# Cross-Embodiment Generalization and Morphological Adaptation for Humanoid Robots: A Prompt Engineering Perspective

AI Research Assistant[1]

[1]Google Gemini

February 3, 2026

## Abstract

This paper explores the critical challenge of cross-embodiment generalization and morphological adaptation in humanoid robotics. While recent advances in imitation learning, exemplified by frameworks like HumanX, enable complex skill acquisition from human demonstrations, transferring these skills across robots with varying morphologies remains a significant hurdle. We propose a novel perspective that frames morphological adaptation as a prompt engineering problem, where descriptions of target robot kinematics, dynamics, and environmental contexts serve as "prompts" to guide skill transfer and refinement. We outline methodologies for developing morphologically-aware policy architectures, leveraging latent space representations, and applying meta-learning for rapid adaptation. The paper discusses the theoretical underpinnings, potential architectures, and evaluates the implications for creating truly versatile and adaptable humanoid robots capable of operating effectively across diverse physical embodiments with minimal re-training.

## 1 Introduction

The field of humanoid robotics is rapidly advancing, with robots demonstrating increasingly sophisticated locomotion and interaction skills in complex environments. A cornerstone of this progress is imitation learning, where robots acquire behaviors by observing human demonstrations. Recent frameworks, such as HumanX [1], have shown remarkable success in synthesizing physically plausible interaction data from human videos and learning generalizable skills for specific humanoid platforms. However, a fundamental limitation persists: the learned policies are often highly specialized for the robot they were trained on. Significant variations in kinematics, dynamics, or even minor changes in joint limits or

link lengths typically necessitate extensive re-training or manual re-engineering, hindering scalability and broad deployment.

This paper addresses the challenge of **cross-embodiment generalization and morphological adaptation**. Our central hypothesis is that by treating morphological descriptions and environmental contexts as structured "prompts," we can guide robotic learning systems to adapt existing skills to novel robot bodies more efficiently and effectively. Drawing inspiration from prompt engineering in Large Language Models (LLMs), where textual prompts steer model behavior, we conceptualize how parametric or descriptive "morphological prompts" can inform a robot's policy to adjust its actions for a new physical form.

The contributions of this paper are:

- To frame morphological adaptation as a prompt engineering problem, highlighting the conceptual parallels with guiding LLMs.

- To propose theoretical frameworks and architectural considerations for developing morphologically-aware robotic policies.

- To discuss methodologies for learning robust, generalizable skill representations that can be adapted via morphological prompts.

- To identify key challenges and outline future research directions towards creating truly adaptable humanoid robots.

## 2 Related Work

### 2.1 Imitation Learning and Humanoid Control

Imitation learning has emerged as a powerful paradigm for teaching complex behaviors to robots. Works like [2, 3] demonstrate learning agile locomotion and manipulation. Specifically, HumanX [1] showcases a full-stack framework for acquiring diverse interaction skills from human videos, emphasizing data generation and imitation learning. While these methods achieve high fidelity for a target robot, they often lack explicit mechanisms for cross-embodiment transfer.

### 2.2 Generalization and Adaptation in Robotics

Research in domain randomization [4] and sim-to-real transfer [5] aims to make policies robust to variations within a single robot's operational envelope. Meta-learning approaches [6, 7] have explored rapid adaptation to new tasks or environmental dynamics. However, explicit morphological adaptation, especially for significant structural changes, remains an active area of research, often relying on learning a separate policy for each morphology [8] or limiting adaptation to minor parameter variations.

## 2.3 Parametric Policies and Morphological Awareness

Some approaches design policies that explicitly take robot parameters as input [9], allowing for continuous variations in morphology. Graph Neural Networks (GNNs) have also been used to represent robot structures [10], offering a more flexible way to encode morphology. Our work builds upon these ideas by integrating a "prompting" mechanism to guide this parameterization and adaptation.

## 2.4 Prompt Engineering in Large Language Models

Prompt engineering [11] is a technique for optimizing the input to LLMs to achieve desired outputs. It involves crafting specific instructions, examples, or contextual information within the prompt. We draw a conceptual parallel, where a "morphological prompt" encodes information about the target robot's physical characteristics, guiding the robot's policy to adapt its learned skills.

# 3 Problem Formulation: Morphological Adaptation as a Prompting Task

We define a robot's morphology by a set of parameters $\mathcal{M} = \{\mathbf{k}, \mathbf{d}, \mathbf{j}\}$, where $\mathbf{k}$ represents kinematic parameters (e.g., link lengths, joint limits), $\mathbf{d}$ represents dynamic parameters (e.g., mass, inertia), and $\mathbf{j}$ represents joint properties (e.g., stiffness, damping). A skill $\mathcal{S}$ is a desired behavior, typically defined by a trajectory or a goal-oriented objective.

Let $\pi_\theta(\mathbf{a}_t|\mathbf{s}_t, \mathbf{p})$ be a robot policy parameterized by $\theta$, which outputs actions $\mathbf{a}_t$ given a state $\mathbf{s}_t$ and a "prompt" $\mathbf{p}$. In our context, the prompt $\mathbf{p}$ encapsulates information about the target robot's morphology $\mathcal{M}$ and potentially the task context $\mathcal{T}$.

The objective of morphological adaptation is to learn a policy $\pi_\theta$ such that for a given skill $\mathcal{S}$ learned on a source morphology $\mathcal{M}_{\text{source}}$, it can effectively execute $\mathcal{S}$ on a target morphology $\mathcal{M}_{\text{target}}$ by leveraging a morphological prompt $\mathbf{p}_{\mathcal{M}_{\text{target}}}$. This can be formally expressed as minimizing a loss function $L(\pi_\theta, \mathcal{S}, \mathcal{M}_{\text{target}})$ across a distribution of target morphologies:

$$\min_\theta \mathbb{E}_{\mathcal{M}_{\text{target}} \sim \mathcal{D}_\mathcal{M}} [L(\pi_\theta(\mathbf{a}_t|\mathbf{s}_t, \mathbf{p}_{\mathcal{M}_{\text{target}}}), \mathcal{S}, \mathcal{M}_{\text{target}})]$$

where $\mathcal{D}_\mathcal{M}$ is the distribution of possible robot morphologies. The prompt $\mathbf{p}_{\mathcal{M}_{\text{target}}}$ could be a vector encoding of $\mathcal{M}_{\text{target}}$, a learnable embedding, or even a textual description processed by an auxiliary encoder.

# 4 Proposed Methodologies for Morphological Adaptation

## 4.1 Morphologically-Aware Policy Architectures

Instead of learning a separate policy for each robot, we propose designing policies that are explicitly conditioned on morphological parameters.

- **Direct Parameter Input:** The policy $\pi_\theta(\mathbf{a}_t|\mathbf{s}_t, \mathcal{M})$ takes the full set of morphological parameters $\mathcal{M}$ as an additional input. This requires the policy network to learn how to interpret and utilize these parameters to adjust its control outputs.

- **Hypernetwork Approach:** A hypernetwork could generate the weights or biases for a base policy network, conditioned on the morphological prompt. This allows for more complex, non-linear adaptations of the policy structure itself.

$$\theta = H(\mathbf{p}_\mathcal{M})$$

  where $H$ is the hypernetwork.

## 4.2 Latent Space Learning for Skills and Morphologies

A powerful approach involves learning a shared latent space where both skills and morphologies are represented.

- **Morphological Embeddings:** An encoder $E_\mathcal{M}$ maps a raw morphology $\mathcal{M}$ into a compact latent vector $\mathbf{z}_\mathcal{M}$. This $\mathbf{z}_\mathcal{M}$ then serves as the morphological prompt.

$$\mathbf{z}_\mathcal{M} = E_\mathcal{M}(\mathcal{M})$$

- **Skill Embeddings:** Similarly, a skill encoder $E_\mathcal{S}$ can map a skill demonstration or description into a latent skill vector $\mathbf{z}_\mathcal{S}$.

- **Conditioned Policy:** The robot's policy then operates in this shared latent space: $\pi_\theta(\mathbf{a}_t|\mathbf{s}_t, \mathbf{z}_\mathcal{S}, \mathbf{z}_\mathcal{M})$. The training objective would encourage the policy to produce effective actions when conditioned on different combinations of skill and morphological embeddings. This can be conceptualized as learning a universal skill representation that can be "decoded" for any given robot morphology.

## 4.3 Meta-Learning for Rapid Adaptation

Meta-learning, or "learning to learn," is particularly well-suited for rapid adaptation to new morphologies.

- **Model-Agnostic Meta-Learning (MAML) [6]:** Train a policy's initial parameters such that it can quickly adapt to a new morphology with only a few gradient steps using limited online interaction data (or simulated data for the new morphology).

- **Contextual Meta-Learning:** The policy learns a context encoder that generates a task/morphology-specific embedding from a few adaptation trials, which then conditions the main policy. The morphological prompt could directly inform this context.

## 4.4 Prompt-Guided Data Augmentation (Inspired by XGen)

Leveraging the data generation capabilities similar to HumanX's XGen, we can augment training data to explicitly cover a distribution of morphologies.

- **Morphological Randomization in Simulation:** During training, simulate skills across a wide range of randomly sampled morphologies. The sampled morphological parameters are then fed into the policy as a "prompt."

- **Prompt-Driven Synthesis:** Given a morphological prompt (e.g., "a robot with 10

# 5 Experimental Setup and Evaluation (Conceptual)

To evaluate the proposed methodologies, a comprehensive experimental setup would be required, ideally leveraging high-fidelity physics simulators like Isaac-Gym [12] (as used in HumanX) to generate diverse morphological variations.

## 5.1 Benchmark Tasks

- **Locomotion:** Walking, running, navigating uneven terrain with varying robot leg lengths, joint strengths, or mass distributions.

- **Manipulation:** Grasping and manipulating objects with different arm lengths, gripper types, or degrees of freedom.

- **Interactive Skills:** Adapting skills like basketball shooting or object passing (as in HumanX) to robots with different arm spans, body masses, or reaction forces.

## 5.2 Metrics

- **Skill Success Rate (SR):** Percentage of successful task completions for novel morphologies.

- **Adaptation Efficiency:** Number of adaptation steps or amount of data required to achieve a target SR on a new morphology.

- **Generalization Gap:** Difference in performance between training morphologies and unseen test morphologies.

- **Robustness to Morphological Perturbations:** How well the policy performs under small, continuous variations around a target morphology.

- **Physical Plausibility/Stability:** Metrics for maintaining balance and executing physically realistic motions.

## 5.3 Levels of Morphological Change

Experiments should cover:

- **Parametric Variations:** Small changes to existing kinematic or dynamic parameters (e.g., $\pm 10\%$ link length).

- **Structural Variations:** More significant changes like adding or removing degrees of freedom, or altering the fundamental topology of a limb.

# 6 Challenges and Future Directions

## 6.1 The Sim-to-Real Gap for Diverse Morphologies

Transferring policies learned in simulation to real robots is challenging, and this gap can be exacerbated when dealing with a wide range of morphologies. Accurate simulation of varying kinematics and dynamics is crucial.

## 6.2 Rich and Diverse Datasets

Training policies that generalize across morphologies requires datasets that cover a broad distribution of robot forms and corresponding skill executions. Generating such datasets, especially for real robots, is resource-intensive. Prompt-guided data augmentation in simulation can mitigate this.

## 6.3 Interpretability of Morphological Prompts

As prompts become more complex (e.g., natural language descriptions), understanding how the policy interprets and acts upon this information becomes important for debugging and safety.

## 6.4 Beyond Static Morphologies

Future work could explore adaptation to dynamically changing morphologies (e.g., reconfigurable robots) or even co-designing robot morphology and control policies.

## 6.5 Hierarchical Prompting

Investigate hierarchical prompting where high-level prompts define the overall robot type, and low-level prompts fine-tune specific parameters.

# 7 Conclusion

This paper has proposed a novel perspective on morphological adaptation for humanoid robots, framing it as a prompt engineering problem. By designing policies that are explicitly conditioned on morphological descriptions, we can move towards more versatile and adaptable robotic systems. We have outlined key methodologies including morphologically-aware architectures, latent space learning, meta-learning, and prompt-guided data augmentation, drawing inspiration from the success of imitation learning (like HumanX) and prompt engineering in LLMs. Addressing the challenges in data generation, sim-to-real transfer, and robust adaptation will pave the way for a new generation of humanoid robots capable of seamlessly transferring and executing skills across a diverse range of physical embodiments, significantly advancing the field of intelligent robotics.

# References

[1] Yinhuai Wang, Qihan Zhao, Yuen Fui Lau, Runyi Yu, Hok Wai Tsui, Qifeng Chen, Jingbo Wang, Jiangmiao Pang, Ping Tan. HumanX: Toward Agile and Generalizable Humanoid Interaction Skills from Human Videos. *arXiv preprint arXiv:2602.02473*, 2026. https://arxiv.org/pdf/2602.02473v1

[2] Xue Bin Peng, Pieter Abbeel, Sergey Levine, and Michiel van de Panne. Deepmimic. *ACM Transactions on Graphics*, 37:1–14, Aug 2018.

[3] Tairan He, Jiawei Gao, Wenli Xiao, Yuanhang Zhang, Zi Wang, Jiashun Wang, Zhengyi Luo, Guanqi He, Nikhil Sobanbab, Chaoyi Pan, et al. Asap: Aligning simulation and real-world physics for learning agile humanoid whole-body skills. *arXiv preprint arXiv:2502.01143*, 2025.

[4] Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. Domain randomization for transferring deep neural networks from simulation to the real world. In *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 23–30. IEEE, 2017.

[5] F. Sadeghi, A. Lee, D. Stanford, A. O. Alden, K. R. S. Bhat, and S. Levine. Cad2rl: Real single-image flight without a single real image. *Robotics and Automation Letters*, 3(4):3020-3026, 2018.

[6] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. *Proceedings of the 34th International Conference on Machine Learning*, 2017.

[7] A. Nagabandi, I. Konogliev, S. Levine, and A. Singh. Learning to Adapt in Dynamic, Real-World Environments Through Meta-Reinforcement Learning. *International Conference on Learning Representations*, 2019.

[8] Z. Li, X. B. Peng, P. Abbeel, S. Levine, G. Berseth, and K. Sreenath. Reinforcement learning for robust parameterized locomotion control of bipedal robots. *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2811-2817, 2021.

[9] S. M. Eslami, N. Heess, Y. Tassa, and J. Schulman. Continuous control with deep reinforcement learning. *International Conference on Learning Representations*, 2016.

[10] F. Locatello, S. O. Alsallakh, and A. Schwing. Graph-based representation for learning about 3D articulated objects. *Advances in Neural Information Processing Systems*, 32, 2019.

[11] L. Reynolds and M. McDonell. Prompt Engineering for Large Language Models. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 2021.

[12] Viktor Makoviychuk, Lukasz Wawrzyniak, Yunrong Guo, Michelle Lu, Kier Storey, Miles Macklin, David Hoeller, Nikita Rudin, Arthur Allshire, Ankur Handa, and Gavriel State. Isaac gym: High performance GPU based physics simulation for robot learning. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021.