>	<b>7</b>
5.1	Human Limitations	7
5.2	Claude Code Limitations	7
5.3	USL Framework Limitations and Future Work	7
<b>6</b>	<b>Conclusion</b>	<b>8</b>
	<b>Acknowledgments</b>	<b>9</b>
	<b>Ethical Disclosures</b>	<b>9</b>
	<b>Rights and Permissions</b>	<b>9</b>
	<b>Cite This Article</b>	<b>9</b>

## 1 Introduction

### 1.1 The Need for Standardized Robot Readiness

Physical AI—the integration of embodied robotic systems with modern artificial intelligence—is poised to transform oncology clinical trials. Robotic systems now participate in surgical procedures, laboratory automation, patient logistics, and specimen handling. As multiple robot platforms from different manufacturers converge on clinical applications, no standard exists to evaluate whether a given robot is ready for *unified*, multi-site deployment.

Existing readiness frameworks address adjacent but different concerns:

- **NASA/DOD TRL** [1]: Evaluates hardware technology maturity (levels 1–9) from basic principles to flight-proven systems. Does not address software interoperability, AI integration, or multi-site clinical coordination.
- **MLTRL** [2]: Extends TRL to machine learning systems with levels covering data readiness, model development, and deployment. Does not address cross-robot sharing or simulation framework switching.
- **TRL for Complex Systems** [3]: Identifies challenges in applying TRL to integrated multi-technology systems—directly relevant but does not provide a scoring mechanism for robotic clinical trial readiness.
- **LLM Recommendations for Oncology** [4]: Recommends LLM usage for oncology trials and motivates the need for standardized AI-robot evaluation in clinical settings.

The Unification Standard Level (USL) fills this gap by providing a quantitative, reproducible scoring framework designed for evaluating robot unification readiness in multi-site oncology clinical trials.

### 1.2 Repository Overview

The USL framework is implemented within the `physical-ai-oncology-trials` repository [5], a comprehensive open-source project for integrating physical AI into oncology. Key statistics (v1.7.0, February 2026):

- **51 Python modules** (40,526 LOC), CI-validated on Python 3.10, 3.11, 3.12
- **69 documentation files**, 28 examples, 5 CLI tools, 1,414+ automated tests
- **18 text diagrams** documenting USL results, meaning, and impact per category

```
physical-ai-oncology-trials/
+-- unification/           # Core unification framework
|   +-- usl/              # USL scoring + 9 robot evaluations
|   |   +-- cobots/       # Franka, Kinova, xArm (v1.4.0)
|   |   +-- surgical/     # dVRK, Hugo, Versius (v1.5.0)
|   |   +-- humanoids/    # Atlas, Digit, Optimus (v1.6.0)
|   +-- simulation_physics/ # Isaac <-> MuJoCo bridge (Dim A)
|   +-- agentic_generative_ai/ # LLM/VLA interfaces (Dim B)
|   +-- cross_platform_tools/ # ONNX export, validation (Dim C)
+-- federation/           # Multi-site federated learning (Dim D)
+-- regulatory/           # FDA, IRB, ISO compliance (Dim D)
+-- digital-twins/        # Patient-specific tumor modeling
+-- privacy/              # HIPAA, de-identification
```

### 1.3 The Unification Directory and Path to USL

The `unification/` directory implements seamless interoperability across four pillars, each mapping directly to a USL dimension:

1. **Simulation Physics** (`simulation_physics/`) → **Dimension A**: Bridges between Isaac Lab [6], MuJoCo [7], Gazebo, and PyBullet; URDF/SDF/MJCF model conversion; physics parameter mapping.
2. **Agentic/Generative AI** (`agentic_generative_ai/`) → **Dimension B**: Unified agent interfaces for CrewAI [8], LangGraph [9], MCP [10]; VLA model integration (GR00T [11]); LLM task planning.

3. **Cross-Platform Tools** (cross\_platform\_tools/) → **Dimension C**: Framework detection, cross-platform validation suites, ONNX policy export.
4. **Federation & Regulatory** (federation/, regulatory/) → **Dimension D**: Federated learning, differential privacy, FDA submission tracking, ISO/IEC compliance [12, 13, 14].

These pillars matured through v1.0.0 (Feb 8, 2026) to v1.3.0 (Feb 16, 2026), establishing the infrastructure upon which USL scoring was built starting with v1.4.0 (Feb 23, 2026).

## 2 Methods

### 2.1 Development Process and AI Tools

This paper and the underlying USL framework were developed collaboratively between Kevin Kawchak and **Anthropic Claude Code Opus 4.6**, with acknowledgment to OpenAI ChatGPT for supplementary assistance. The primary development tool was Claude Code operating within the Claude Cowork environment, which generated code, documentation, scoring frameworks, and this paper based on structured prompts from Mr. Kawchak. Each version was generated through prompts archived in `unification/usl/prompts.md`.

**AI manufacturer and model:** Primary—Anthropic Claude Code Opus 4.6 (autonomous file creation, git operations, web search, code execution). Secondary—OpenAI ChatGPT 5.2 Thinking.

**Development timeline and versions:**

**Table 1:** Development Timeline (USL-relevant versions)

Version	Date	Milestone
v1.0.0	Feb 8, 2026	Initial release: 51 modules, 40,526 LOC
v1.3.0	Feb 16, 2026	Visualization suite: 30 interactive charts
v1.4.0	Feb 23, 2026	<b>USL for Cobots:</b> Franka, Kinova, xArm evaluated
v1.5.0	Feb 24, 2026	<b>USL for Surgical:</b> dVRK, Hugo, Versius evaluated
v1.6.0	Feb 24, 2026	<b>USL for Humanoids:</b> Atlas, Digit, Optimus evaluated
v1.7.0	Feb 24, 2026	USL Restructure: 18 diagrams, category READMEs
v1.8.0	Feb 26, 2026	<b>This paper:</b> comprehensive USL publication

This paper (v1.8.0) was primarily generated by Claude Code Opus 4.6 based on Mr. Kawchak’s prompt(s) and prior Claude Code pull requests (v1.4.0–v1.7.0).

### 2.2 USL Scoring Methodology

#### 2.2.1 Dimension Definitions and Weights

USL evaluates robots across four equally weighted dimensions (25% each):

**Table 2:** USL Dimension Definitions and Weights

Dim	Name	Wt	What It Measures
A	Simulation Switching	25%	Ability to transfer policies between sim engines (Isaac Lab, MuJoCo, Gazebo, PyBullet, Drake)
B	AI Integration	25%	Integration with LLMs, VLAs, diffusion policies, MCP, agentic frameworks
C	Cross-Robot Sharing	25%	Intra- and inter-organization progress sharing via ONNX, ROS 2, federated learning
D	Clinical Trial Collab	25%	Regulatory compliance, federated deployment, audit trails across clinical sites

#### 2.2.2 Score Computation

The final USL score is the weighted average, clamped to  $[1.0, 10.0]$  and rounded to nearest 0.1:

$$\text{USL}_{\text{final}} = \text{clamp}\left(\frac{w_A \cdot S_A + w_B \cdot S_B + w_C \cdot S_C + w_D \cdot S_D}{w_A + w_B + w_C + w_D}, 1.0, 10.0\right) \quad (1)$$

where each  $w_X = 0.25$ . Scores map to five bands (Table 3) and ten granular levels (Table 4).

**Table 3:** USL Score Bands

Score	Band	Description
9.0–10.0	Exemplary	Fully unified, multi-site clinical trial ready
7.0–8.9	Advanced	Strong unification, near clinical-trial ready
5.0–6.9	Intermediate	Partial unification, significant work remaining
3.0–4.9	Foundational	Basic interoperability, major gaps exist
1.0–2.9	Initial	Minimal unification capability

**Table 4:** USL Level Definitions (1–10)

Lvl	Name	Description
1	Conceptual	Robot exists; no simulation or AI integration attempted
2	Exploratory	Single framework tested; basic model available
3	Basic	2+ frameworks; initial AI experiments conducted
4	Developing	Cross-framework transfer demonstrated; AI planning tested
5	Functional	3+ frameworks; agentic AI operational; intra-org sharing
6	Integrated	Multi-framework validated; LLM planning; inter-org sharing
7	Advanced	GPU sim + policy transfer; MCP/VLA; skill sharing
8	Clinical-Ready	Multi-site tested; regulatory docs; federated learning
9	Validated	Full regulatory compliance; multi-site trials active
10	Exemplary	Production deployment; open consortium; continuous improvement

### 2.3 Category-Specific Scoring Engines

Each category uses a specialized Python scoring engine that adapts USL dimensions with category-specific criteria.

**Cobot Scoring** (`usl_scoring_framework.py`, 865 lines): The foundational engine defines Python dataclasses for each dimension with enumerated criteria. Dimension A scores per-framework support (official models, URD-F/MJCF/SDF availability, ROS 2, GPU capability, examples, bidirectional transfer), then adds breadth, transfer testing, and format coverage bonuses:

**Listing 1:** USL Dimension A computation from the cobot scoring engine

```

1 def compute_score(self) -> float:
2     """Compute dimension A score (0.0-10.0)."""
3     if not self.framework_scores:
4         self.score = 1.0
5         return self.score
6     avg_fw = sum(fs.compute_score()
7                 for fs in self.framework_scores
8                 ) / len(self.framework_scores)
9     breadth_bonus = min(
10         self.num_frameworks_supported / 6.0, 1.0) * 2.0
11     transfer_bonus = (1.0
12                     if self.cross_framework_transfer_tested else 0.0)
13     format_bonus = min(
14         len(self.model_format_coverage) / 4.0, 1.0)
15     raw = (avg_fw * 0.5 + breadth_bonus
16          + transfer_bonus + format_bonus)
17     self.score = round(max(1.0, min(raw, 10.0)), 1)
18     return self.score

```

**Surgical Scoring** (`usl_surgical_scoring.py`, 852 lines): Extends with surgical-specific criteria per dimension:

- **Dim A:** tissue deformation simulation, haptic feedback, instrument model coverage
- **Dim B:** surgical video AI, imitation learning, instrument segmentation, phase recognition
- **Dim C:** open-source platform availability, instrument interchangeability, community activity
- **Dim D:** FDA clearance, CE marking, remote proctoring, IEC 80601 [14]

**Humanoid Scoring** (`usl_humanoid_scoring.py`, 896 lines): Extends with:

- **Dim A:** full-body model fidelity, locomotion/manipulation sim, terrain simulation
- **Dim B:** foundation model integration (GR00T [11]), whole-body learning
- **Dim C:** standardized morphology, locomotion/manipulation transferability
- **Dim D:** autonomous navigation safety, hospital pilot testing, ISO 13482 [12]

## 2.4 Individual Robot Evaluation Modules

Each of the nine robots has a dedicated Python module (400–650 lines) containing hardware specifications, kinematic models (DH parameters, joint limits), framework configurations, oncology-specific task definitions (observation/action spaces, force limits), cross-organization sharing interfaces, and a USL evaluation function returning scores with strengths, gaps, and recommendations.

## 3 Results

### 3.1 Complete USL Scores

Table 5 presents the complete evaluation results for all nine robots.

**Table 5:** Complete USL Scores for All Nine Evaluated Robots

Robot	Manufacturer	A	B	C	D	USL	Band
<i>Collaborative Robots (Cobots) — v1.4.0</i>							
Franka Panda	Franka Robotics	8.0	7.7	8.5	5.5	<b>7.4</b>	Advanced
Kinova Gen3	Kinova Robotics	6.3	5.2	5.8	5.5	<b>5.7</b>	Intermediate
xArm 7	UFACTORY	4.6	3.2	3.8	2.1	<b>3.4</b>	Foundational
<i>Surgical Robots — v1.5.0</i>							
da Vinci dVRK	Intuitive Surgical	7.5	7.2	6.8	7.0	<b>7.1</b>	Advanced
Hugo RAS	Medtronic	4.5	4.0	3.5	5.8	<b>4.5</b>	Foundational
Versius	CMR Surgical	3.2	2.8	2.5	5.2	<b>3.4</b>	Foundational
<i>Humanoid Robots — v1.6.0</i>							
Atlas Electric	Boston Dynamics	7.0	7.4	4.5	4.2	<b>5.8</b>	Intermediate
Digit	Agility Robotics	5.8	5.4	3.0	2.7	<b>4.2</b>	Foundational
Optimus Gen 2	Tesla	3.4	5.0	1.5	4.4	<b>3.6</b>	Foundational

### 3.2 Cobot Results and Score Rationale

**Franka Emika Panda — USL 7.4 (Level 7: Advanced).** Highest-scoring robot across all categories.

- **Dim A (8.0):** 5 frameworks (MuJoCo [7], Isaac Lab [6], Gazebo, PyBullet, Isaac Sim). Official models in MuJoCo Menagerie [15] and Isaac Lab. GPU via MJX. Bidirectional policy transfer. All 4 model formats (URDF, MJCF, SDF, USD).
- **Dim B (7.7):** 5 AI capabilities (VLA, diffusion policy, LLM planning, agentic AI, generative AI). Natural language control. AI safety constraints.
- **Dim C (8.5):** 4 sharing methods (ONNX, ROS 2 actions, checkpoints, shared state). Inter-org ONNX transfer with Kinova/UR5e/xArm. Skill library. Real-time sync.
- **Dim D (5.5):** Federated compatible. ISO 13482 aligned. *No multi-site trial deployment, HIPAA tools, or 21 CFR Part 11 audit trail.*

**Kinova Gen3 7DoF — USL 5.7 (Level 5: Functional).** Lightest 7-DOF cobot (8.2 kg) with Intel RealSense depth camera. *A (6.3):* 4 frameworks, partial GPU. *B (5.2):* LLM planning, no VLA/diffusion. *C (5.8):* Intra-org Kortex API, ONNX. *D (5.5):* Clinical workflow ready, ISO aligned. Heritage in assistive robotics (JACO arm).

**UFACTORY xArm 7 — USL 3.4 (Level 3: Basic).** Most affordable 7-DOF cobot with built-in collision detection, IP51 rating. *A (4.6):* 3 frameworks, no GPU. *B (3.2):* Basic AI only. *C (3.8):* Intra-org SDK sharing across xArm family (5/6/7/Lite 6/850). *D (2.1):* Remote monitoring only, no regulatory documentation.

### 3.3 Surgical Robot Results and Score Rationale

**da Vinci (dVRK) — USL 7.1 (Level 7: Advanced).** FDA cleared, 9,000+ systems worldwide, 14M+ procedures across 69 countries [16].

- **Dim A (7.5):** 5 frameworks (ORBIT-Surgical/Isaac Lab [17], SurRoL/PyBullet [18], AMBF [19], Gazebo, MuJoCo). GPU surgical RL with tissue deformation and haptic feedback.
- **Dim B (7.2):** 9 AI capabilities including surgical video AI, instrument segmentation, phase recognition, imitation learning, VLA compatibility.
- **Dim C (6.8):** Open-source via JHU dVRK (45+ institutions). Inter-org transfer proven. Active community. ONNX export. Real-time state sync.
- **Dim D (7.0):** FDA cleared. Remote proctoring. IEC 80601. Audit trail. **Highest clinical readiness score.**

**Hugo RAS — USL 4.5 (Level 4: Developing).** *A (4.5):* Limited open sim. *B (4.0):* Developing AI. *C (3.5):* Primarily proprietary. *D (5.8):* CE marked, EU multi-center trials ongoing.

**CMR Versius — USL 3.4 (Level 3: Basic).** *A* (3.2): Minimal framework support. *B* (2.8): Basic AI. *C* (2.5): Proprietary. *D* (5.2): CE marked, NHS deployed, modular arm design supports scalable multi-site deployment.

### 3.4 Humanoid Robot Results and Score Rationale

**Atlas Electric — USL 5.8 (Level 5: Functional)** [20].

- **Dim A (7.0):** 4 frameworks (Isaac Lab, MuJoCo, Drake [21], Gazebo). Full-body GPU sim. Terrain simulation.
- **Dim B (7.4):** Highest humanoid AI score. Foundation model potential (GR00T). Whole-body control. Locomotion + manipulation policies. LLM planning (BDAIL).
- **Dim C (4.5):** Proprietary. Intra-org BD/TRI only. ONNX export available.
- **Dim D (4.2):** No healthcare deployment. Safety cert potential.

**Agility Digit — USL 4.2 (Level 4: Developing)** [22]. GR00T N1 [11] partnership with NVIDIA. First commercial humanoid deployed (Amazon). *A* (5.8), *B* (5.4), *C* (3.0), *D* (2.7): no hospital deployment.

**Tesla Optimus Gen 2 — USL 3.6 (Level 3: Basic).** Most capable hands (11 DOF, 22 joints) and mass production potential. *A* (3.4): proprietary, no public models. *B* (5.0): FSD-derived perception, Dojo training. *C* (1.5): **lowest sharing score of all 9 robots**—fully proprietary. *D* (4.4): mass production projected 2027.

### 3.5 Cross-Category Visual Comparison

**Listing 2:** Dimension A and D scores across all nine robots

Dim A: Simulation Framework Switching (25% weight)		
-----		
Franka Panda	[=====]	8.0 5 fwks, GPU, transfer
da Vinci dVRK	[=====]	7.5 ORBIT-Surgical, tissue
Atlas Electric	[=====]	7.0 Drake, full-body GPU
Kinova Gen3	[=====]	6.3 4 fwks, partial GPU
Digit	[=====]	5.8 GR00T sim-to-real
xArm 7	[=====]	4.6 3 fwks, basic models
Hugo RAS	[=====]	4.5 Limited open sim
Optimus Gen 2	[=====]	3.4 Proprietary only
Versius	[=====]	3.2 Minimal framework
Dim D: Clinical Trial Collaboration (25% weight)		
-----		
da Vinci dVRK	[=====]	7.0 FDA, 14M+ procedures
Hugo RAS	[=====]	5.8 CE marked, EU trials
Franka Panda	[=====]	5.5 Federated, ISO aligned
Kinova Gen3	[=====]	5.5 Clinical workflow
Versius	[=====]	5.2 CE marked, NHS deployed
Optimus Gen 2	[=====]	4.4 Mass production
Atlas Electric	[=====]	4.2 Safety cert potential
Digit	[=====]	2.7 Amazon warehouse only
xArm 7	[=====]	2.1 Monitoring only

### 3.6 USL Impact on Future Trials

**Listing 3:** Phased timeline for physical AI oncology trials

Phase 1 (2026): Category-specific single-site trials
Cobots --> Lab automation (Franka leads)
Surgical --> dVRK-based research trials (45+ inst)
Humanoid --> Logistics pilots (Atlas, Digit)
Phase 2 (2027): Cross-site policy transfer
Franka (train) --ONNX--> Kinova (validate)
dVRK (Site A) --ONNX--> dVRK (Site B)
Standard action spaces enable multi-site
Phase 3 (2028+): Cross-category integration
Surgical (dVRK) <--> Cobot (Franka)
Humanoid (logistics) supports surgical teams
Unified multi-site consortium operations

## 4 Discussion

### 4.1 Open-Source Ecosystem Size Predicts USL Score

The strongest predictor of USL score is open-source ecosystem size and maturity. The two highest-scoring robots—Franka Panda (7.4) and da Vinci dVRK (7.1)—have the largest communities in their categories: Franka has 1,000+ GitHub stars across franka\_ros2, panda-gym, and libfranka; dVRK has 45+ institutions using the open-source platform. Larger ecosystems directly enable higher Dim A (more framework support), Dim B (more AI research), and Dim C (more sharing infrastructure).



## 4.2 Hardware Excellence Does Not Equal Readiness

The xArm 7 has the best environmental protection (IP51) and built-in collision detection among cobots, yet scores lowest (3.4). Optimus Gen 2 has the most capable hands (11 DOF, 22 joints) and mass production potential, yet scores only 3.6. USL measures *software interoperability*—simulation coverage, AI integration depth, and sharing infrastructure—not hardware quality alone.

## 4.3 Clinical Trial Readiness Lags Field-Wide

Dimension D is the weakest dimension for 7 of 9 robots. Even Franka (7.4 overall) scores only 5.5 on Dim D. Only da Vinci dVRK achieves  $\text{Dim D} \geq 7.0$ , reflecting FDA clearance and 20+ years of clinical deployment. This reveals a field-wide gap: robots mature rapidly in simulation and AI but clinical trial infrastructure (federated learning, regulatory docs, audit trails, HIPAA compliance) remains underdeveloped.

## 4.4 Why Scoring Differs Across Categories

Each scoring engine adapts USL dimensions to reflect fundamentally different clinical roles. Cobots focus on laboratory manipulation (needle insertion, pipetting, sample handling), emphasizing ONNX policy portability, ROS 2 interfaces, and control frequency (250–1000 Hz). Surgical robots focus on intraoperative procedures, adding tissue deformation, haptic feedback, surgical video AI, FDA/CE clearance [23], and remote proctoring. Humanoid robots focus on hospital logistics (supply delivery, specimen transport, decontamination), adding whole-body locomotion, foundation model integration, navigation safety, and hospital pilot testing.

## 4.5 Why Individual Robot Code Differs

Each robot’s evaluation module differs due to unique: (1) kinematic structures—Franka uses standard DH parameters with 7 revolute joints; Kinova uses modified DH with continuous rotation joints; Atlas has 28 DOF across 8 joint groups; Digit has backward-bending knees with spring energy models; (2) control interfaces—libfranka C++ at 1 kHz vs. Kortex API multi-mode vs. xArm Python SDK at 250 Hz; (3) oncology task definitions with different observation/action dimensions (Franka `needle_insertion`: obs=26, act=7 vs. Kinova `medication_dispensing`: obs=20, act=7 vs. Atlas `supply_transport`: obs=48, act=12); and (4) sharing ecosystems ranging from Franka’s broad inter-org ONNX sharing to Optimus’s fully proprietary platform with no public APIs.

# 5 Limitations and Future Work

## 5.1 Human Limitations

- **Single evaluator:** All USL scores assigned by one researcher (Kawchak) with AI assistance, introducing potential evaluator bias. Multi-evaluator validation is needed.
- **No physical robot access:** Evaluations based on documentation, code, and published research—not hands-on hardware testing.
- **Limited clinical expertise:** Author is not a clinical oncologist; Dim D assessments need input from practicing oncologists and clinical trial coordinators.
- **Scope constraints:** Only 3 robots per category. Expansion to additional manufacturers (Universal Robots, ABB, Stryker Mako, Figure AI) would improve coverage.

## 5.2 Claude Code Limitations

- **Knowledge cutoff:** Claude Code’s training data cutoff (May 2025) means developments after that date required explicit web search verification.
- **Static scoring:** Scoring functions are embedded in evaluation code rather than dynamic real-time scoring against live robot interfaces.
- **Prompt dependency:** Generated code quality is bounded by prompt specificity. Under-specified prompts can lead to incomplete evaluations.
- **Hallucination risk:** AI-generated content requires careful verification against primary sources. All scores and specifications were cross-referenced with manufacturer documentation and open-source repositories.
- **Single-session generation:** Each version (v1.4.0–v1.8.0) was generated in a single Claude Code session, limiting iterative refinement within a version.

## 5.3 USL Framework Limitations and Future Work

- **Equal weights:** The 25% equal weighting may not reflect clinical priorities; Dim D may warrant higher weight for deployment decisions.
- **No real-world validation:** USL scores have not been validated against actual multi-site clinical trial outcomes.
- **Rapidly evolving field:** Scores are a February 2026 snapshot requiring periodic re-evaluation.
- **Binary criteria:** Many scoring criteria are yes/no, limiting nuance (e.g., partial FDA engagement vs. full clearance).

### Proposed future directions:

1. Multi-evaluator validation with inter-rater reliability assessment
2. Dynamic scoring querying live APIs, repositories, and regulatory databases
3. Expanded categories: AMRs, surgical navigation systems, telepresence robots
4. Context-specific weight profiles (research-focused vs. deployment-focused)
5. Clinical validation with oncology trial sites
6. International regulatory adaptation (EU MDR, PMDA Japan, NMPA China)

## 6 Conclusion

The Unification Standard Level (USL) provides the first standardized, quantitative framework for evaluating physical AI robot readiness for multi-site oncology clinical trials. Scoring nine robots across three categories on four dimensions reveals:

1. **Open-source ecosystem maturity is the strongest predictor of unification readiness.** Franka Panda (7.4) and da Vinci dVRK (7.1) lead their categories because of their large, active open-source communities.
2. **Clinical trial readiness is the weakest dimension field-wide.** Seven of nine robots score lowest on Dim D—regulatory compliance, federated learning, and multi-site deployment infrastructure are primary bottlenecks.
3. **Hardware quality alone does not determine unification readiness.** Robots with superior hardware (xArm IP51, Optimus 22-DOF hands) score low with immature software ecosystems.
4. **Category-specific scoring is essential.** Surgical robots require tissue deformation and FDA criteria; humanoids require locomotion and navigation safety; cobots require manipulation precision and lab automation criteria.

The complete USL framework—scoring engines, robot evaluation modules, text diagrams, and documentation—is open-source at <https://github.com/kevinkawchak/physical-ai-oncology-trials> under the MIT license, enabling the community to evaluate additional robots, refine scoring criteria, and drive progress toward unified, multi-site physical AI oncology clinical trials.

## References

- [1] John C. Mankins. Technology readiness assessments: A retrospective. White paper, NASA, 2004. Original 9-level TRL scale for evaluating technology maturity.
- [2] Alexander Lavin, Ciarán M. Gilligan-Lee, Alessya Visber, Sharath Guntuku, Thomas Olovsson, Miroslav Loncarevic, Cosmin Paduraru, Andrei Georgescu, Wilfried Senn, Julia Stout, et al. Technology readiness levels for machine learning systems. GitHub: <https://github.com/ai-infrastructure-alliance/mltrl>, 2021. Extends TRL to ML systems covering data readiness, model development, deployment, and monitoring.
- [3] Katja Tomaschek, Alison Olechowski, Steven Eppinger, and Nitin Joglekar. A survey of technology readiness level users. In *Proceedings of PICMET 2015: Portland International Center for Management of Engineering and Technology*, 2015. Identifies challenges in applying TRL to integrated multi-technology systems.
- [4] Kevin Kawchak. Physical AI for clinical oncology trials, 2025. Recommends LLM usage for upcoming oncology trials.
- [5] Kevin Kawchak. End-to-end physical AI unification of oncology clinical trials, 2026. Repository: <https://github.com/kevinkawchak/physical-ai-oncology-trials>.
- [6] NVIDIA Corporation. NVIDIA isaac lab. <https://github.com/isaac-sim/IsaacLab>, 2024. Version 2.3.1. GPU-accelerated robot learning framework.
- [7] Google DeepMind. MuJoCo: Multi-joint dynamics with contact. <https://github.com/google-deepmind/mujoco>, 2024. Version 3.4.0. Physics simulation engine.
- [8] CrewAI Inc. CrewAI: Multi-agent orchestration framework. <https://github.com/crewAIInc/crewAI>, 2026. Version 1.6.1.
- [9] LangChain AI. LangGraph: Stateful multi-actor llm applications. <https://github.com/langchain-ai/langgraph>, 2026. Version 1.1.0.
- [10] Linux Foundation AI and Data Foundation. Model context protocol. <https://modelcontextprotocol.io/>, 2025. Agent-tool communication standard.



- [11] NVIDIA Corporation. GR00T N1: Humanoid robot foundation model. <https://developer.nvidia.com/isaac/humanoid>, 2026. Version N1.6.
- [12] ISO 13482:2014 — robots and robotic devices — safety requirements for personal care robots, 2014.
- [13] IEC 62304:2006+AMD1:2015 — medical device software — software life cycle processes, 2015.
- [14] IEC 80601-2-77:2019 — particular requirements for robotically assisted surgical equipment, 2019.
- [15] Google DeepMind. MuJoCo menagerie: Curated robot models. [https://github.com/google-deepmind/mujoco\\_menagerie](https://github.com/google-deepmind/mujoco_menagerie), 2024. Standardized MJCF robot models.
- [16] Johns Hopkins University. dVRK: da vinci research kit. <https://github.com/jhu-dvrk/sawIntuitiveResearchKit>, 2026. Version 2.4.0. Open-source research platform for da Vinci surgical systems.
- [17] Stanford and JHU. ORBIT-Surgical: Gpu-accelerated surgical robot learning. <https://github.com/orbit-surgical/orbit-surgical>, 2024. Isaac Lab-based surgical task benchmarking.
- [18] MedAIR Lab. SurRoL: Surgical robot learning benchmarks. <https://github.com/med-air/SurRoL>, 2024. PyBullet-based surgical robot learning.
- [19] WPI-AIM Lab. AMBF: Asynchronous multi-body framework. <https://github.com/WPI-AIM/ambf>, 2024. Tissue deformation simulation for surgical robotics.
- [20] Boston Dynamics. Atlas electric: A new era for atlas. <https://bostondynamics.com/blog/electric-new-era-for-atlas/>, 2024. Next-generation electric humanoid.
- [21] Toyota Research Institute. Drake: Model-based design and verification for robotics. <https://github.com/RobotLocomotion/drake>, 2024. Open-source simulation and planning framework.
- [22] Agility Robotics. Digit: The first commercial humanoid robot. <https://agilityrobotics.com/robots>, 2024. Commercial warehouse and logistics humanoid.
- [23] U.S. Food and Drug Administration. Marketing submission recommendations for a predetermined change control plan for AI/ML-enabled device software functions. FDA Guidance Document, August 2025.

## Acknowledgments

The author would like to acknowledge Anthropic for providing access to Claude Code, Claude Cowork; and OpenAI for providing access to ChatGPT.

## Ethical Disclosures

The author of the article declares no competing interests.

## Rights and Permissions

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and source are properly credited, a link to the Creative Commons license is provided, and any modifications made are indicated. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>.

## Cite This Article

Kawchak K. Unification Standard Level for Physical AI Oncology Trial. Zenodo. 2026; [10.5281/zenodo.18778220](https://doi.org/10.5281/zenodo.18778220).