

Advancements in Artificial Intelligence for Real-World Problem Solving: Foundations, Methods, and Applications

Meera DC

DOI: <https://doi.org/10.51583/IJLTEMAS.2025.1410000159>

Abstract: This paper surveys recent advancements in artificial intelligence (AI) that have directly improved the capability of systems to solve real-world problems. We review progress in foundation models and multimodal systems, generative models (diffusion and transformer families), human-in-the-loop alignment (RLHF), and privacy/resilience techniques for deploying AI at the edge (federated/TinyML). Building on the literature, we identify important gaps in robustness, evaluation, and societal alignment, then propose a methodology combining multimodal pretraining, task-specific fine-tuning with human feedback, and privacy-preserving edge deployments to address practical tasks in healthcare triage, environmental monitoring, and robotics. Experimental designs, datasets, metrics, and ethical safeguards are provided to enable reproducible, responsible research.

I. Introduction

AI capability has advanced rapidly due to architectural innovations (transformers), scaling laws, and large-scale pretraining, producing models that generalize across tasks and modalities. These advances enable practical solutions in image/video analysis, natural language understanding, and embodied systems (robotics/autonomous agents). However, real-world deployment raises challenges in robustness, data privacy, alignment to human values, and resource constraints at the edge. This paper synthesizes the state of the art and proposes research directions to close gaps between lab progress and reliable, practical systems.

II. Literature Review — Key Recent Advances

Multimodal & Foundation Models

Large language models extended to handle multiple modalities (images, video, sensor data) have shown emergent capabilities—performing tasks that require cross-modal reasoning and enabling richer human–AI interaction. Surveys show MLLMs (multimodal large language models) are an active, high-impact area with rapid development.

Generative Models (Diffusion, Transformers, GANs)

Diffusion models and scaled transformer variants now dominate high-fidelity generation (images, audio). These generative advances are crucial for data augmentation, simulation, and synthetic-data generation for low-resource domains. Comprehensive reviews summarize theoretical progress and applications.

Human-Centered Alignment (RLHF and Related Methods)

In tasks where objective reward functions are hard to specify, reinforcement learning from human feedback (RLHF) and instruction tuning have become central for aligning models with human preferences and safety constraints; surveys and practical guides outline the technique and open problems.

Edge, Privacy, and Federated Learning

Deploying AI in privacy-sensitive, resource-constrained environments (IoT, mobile health) motivates federated learning, TinyML, and compressed models that run on-device while preserving user privacy. Recent reviews examine integrating federated techniques with TinyML for practical edge deployments.

Embodied & Robotics Systems

Embodied multimodal models (e.g., models combining language and sensory inputs) demonstrate transfer to robotics tasks, suggesting end-to-end pretraining can accelerate robotic perception and planning.

Research Gaps & Challenges

1. **Robustness & Distribution Shift:** Models often fail when domain shifts occur; generalization guarantees are limited.
2. **Evaluation:** Benchmarks can be narrow; real-world performance needs richer, task-specific metrics.
3. **Alignment & Safety:** RLHF reduces some failure modes, but is expensive and can encode biases.
4. **Privacy & Resource Constraints:** Large models are costly to run and raise privacy concerns for sensitive data.
5. **Reproducibility & Sim-to-Real Transfer:** Simulated training often does not transfer cleanly to physical systems (robots, sensors).

Objectives

- Survey and synthesize the most promising technical advances that enable real-world problem solving.
- Propose and evaluate a modular pipeline that merges multimodal foundation models, RLHF alignment, and federated edge deployment to solve concrete tasks.
- Measure gains in robustness, privacy preservation, and real-world utility across three application domains: healthcare triage, environmental monitoring, and robotic manipulation.

III. Proposed Methodology

System Overview

Pretraining & Foundation Component: Start from a multimodal pretrained model (text + vision; optionally sensors). Use contrastive and generative pretraining objectives to obtain a versatile backbone.

Task Adaptation: Fine-tune with a combination of supervised data and synthetic data generated by diffusion/transformer generative pipelines for data augmentation.

Human-in-the-Loop Alignment: Apply RLHF or preference learning to align outputs to domain experts (e.g., clinicians for triage). Use smaller, specialist annotator pools and active learning to reduce labeling cost.

Privacy/Edge Deployment: Use federated fine-tuning and model distillation into TinyML footprints to allow on-device inference and privacy preservation. Aggregate model updates via secure aggregation.

Experimental Tasks & Datasets

Healthcare Triage: multimodal clinical notes + imaging (public datasets: MIMIC-CXR, CheXpert; synthetic augmentation when necessary).

Environmental Monitoring: remote sensing images (Sentinel), sensor time series (air quality), anomaly detection via foundation-model features.

Robotics / Embodied Tasks: simulation environments (e.g., Habitat, MuJoCo) for training; real robot transfer tests using PaLM-E-style embodied architectures.

Evaluation Plan

- **Utility Metrics:** task accuracy/F1, AUC, and domain-specific clinical/operational metrics.
- **Robustness Tests:** distribution shift (corruptions, adversarial), calibration metrics, OOD detection.
- **Privacy/On-Device Metrics:** communication cost, model size, latency, federated convergence rate, membership-inference risk.
- **Human Alignment:** human preference win-rate, qualitative error analysis, and annotation cost per improvement.

Experiments (Suggested)

Ablation of multimodal pretraining vs. single-modality baselines on downstream tasks.

RLHF vs. supervised fine-tuning: measure alignment gains and annotation budget tradeoffs.

Federated vs. centralized training: compare privacy metrics and model performance for on-device scenarios.

Sim-to-real transfer test for robotic tasks: evaluate a curriculum combining simulation pretraining and real-world fine-tuning.

Expected Contributions

- A unified, practical pipeline showing how foundation/multimodal models + RLHF + federated/TinyML techniques can improve real-world task performance while respecting privacy and resource constraints.
- Empirical benchmarks across three application areas demonstrating tradeoffs (accuracy vs. cost, alignment vs. data needs).
- Open-source code, models (distilled), and a reproducible experimental suite for community reuse.

Ethical, Legal & Social Considerations

- **Bias & Fairness:** continuously audit deployed systems for disparate impacts; use diverse expert panels during RLHF.
- **Privacy:** adopt differential privacy where feasible, and secure aggregation for federated updates.
- **Accountability:** maintain human-in-the-loop decision thresholds for high-risk domains (e.g., healthcare).

- **Transparency:** provide model cards and documentation about limitations and expected failure modes.

Limitations

- RLHF can be costly and may codify annotator biases.
- Federated setups require nontrivial engineering and may not be practical for all deployments.
- Sim-to-real transfer remains an active challenge—physical safety and cost limit large-scale robot trials.

IV. Conclusion

Integrating recent AI advances—multimodal foundation models, powerful generative models for augmentation, human-aligned fine-tuning (RLHF), and privacy-aware edge deployments—offers a promising path to practical, responsible AI for real-world problem solving. The proposed pipeline and experiments aim to quantify tradeoffs and provide reproducible building blocks that move research from lab benchmarks toward operational, trustworthy systems.