AI for Robotics

Toward Embodied and General Intelligence in the Physical World

Alishba Imran
Keerthana Gopalakrishnan

Apress®

AI for Robotics

Toward Embodied and General Intelligence in the Physical World

Alishba Imran
Keerthana Gopalakrishnan

Apress®

AI for Robotics: Toward Embodied and General Intelligence in the Physical World

Alishba Imran
San Francisco, CA, USA

Keerthana Gopalakrishnan
San Francisco, CA, USA

ISBN-13 (pbk): 979-8-8688-0988-0
ISBN-13 (electronic): 979-8-8688-0989-7
https://doi.org/10.1007/979-8-8688-0989-7

Managing Director, Apress Media LLC: Welmoed Spahr
Acquisitions Editor: Celestin Suresh John
Development Editor: Laura Berendson
Coordinating Editor: Kripa Joseph
Copy Editor: Kezia Endsley

Cover designed by eStudioCalamar

Distributed to the book trade worldwide by Springer Science+Business Media New York, 1 New York Plaza, New York, NY 10004. Phone 1-800-SPRINGER, fax (201) 348-4505, e-mail orders-ny@springer-sbm.com, or visit www.springeronline.com. Apress Media, LLC is a Delaware LLC and the sole member (owner) is Springer Science + Business Media Finance Inc (SSBM Finance Inc). SSBM Finance Inc is a **Delaware** corporation.

For information on translations, please e-mail booktranslations@springernature.com; for reprint, paperback, or audio rights, please e-mail bookpermissions@springernature.com.

Apress titles may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Print and eBook Bulk Sales web page at http://www.apress.com/bulk-sales.

Any source code or other supplementary material referenced by the author in this book can be found here: https://www.apress.com/gp/services/source-code.

If disposing of this product, please recycle the paper

Dedicated to my father,
for his unconditional love and unwavering support

—Keerthana

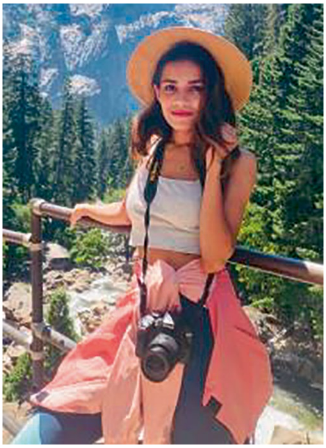
Dedicated to my family and friends who have
supported me on this journey

—Alishba

Table of Contents

About the Authors

Alishba Imran is a machine learning and robotics developer specializing in robot learning for manipulation and perception. She is currently conducting research in reinforcement learning and unsupervised learning with Pieter Abbeel at the Berkeley AI Research Lab. Her past experience spans across top robotics companies—she's worked on advanced perception systems at Cruise, developed simulation-based manipulation methods at NVIDIA, and led impactful initiatives to reduce the cost of prosthetics. At Hanson Robotics—the creators of Sophia the Robot, the world's most advanced human-like robot—Alishba co-led neuro-symbolic AI research and developed low-cost hardware system for humanoids.



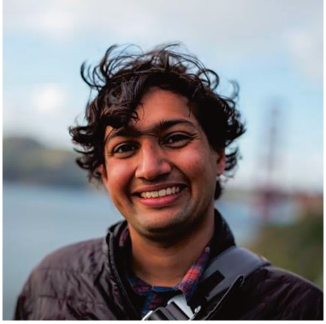
Keerthana Gopalakrishnan is a Senior Research Scientist at Google DeepMind working on robot manipulation and the Gemini project. She was educated at Carnegie Mellon University and the Indian Institute of Technology. Her research concerns large language models for robotic planning, scaling visual language models for low-level control, and cross-embodiment robot learning.

About the Technical Reviewers

Lentin Joseph is the author of ten books in ROS and the cofounder/CTO of RUNTIME Robotics from India/Kerala. He is an accomplished robotics engineer with 14 years of experience in the field. He specializes in the Robot Operating System (ROS), contributing significantly to the development of various robotics applications and solutions. He holds a degree in electronics and communication engineering and an M.Tech in robotics and automation, from which he graduated with honors. Lentin has authored and reviewed more than ten books on ROS, including *Mastering ROS for Robotics Programming, ROS Robotics Projects,* and *Robot Operating System for Absolute Beginners,* making significant contributions to the robotics community.

LinkedIn profile: `https://www.linkedin.com/in/lentinjoseph/`

Company website: `https://runtimerobotics.com/`



Sean Kirmani is a research scientist at Google DeepMind working on problems in vision, language, and action. He was also the technical lead for semantic perception at The Everyday Robot Project within Google[x]. He holds degrees in electrical engineering and computer science from The University of Texas at Austin, where he worked in several robotics labs. Sean has co-authored over 20 publications in robotics and AI.

Acknowledgments

We'd like to acknowledge Chris Paxton for his valuable contribution to the Reinforcement Learning chapter of this book. We'd like to thank the robotics research community for allowing us to tell the story of this field, which was collectively built, and for graciously permitting us to feature their work in this book. We are also grateful for Marco Mascorro and Vivek Aithal for reviews on earlier versions of this book.

Introduction

What This Book Is About

AI for Robotics is the reimagination of robotics as an artificial intelligence problem. Modern robotics is steadily transformed by breakthroughs in AI. This book is your comprehensive guide to framing traditional robotics problems as AI problems and approaching them with deep learning techniques. Whether you're a coder, an enthusiast, or an investor, *AI for Robotics* gives you the blueprint to create generalizable and data-driven robotic intelligence that learns, evolves, and tackles challenges we once thought impossible in dynamic, real-world environments.

Who This Book Is For

This book empowers:

- **Software and AI engineers:** If you have a background in machine learning but are new to robotics, this book bridges the gap, showing you how to build robots that learn and adapt.
- **Robotics and mechanical engineers:** Stay ahead of the curve by learning how to integrate AI and data-driven approaches into your designs, ensuring your robots are at the forefront of innovation.
- **Investors, executives, and decision-makers:** Gain a clear understanding of the AI-robotics landscape. Make informed choices about which technologies to bet on.

No matter your background, if you're ready to shape the future of robotics, this book is your guide.

The Structure of the Book

This book is structured to gradually build your understanding of the use of artificial intelligence for robotics, starting with fundamental concepts and progressing to advanced applications.

- **Chapter 1: Introduction to General Purpose Robotics**

  Provides an overview of the current state and future directions of robotics, emphasizing the role of machine learning in enabling more versatile and intelligent systems.

- **Chapter 2: Robot Perception: Sensors and Image Processing**

  Covers the basics of how robots perceive their environment through sensors and image-processing techniques, focusing on learning representations for vision tasks.

- **Chapter 3: Robot Perception: 3D Data and Sensor Fusion**

  Explores how to process and integrate 3D data from various sensors to create a coherent understanding of the robot's surroundings.

- **Chapter 4: Foundation Models in Robotics**

  Discusses the application of large, pretrained models in robotics, including language models and visual language models, and how they can be adapted for robotics.

- **Chapter 5: Simulation**

  Details the use of synthetic data and simulation environments for training and testing robots, including simulated-to-real transfer techniques.

- **Chapter 6: Mapping, Localization, and Navigation**

  Focuses on the techniques robots use to map the surroundings, understand where they are, and navigate their environments.

- **Chapter 7: Reinforcement Learning and Control**

  Introduces reinforcement learning and control strategies for teaching robots to self-improve and learn from trial and error.

- **Chapter 8: Self-Driving Vehicles**

  Explains the design, safety considerations, and technical challenges involved in building autonomous vehicles.

- **Chapter 9: Industrial Robotics**

  Covers the application of robotics in industrial settings, including manufacturing and warehouse automation, and the integration of machine learning to enhance these processes.

- **Chapter 10: Humanoid Robotics**

  Delves into the unique challenges and opportunities in developing humanoid robots, including perception, hardware, and software design.

- **Chapter 11: Data-Driven Robotics in Practice**

  Discusses the infrastructure required to support data-driven robotics, including important considerations, safety issues, and future directions.

What You Will Learn

By the end of this book, you'll gain expertise in the following:

- Applying machine learning to key robotics areas, including perception, mapping, control, and decision-making.
- Designing and implementing robotic systems for diverse industries, including self-driving cars, manufacturing, and humanoid robots.
- Overcoming the specific hurdles of integrating machine learning with robotics, understanding the future trends of robotics, and learning about the ongoing impact of machine learning.

What You Need to Know Before You Start

This book requires some Python programming knowledge and familiarity with libraries like NumPy, PyTorch/Jax, or ROS. A basic understanding of neural networks and machine learning is also necessary, either through an introductory course or self-study. If you lack this background, consider taking Andrew Ng's ML course or the Deep Learning Specialization on Coursera.

Why We Wrote This Book

The last half decade has shown robotics being disrupted by machine learning methods and evidence is stronger than ever that the path to building generally intelligent robots is paved heavily with AI. We believe that the field of robotics is at a special moment today: one that is ready to be disrupted by breakthroughs in AI research. Machine learning has fundamentally transformed how we design and build robots, opening up a world of possibilities to create intelligent machines that effortlessly navigate and interact with our complex world.

Innovation at a rapid pace has created a gap in literature, where most textbooks on robotics taught at schools tread in classical methods and most ML textbooks rarely address embodied AI, therefore restricting the knowledge of designing data-driven robotics to privileged conferences, research labs, and academic papers. We are writing this book to democratize access to the practice and know-how of modern robotics.

We aim to break down barriers, making the fusion of AI and robotics comprehensible to a broader audience and inspiring a new generation of roboticists. We write this book to spark innovation, ignite new ideas, and invite more people to contribute to this thrilling field.

It's time to learn the new robotics, where AI leads the way.

CHAPTER 1

Introduction to General Purpose Robotics

People have dreamt of making intelligent machines that behave and think like humans for centuries. From the industrial revolution to Asimov's "I, Robot" and the world's first humanoids built a century ago, robots have occupied our collective imagination for a long time. Robots have transcended from being a figment of science fiction to being realized in the present, with accelerating capabilities.

What led to this transformation? Advances in artificial intelligence have disrupted various industries in the last decade by unlocking new capabilities with machine learning. Robotics has escaped its constrained and narrow applications in well-structured industrial and research environments and is now integrated into our daily lives. Robots drive competition through automation in large-scale manufacturing[1], space and underwater exploration[2], agriculture[3], and healthcare[4], among other industries. In the future, we expect to see robots handling fine-manipulation tasks in industries, performing household chores in homes, and autonomously operating on public roads and in hospitals.

As the capability has increased in the last few decades, the cost has decreased. Over the past 30 years, the average cost of robotics has fallen by half in Consumer Price Index (CPI)-adjusted terms (after accounting

A. Imran and K. Gopalakrishnan, *AI for Robotics*,
https://doi.org/10.1007/979-8-8688-0989-7_1

for inflation), according to a recent McKinsey & Company report[5]. Costs have fallen even further in relation to their capabilities due to Moore's law, the ubiquity of GPUs, and the falling cost yet rising capacity of batteries and onboard computers. The widespread adoption of robots is motivated by increased economic expansion, the rising cost of human labor, the falling cost of robots, and the increase in their capabilities.

However, as robots move from research labs and constrained industrial settings to the real world, they face new challenges. Consider, for example, a household cleaning robot. This robot would have to engage in many tasks, including cleaning the floors, dusting counters, and washing dishes. To accomplish this, it must know how to:

- Traverse indoor environments while perceiving and avoiding obstacles.
- Handle fragile, soft, and sometimes heavy objects with irregular shapes, including objects it may never have encountered before.
- Manipulate scenes it may have never experienced before, since each home looks different, has different lighting, layouts, and so on.
- Reason about interactions with household objects, humans, and pets and past configurations of the space.

The challenge is designing an approach that can adapt to changes in the real world and the variety of situations it will encounter. Before the advent of deep learning, a software stack to solve any of these tasks would be written as a state machine with "hard-wired" motion primitives resembling traditional controls for that particular task. This approach cannot handle unseen situations very well, doesn't scale, and limits the utility of hard-programmed robots. Additionally, even for simple pick-and-place, translating the wide repository of human intuitions into transitional controls is challenging, if not close to impossible.

The breadth and universality of perception, reasoning, and controls required for general-purpose robotics is best handled by universal function approximators: *neural networks.* Instead of hand-coding a control system, we use machine learning to allow a robot to learn the relevant features and their relationships from training data.

This approach has yielded results in many other areas. The recent success of ChatGPT and language models in general has minted multibillion dollar AI companies. The research behind these products shows that scaling data, compute, and models programmatically leads to general capabilities in the language/vision/audio spaces. By converting data into tokens, similarities between them can be identified, and those similarities can be transferred to other domains. A wide array of capabilities have been unlocked as a result:

1. Creation of custom and realistic images[37] and videos[38] on demand, which promises to transform the film, marketing, and advertising industries.

2. Language generation, including translations, creative writing, copywriting, code generation[41], and transcription[40].

3. Audio generation. Creating on-demand podcasts[42] and music[43].

4. Multimodal reasoning. Solving mathematics[45], graduate-level science problems, and law and medical problems[44].

These capabilities have led to an AI spring, with generative AI companies raising 25.9 billion dollars in funding in 2023 alone[39]. These trends, as well as recent breakthroughs in spatial intelligence and robot foundation models, show that robotics can also be framed and solved as an AI problem. In an era in which we have generic intelligence, the ingredients needed to build generally useful robots are mostly present.

The success of AI and the promise of emergent capabilities has led to a boon in machine learning powered robotics and a rising demand for talent in the labor market. According to research from *Mordor Intelligence*[6], the Global Robotics Market was valued at USD 27.73 billion in 2020 and is expected to reach USD 74.1 billion by 2026, registering a Compound Annual Growth Rate (CAGR) of 17.45 percent. While this increase mostly accounts for the boon in industrial robots, the AI robotics market is expected to grow at a CAGR of 38.6 percent from 6.9 billion USD in 2021 to 35.3 billion USD in 2026 according to this[7] report. To capitalize on this opportunity, large tech companies, startups, and research labs are increasingly seek qualified AI and robotics engineers for their robotics R&D, autonomous cars R&D, and manufacturing divisions. To start contributing to these companies' machine learning efforts, you'll need to understand:

- How to formulate a robotics problem in the context of machine learning
- Which machine learning methods can be used to solve different problems in robotics and the tradeoffs between them
- At what point in the robotics stack you should use machine learning

Moravec's paradox is one of the main challenges of machine learning for robotics. As Steven Pinker described in 1994[8], "The main lesson of 35 years of AI research is that the hard problems are easy and the easy problems are hard." Artificial intelligence, especially neural nets, is a fairly different form of intelligence than the human brain and, as such, has different strengths. Things that may seem very difficult for humans—such as generative imaging, language compression, and sequential projection like stock analytics—are quite easy for AI. However, tasks that even a four year old child can do easily via sensorimotor and perceptual reasoning—such as taking a walk and lifting a pencil—are much harder. In the history of scientific innovation, all problems seem hard before they are solved, and the authors of this book are optimistic that mapping and fixing the real challenges in robotic learning can put a dent in advancing physical intelligence.

This chapter starts with defining the two premises of this book: robots and AI. We present general motivations for why one needs to use AI for robotics and the challenges in doing so. Subsequent chapters map out key areas in the development of AI for robotics, such as machine learning perception, language in robotics, training robots in simulations, and building infrastructures for scalable robot learning. Then we explain how to practically design and implement these principles in a few select applications—self-driving cars, industrial robots, and humanoids.

Let's get into it!

A Robot System

A robot is defined as an interactive machine that takes in a world model and outputs actions. Unlike many machine learning applications, a robot is characterized by agency and a closed loop feedback in a real or simulated world.

A robot typically senses the world through its suite of sensors, including cameras, LiDAR, inertial measurement units (IMU), voice detectors, and/or radars, as a few examples. A robot brain, typically executing on an onboard microcontroller, processes the inputs from sensors and calculates actions, which are sent as signals to the robot's actuators. These actuators can be direct current motors that cause its joints to move or compliant materials in the case of soft robotics.

The action space of a robot is determined by its application. For example, the action space for self-driving cars and navigating robots is the acceleration and steering angle. A robot arm could be designed as positions or velocities of the joints on the arm. Additional action spaces for robots that interact with humans could include natural language via a chat interface, gestures, and facial expressions.

Figure 1-1 shows a high-level diagram of a robot system.

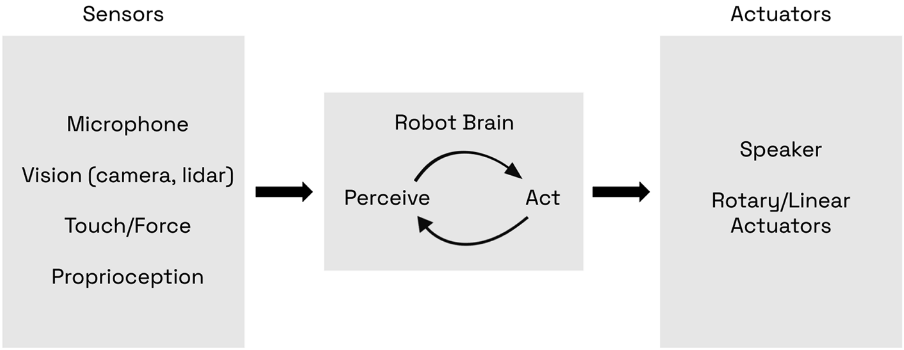


Figure 1-1. *The main components of a robot: sensors (microphones, vision systems like cameras and LiDAR, touch/force sensors, and proprioception), which are used to perceive the environment; the robot brain where perception data is processed in a continuous perception-action loop; and actuators (speakers and rotary/linear actuators), which carry out the robot's actions. This perception-action loop is critical for robotic learning*

Common Types of Robots

Robots come in many sizes and shapes. We can segregate them by vertical (or the sector in which they're deployed), as shown in Table 1-1 (curated with assistance from AI).

Table 1-1. *Types of Robots by Vertical*

	Type	Definition	Examples Industrial robots
1.	Industrial robots	Robots used in manufacturing processes such as assembly, painting, welding, and packaging	Robotic arms, gantry robots Service robots
2.	Service robots	Robots that perform tasks to assist humans in various environments such as hospitals, hotels, and restaurants	Delivery robots, cleaning robots, telepresence robots
3.	Medical robots	Robots used in healthcare settings to assist with surgeries, diagnostics, and patient care	da Vinci surgical system, rehabilitation robots, pharmacy automation systems
4.	Military and defense robots	Robots designed for use in military applications, such as reconnaissance, surveillance, and combat support	Unmanned aerial vehicles (UAVs), unmanned ground vehicles (UGVs), bomb disposal robots
5.	Agricultural robots	Robots used in farming to automate tasks like planting, harvesting, and monitoring crop health	Autonomous tractors, drones for crop monitoring, fruit-picking robots
6.	Domestic robots	Robots designed for use in homes to help with chores and other tasks	Roomba vacuuming robots, lawn-mowing robots, personal assistant robots like Jibo

(*continued*)

Table 1-1. (*continued*)

	Type	Definition	Examples Industrial robots
7.	Educational robots	Robots used in educational settings to help teach various subjects or skills	LEGO Mindstorms, social robots like Pepper, Sphero
8.	Research robots	Robots used in scientific research, including exploring remote or hazardous environments and developing new robotic technologies	Underwater robots, Mars Rovers, humanoid robots like ASIMO
9.	Entertainment robots	Robots designed for amusement or companionship	Robotic pets like Aibo, interactive toys like Furby, robots used in theme parks or movies
10.	Swarm robots	Robots that work together in large groups, coordinating their actions to complete tasks more efficiently	Swarm robots used in research, agriculture, search and rescue, and environmental monitoring

A second way to split robots is by the nature of their embodiment, as shown in Table 1-2 (curated with assistance from AI).

Table 1-2. *Types of Robots by Embodiment*

Serial Num	Embodiment	Explanation	Examples
1	Wheeled robots	Robots using wheels for locomotion, often used on flat surfaces	Roomba, TurtleBot, self-driving cars like Waymo and Cruise
2	Tracked robots	Robots utilizing tracks for movement, providing greater traction and stability on rough or uneven terrain	Mars rovers, bomb disposal robots
3	Legged robots	Robots using legs for locomotion, navigating complex environments like stairs and uneven terrain	Boston Dynamics' Spot, ASIMO
4	Flying robots	Robots capable of flight, typically using rotors or wings, for aerial surveillance, inspection, and photography.	Quadcopter drones, fixed-wing UAVs
5	Underwater robots	Robots designed for underwater operation, used for exploration, inspection, and monitoring tasks	Bluefin Robotics AUV, SeaBED
6	Snake robots	Robots with long, flexible bodies, for moving through tight spaces and navigating around obstacles	CMU's Biorobotics Lab's snake robots, OC Robotics' snake-arm robots

(*continued*)

Table 1-2. (*continued*)

Serial Num	Embodiment	Explanation	Examples
7	Robotic arms	Robots consisting of a series of joints and links, resembling a human arm, used in industrial settings	KUKA, Fanuc, and ABB robotic arms
8	Humanoid robots	Robots with human-like forms, used in research, entertainment, and service applications	SoftBank Robotics' Pepper, Hanson Robotics' Sophia
9	Soft robots	Robots that mimic locomotion mechanisms of deformable matter such as fluids, gels, and elastomers for greater flexibility. Commonly used in biomedical applications such as soft tools for surgery, rehabilitation devices, and drug delivery	Harvard's Wyss Institute's soft robots, Octobot

Despite the variety in robots, they share many similarities that can be used to build a common framework and science for robotics, which is extensible with modifications to fit the deployment conditions of a robot.

Common Concepts in Robot Design

This section explains a few ubiquitous concepts that are used in robot design.