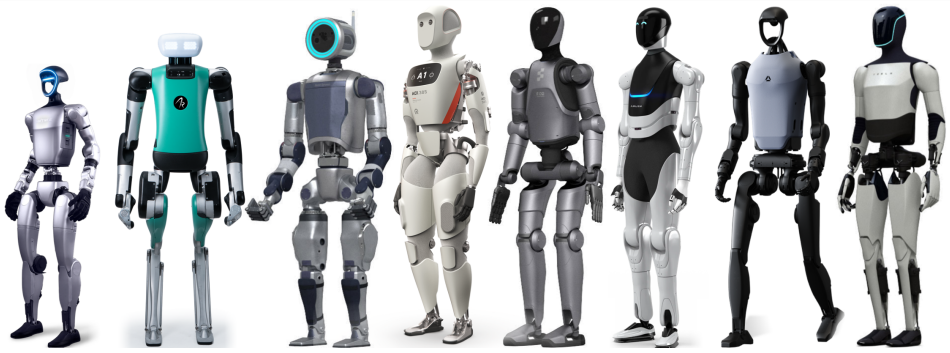
White Paper – Regionale Einsatzpotenziale humanoider Robotik

D. Bank[a,*], R. Kühn[a], J. Cordes[a], S.F.G. Ehlers[a], M. Schappler[a], T. Seel[a]

[a]*Institut für Mechatronische Systeme, Leibniz Universität Hannover, An der Universität 1, Garbsen, 30823, Lower Saxony, Germany*

Kurzfassung

Humanoide Robotik hat in den letzten Jahren bedeutende Fortschritte gemacht, sowohl auf Hardware- als auch auf Softwareebene. Während erste Modelle bereits in Serienproduktion gehen, entstehen kontinuierlich neue Entwicklungen durch zahlreiche Startups weltweit. Ebenfalls bemerkenswert sind Fortschritte bei der Software, insbesondere bei den Visual Language Action Models (VLAs), die es humanoiden Robotern ermöglichen, neue Aufgaben mit minimalem Training robust zu bewältigen und Anweisungen in natürlicher Sprache zu folgen. Dieses Whitepaper befasst sich mit aktuellen Entwicklungen zum Stand Mai 2025 und zeigt verschiedene Einsatzmöglichkeiten auf.



Bildquelle: Hersteller

Deutschland erlebt durch den demografischen Wandel einen stetig wachsenden Fachkräftemangel. Prognosen gehen davon aus, dass diese Lücke bis 2027 auf 728.000 Arbeitskräfte anwachsen könnte, was mit einem möglichen Produktionspotenzialverlust von bis zu 74 Milliarden Euro verbunden ist. Während Automatisierung und Zuwanderung in der Vergangenheit Abhilfe geschaffen haben, eröffnet humanoide Robotik nun eine neue Dimension der Automatisierung. Aufgaben, die bisher nicht kosteneffizient automatisiert werden konnten, könnten durch humanoide Roboter bearbeitet werden, womit wirtschaftliche Vorteile einhergehen. Die Vielseitigkeit humanoider Roboter spielt dabei eine entscheidende Rolle, da sie mit ihren menschenähnlichen Händen unterschiedliche Objekte greifen und normale Werkzeuge bedienen können, ohne dass spezifische Hardwareanpassungen und somit Umrüstzeiten und Kosten erforderlich sind. Diese Flexibilität erlaubt eine rasche Integration in bestehende Arbeitsprozesse, ohne große Umstellungen vornehmen zu müssen.

Die jüngsten Fortschritte in der Software haben maßgeblich dazu beigetragen, dass humanoide Roboter selbstständig lernen und sich an neue Aufgaben anpassen können. Während ältere Ansätze oft eine monatelange Programmierung erforderten, gibt es heute Methoden wie Imitation Learning und VLAs, die es Robotern ermöglichen, aus Demonstrationen eigenständig zu lernen. Indem sie mit Kameras ihre Umgebung wahrnehmen, können sie eigenständig Aufgaben anlernen sowie Fehler erkennen und korrigieren. Beispielsweise kann ein Roboter, der ein Bauteil falsch positioniert hat, durch eigene Analyse der Situation das Problem beheben, indem er das Objekt erneut aufnimmt oder an die korrekte Stelle verschiebt. Dieses selbstständige Lernverhalten wird durch Daten aus Demonstrationen erlangt, bei denen ein Mensch den Roboter teleoperiert und so relevante Trainingsdaten generiert.

Neben der industriellen Produktion bieten humanoide Roboter enorme Potenziale in verschiedenen Bereichen. Gerade in Bereichen, die für den Menschen gefährlich oder schwer zugänglich sind, könnten sie eingesetzt werden und physisch anstrengende Tätigkeiten übernehmen. Fortschrittliche Bewegungssteuerungen erlauben es humanoiden Robotern, sich auch in unstrukturierten und unwegsamen Umgebungen sicher zu bewegen.

Zusammenfassend ist festzustellen, dass humanoide Robotik durch die Fortschritte in künstlicher Intelligenz und maschinellem Lernen ein enormes Potenzial erreicht hat. Die Möglichkeit, aus Demonstrationen zu lernen und sich an neue Umgebungen anzupassen, macht sie zu einer vielversprechenden Lösung für die Automatisierung unterschiedlicher Prozesse. Unternehmen, die frühzeitig humanoide Roboter einsetzen, könnten Wettbewerbsvorteile erzielen und langfristig ihre Effizienz steigern.

1. Einleitung

Durch den demographischen Wandel haben in Deutschland immer mehr Unternehmen Probleme mit einem Mangel an Arbeitskräften. Das Institut für Wirtschaft geht bis 2027 von einem Anstieg der Fachkräftelücke auf 728.000 aus [1]. Bereits 2024 wurde von einem Produktionspotenzialverlust von 49 Milliarden Euro ausgegangen, was sich bis zum Jahre 2027 auf 74 Milliarden Euro pro Jahr steigern könnte [2]. Selbst bei einfachen Hilfstätigkeiten berichten laut dem WSI bereits ca. 20% der Unternehmen von Problemen die Stellen zu besetzen. Wobei bei mehr als der Hälfte der Grund war, dass keine geeigneten Bewerber gefunden wurden [3].

Zuwanderung und Automatisierung haben in der Vergangenheit geholfen den Fachkräftemangel abzuschwächen. Insbesondere durch die humanoide Robotik ergeben sich nun neue Möglichkeiten, Aufgaben zu automatisieren, die vorher gar nicht oder nicht kostendeckend automatisierbar waren. Dies bietet eine entscheidende Chance, wobei die frühe Adaptierung entscheidende Vorteile bringen kann.

Einfache Hilfstätigkeiten könnten in Zukunft fast vollständig automatisiert werden und Fachkräfte durch zusätzliche Unterstützung entlastet. Ein besonderer Vorteil der humanoiden Robotik stellt hierbei die Vielseitigkeit der Systeme dar. Während aktuell Robotersysteme in der Regel durch spezielle Erweiterungen an bestimmte Aufgaben hardwareseitig angepasst werden, können humanoide Roboter durch ihre menschenähnlichen Hände diverse unterschiedliche Objekte greifen und auch normale, für den Menschen gemachte, Werkzeuge benutzen. Eine aufgabenspezifische Hardwareänderung entfällt dementsprechend. Dies erlaubt, dass die Roboter schnell für unterschiedliche Aufgaben eingesetzt werden können.

Auch softwareseitig gab es in den vergangenen Jahren große Fortschritte in der Robotik. Hierbei ist der klassische Ansatz, dass Roboter zeitintensiv programmiert werden, um eine bestimmte Aufgabe auszuführen, wobei eine Änderung der Umgebung oder der Aufgabe zu erneutem hohem Entwicklungsaufwand führt. Seit einigen Jahren gibt es die Möglichkeit, Roboter durch das Vormachen von Bewegungsabläufen anzulernen, wobei diese dann starr definiert sind. Auch hier bedarf eine kleine Änderung der Umgebung oder der Aufgabe eines erneuten Anlernens. Die Systeme sind also nicht robust und führen immer nur den exakt definierten Bewegungsablauf durch. Neue Ansätze aus der Forschung umfassen das Imitation Learning und Visual Language Action Models (VLAs). Beide ermöglichen, dass die Roboter aus Demonstrationen selbständig lernen. Im Gegensatz zum klassischen Anlernen verbinden diese Systeme jedoch eine Kamera und benötigen mehrere Demonstrationen. Dies erlaubt jedoch, dass die Systeme robust gegen Störeinflüsse werden und Fehler selbständig beheben können. Sollte z. B. bei einer Positionierungsaufgabe ein Bauteil vom Roboter falsch positioniert worden sein, so versucht dieser anschließend dieses Problem selbständig zu lösen, z. B. indem er das Bauteil an die richtige Stelle schiebt oder erneut anhebt und positioniert.

Im folgenden soll zuerst in Abschnitt 2 das aktuelle Marktumfeld und Fortschritte in der humanoiden Robotik aufgezeigt werden, danach werden zuerst verschiedene Systeme und dann der aktuelle Stand der Forschung für die Bewegung in komplexen Umgebungen in den Abschnitten 3 und 4 aufgezeigt. In Abschnitt 5 wird diskutiert, wie sich Robotersysteme neue Aufgaben selbständig beibringen können, gefolgt von der Bewertung der Einsatzmöglichkeiten humanoider Roboter in verschiedenen Industriezweigen in Abschnitt 6. Zuletzt werden in Abschnitt 7 die Limitationen der aktuellen Systeme diskutiert und in Abschnitt 8 eine Zusammenfassung gegeben.

2. Aktuelles Marktumfeld in der humanoiden Robotik

Das aktuelle Marktumfeld in der humanoiden Robotik ist hochdynamisch, viele unterschiedliche etablierte Unternehmen, sowie Startups arbeiten aktiv an der Entwicklung humanoider Roboter. Der folgende Überblick über die dominierenden Firmen ist daher nur eine Momentaufnahme.

Das amerikanische Unternehmen Boston Dynamics ist eines der bekanntesten Unternehmen, wenn es um die Forschung an humanoider Robotik geht. Ihr Robotersystem Atlas wird unter anderem bei Hyundai vertestet, ist aber zum aktuellen Zeitpunkt der Studie noch nicht kaufbar. Andere Robotersysteme von Boston Dynamics, wie der Quadruped Spot, finden bereits heute weltweit diverse Anwendungen.

Agility Robotics ist ein weiteres US-Amerikanisches Unternehmen. Sie verkaufen ihr Robotersystem Digit schon seit einigen Jahren, dieses ist insbesondere für Logistikaufgaben wie das Transportieren von Boxen entwickelt worden und findet bereits jetzt in verschiedenen Unternehmen Anwendung.

Auch die Startups Figure und Apptronik haben ihren Sitz in den USA. Mit ihren Systemen Figure.02 und Apollo testen aktuell die Fahrzeughersteller BMW und Mercedes. Beide Systeme sind jedoch aktuell nicht käuflich erwerblich und befinden sich noch in der aktiven Entwicklung.

Ein weiterer Startup-Schwerpunkt befindet sich im asiatischen Raum, vor allem in China, wo es aktuell diverse Startups gibt, die an der Entwicklung humanoider Roboter arbeiten. Insbesondere sind hier Fourier Robotics und Unitree zu nennen. Unitree verkauft bereits seit einigen Jahren ihr System H1 und hat im Jahr 2024 den G1 auf den Markt gebracht. Diese Systeme sind vor allem als Hardware-Plattformen gedacht, damit Drittentwickler und Forscher eigene Software für diese schreiben können.

In Deutschland arbeitet vor allem Neura Robotics an einem humanoiden Roboter und auch Igus hat ein Konzept für einen humanoiden Roboter vorgestellt, diese sind zum Zeitpunkt der Veröffentlichung jedoch noch nicht offiziell vorgestellt worden.

In Abbildung 1 werden verschiedene kaufbare und quelloffene Systeme hinsichtlich ihres Kaufpreises und ihres maximalen Drehmoments verglichen.

*Korrespondierender Autor
Email address: dennis.bank@imes.uni-hannover.de (D. Bank)

Abkürzungen

ACT	Action Chunking with Transformers
FFG	Force Feedback Gloves
FG	Freiheitsgrad
IL	Imitation Learning
LLM	Large Language Model
RaaS	Robots as a Service
RL	Reinforcement Learning
VLA	Visual Language Action Model
VR	Virtual Reality

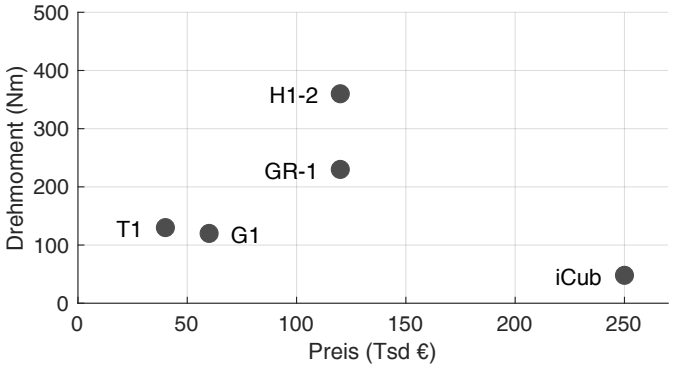


Abbildung 1: Vergleich verschiedener humanoider Robotersysteme, die zum Zeitpunkt der Studie erwerbbar oder Open-Source sind, anhand des Drehmoments des stärksten Motors und des Preises.

Tabelle 1: Übersicht über verschiedene, aktuell kaufbare Robotersysteme mit ihrem Preis, Anzahl an Freiheitsgraden und dem maximalen Drehmoment des stärksten Motors.

Name (Firma)	Preis in T€	FG	Größe in m	Max M in Nm:
iCub (Quelloffen)	250	53	1,05	48
Digit (Agility Rob.)	180	28	1,75	–
GR-1 (Fourier Rob.)	120	44	1,68	230
T1 (Booster Rob.)	40	23	1,20	130
G1 (Unitree Rob.)	60	43	1,30	120
H1-2 (Unitree Rob.)	120	31	1,80	360

3. Systeme zur Fortbewegung in komplexen Umgebungen

Humanoide Roboter sind dazu vorgesehen, in von Menschen genutzten Umgebungen zu agieren, von Wohnungen zu Fabrikhallen, aber auch in Katastrophengebieten [4]. Sie können neben alltäglichen Aufgaben außerdem besondere Einsätze in Bereichen übernehmen, die für Menschen gefährlich oder schwer zugänglich sind. Dazu gehören unter anderem [5, 6, 7]:

- **Alltagshilfe**: Unterstützung eingeschränkter Menschen im Haushalt, Bedienung von Geräten oder das Holen und Bringen von Gegenständen.
- **Industrie**: Einsatz als flexible Arbeiter in Fertigungsanlagen oder Lagerhallen, wo sie Werkzeuge und Maschinen ähnlich wie ein Mensch bedienen können.
- **Katastrophenschutz und Rettung**: Vordringen in eingestürzte Gebäude, Kriegsgebiete oder andere gefährliche Bereiche, um z. B. nach Verschütteten zu suchen oder Arbeiten an für Menschen gefährlichen Orten durchzuführen
- **Extrembedingungen**: Erkundung von Umgebungen mit extremen Temperaturen oder auf anderen Planeten, wo herkömmliche Fahrzeuge an ihre Grenzen stoßen

Um diese Aufgaben bewältigen zu können, ist eine zuverlässige, sichere Fortbewegung in teils komplexen und unübersichtlichen Umgebungen die grundlegende Voraussetzung. Zu den Herausforderungen zählen unter anderem unsichere Fußtritte, Vegetation, erschwerte Sicht- und Beleuchtungsverhältnisse im unbekannten Außengelände, aber auch physische Hindernisse wie Treppen oder Engpässe sowie schwer wahrnehmbare transparente und spiegelnde Flächen in Gebäuden. Zur Lösung der beschriebenen Herausforderungen ist ein hochflexibles Fortbewegungskonzept notwendig.

Quadrupeds / Vierbeinige Laufroboter verbinden Stabilität mit Geländegängigkeit. Mit vier Beinen können sie, anders als Zweibeiner, auch statisch stabil stehen und besitzen eine geringere Sensitivität des Körperschwerpunkts. Sie können gezielt Beine anheben, um Hindernisse und unebenes Gelände abhängig von der Ausprägung zu überwinden. Quadrupeds wie z. B. *ANYmal* (Firma ANYbotics) oder *Spot* (Firma Boston Dynamics) können Treppen steigen, über Geröll laufen und sich, sollten sie durch externe Einflüsse aus dem Gleichgewicht gebracht werden, oft selbst abfangen. Nachteilig ist die Bodennähe des Rumpfs bei kleinen Systemen, welche Manipulationsaufgaben erschwert, und die große Grundfläche bei größerer Bodenhöhe.

Wheeled Quadruped (Hybrid): Sind eine Mischform, die Beine mit angetriebenen Rädern kombinieren. Hierbei sitzen an den Enden der Beine Räder, sodass der Roboter auf ebenen Flächen fahren bzw. rollen kann, aber bei Bedarf die Füße hebt, um z. B. eine Stufe zu überwinden. Solche Konzepte versprechen das Beste aus beiden Welten: Effizienz auf glattem Boden und Beweglichkeit auf schwierigem Terrain. Die Herausforderung ist hier allerdings der große Konstruktions- und Steuerungs-Aufwand. Diese Hybriden sind mechanisch aufwändig und das Zusammenspiel zwischen Roll- und Schreitschritten ist komplex.

Abbildung 2: Spot (Firma Boston Dynamics) als Beispiel für einen quadrupeden Roboter. Quelle: Boston Dynamics

Abbildung 3: Der B2-W (Firma Unitree) hat angetriebene Räder an seinen Füßen und kann so dynamisch zwischen Gehen und Fahren wechseln. Quelle: Unitree

Die Kombination einer stabilen Plattform (radgebunden, vierbeinig oder einer Mischform daraus) mit einem humanoiden Oberkörper bietet Zugang zu Aufgaben, die für klassische Vierbeiner nicht erreichbar sind. Beispiele dafür sind fahrbare Dual-Arm-Cobot-Systeme wie MiPA (Firma Neura Robotics) oder die Forschungsplattform Rollin' Justin (DLR)

Abbildung 4: Rollin' Justin (DLR) kombiniert einen menschenähnlichen Oberkörper mit einer fahrenden Plattform und muss sich daher nicht aktiv balancieren. Quelle: DLR

Bipedal / Humanoid (Zweibeiner): Bipedale humanoide Roboter zeichnen sich durch einen menschenähnlichen zweibeinigen Aufbau aus. Dieser Aufbau ist vorteilhaft, weil er prinzipiell den Zugang zu all den Orten ermöglicht, die auch Menschen erreichen können, über Treppen, Leitern und unebenes Gelände. Gleichzeitig können humanoide Roboter Werkzeuge und Bedienelemente nutzen, die für Menschen entwickelt wurden (z. B. Türgriffe oder Fahrzeuge), was radgetriebene Roboter ohne humanoide Gestalt nicht in jedem Fall können. Daraus folgen vielfältige Einsatzgebiete und Aufgabenpotenziale von humanoiden Robotern. Auch soziale Aspekte spielen eine Rolle, denn beim Einsatz in der Pflege fördert ein menschenähnliches Erscheinungsbild die Akzeptanz durch die Patienten [8].

Mit den Vorteilen geht allerdings eine erhöhte Komplexität humanoider Roboter im Vergleich zu anderen Robotertypen einher. Der inhärent instabile Körperschwerpunkt auf zwei Beinen führt zu einer hohen Sensitivität bei der Verlagerung des Körperschwerpunkts und muss kontinuierlich ausgeregelt werden, um Stürze zu vermeiden. Zudem führt die hohe Anzahl an aktuierten Gelenken bzw. Freiheitsgraden zu dementsprechend erhöhter Komplexität in der Bewegungsplanung und -regelung [9]. Dadurch sind humanoide Roboter bisher anfälliger für Stürze und Fehlfunktionen in komplexen Umgebungen als beispielsweise Quadrupeds.

4. Locomotion

Der Begriff *Locomotion* bezeichnet die aktive Fortbewegung mit dem Ziel der Ortsveränderung. Ansätze zur Fortbewegung in der humanoiden Robotik sind weiterhin Gegenstand aktueller Forschung. Bisher fand die Entwicklung von Laufrobotern überwiegend in kontrollierten Laborumgebungen statt [4]. Berühmte humanoide Roboter wie *ASIMO* oder *Atlas* demonstrieren beeindruckende Läufe und Sprünge, jedoch meist auf vorbereiteten, flachen Böden und klar definierten Parcours. In unbekannten, komplexen Umgebungen stießen Systeme schnell an ihre Grenzen. Viele heute verfügbare zweibeinige Roboter können nur unter idealen Bedingungen zuverlässig gehen und laufen.

Die bisherigen Ansätze basieren häufig auf klassischer Regelungstechnik: Es werden meist programmierte explizite Gleichgewichtsregler und Fußschrittplaner mithilfe von vereinfachten Annahmen über den Untergrund verwendet. Solche Controller funktionieren gut für vorhergesehene Situationen, versagen aber oft, sobald unvorhergesehene Störungen auftreten, wie veränderte Reibungskräfte und Höhenprofile durch bspw. einen Stein unter dem Fuß oder einen unerwarteten Zusammenstoß mit einem anderen Objekt. Entsprechend war der bisherige Anwendungsbereich vorrangig auf einfache, statische Umgebungen limitiert, was eine Lücke bei der Beherrschung wirklich komplexer, dynamischer Terrains hinterlässt [4].

Im Folgenden werden in Abschnitt 4.1 die grundsätzlichen Ideen zum Reinforcement Learning im Bereich der Locomotion vorgestellt. Danach folgt in Abschnitt 4.2 die Erklärung von Policies, die keine visuellen Informationen nutzen und in Abschnitt 4.3 werden Policies, die auch visuelle Informationen nutzen, vorgestellt. Abschließend wird ein Zwischenfazit gezogen.

4.1. Reinforcement Learning

Aktuelle Ansätze sehen keine klassische Regelung mehr vor, sondern nutzen *Reinforcement Learning* (Verstärkungslernen), eine Methode des maschinellen Lernens. Im *Reinforcement Learning* lernt ein Agent durch das Interagieren mit seiner Umgebung nach dem *Trial-and-Error-Prinzip* ein optimales Steuerungsverfahren, die *Policy* für sequentielle Entscheidungsprozesse. Der wichtige Unterschied ist, dass ein regelbasierter Controller außerhalb seines Designfalls üblicherweise

versagt, während eine gut trainierte *Reinforcement-Learning-Policy* auch in solchen Situationen adäquat reagieren kann. [10]

Das Problem der Locomotion wird im Regelfall als *POMDP* (*Partially Observable Markov Decision Process*) – partiell beobachtbarer Markov-Entscheidungsprozess – beschrieben. Dabei beschreibt der Zustandsraum den vollständigen Zustand von Roboter und Umgebung, der Aktionsraum mögliche Gelenkansteuerungen und der Beobachtungsraum die beschränkten, oft verrauschten Sensordaten des Roboters. Eine *Policy* stellt einen Zusammenhang zwischen den Beobachtungen und den Aktionen her, sodass die erwartete kumulative Belohnung maximiert wird. Die Belohnung wird anhand einer Belohnungsfunktion berechnet, die in der Planung ausgelegt werden muss. Die zentrale Idee ist somit, den Roboter durch *Trial-and-Error* selbstständig lernen zu lassen, welche Bewegungen vorteilhaft sind. [11, 9, 12]

Die Belohnungsfunktion setzt sich meist aus mehreren Termen zusammen, die verschiedenen Gruppen abhängig von ihrer Zielsetzung zugeordnet werden können. In [9] wird eine Unterteilung in folgende Gruppen vorgenommen, es sind beispielhaft einige übliche Belohnungsterme angeführt vgl. [11, 9, 13, 14, 15, 16, 17, 4, 18, 19, 20, 21]:

- **Aufgabenerfüllung**: Terme zum sicheren Fortbewegen in einer vorgegebenen Geschwindigkeit bzw. Erreichen eines Zielpunkts ohne Sturz
  - Belohnung proportional zum Übereinstimmungsgrad der Orientierung
  - Belohnung proportional zum Übereinstimmungsgrad der Zielgeschwindigkeit des Roboters
  - Grundlegende Belohnung pro Zeitintervall ohne Terminierung für sichere Fortbewegung
  - Hohe Bestrafung bei Terminierung durch Sturz o.ä.
- **Regularisierung**: Terme zur Gewährleistung ruckfreier Bewegungsabläufe und Einhaltung der Hardware Grenzen (bspw. Motordrehmomente, -positionen)
  - Einschränkung von Motordrehmomenten
  - Einschränkung von Gelenkpositionen
  - Glätten und Skalieren von extrem ruckartigen Motorbefehlen
- **Bewegungsstil**: Terme für natürliche humane Fortbewegung und leichte Bodenkontakte
  - gleichmäßige Schrittdauern, -höhen, -weiten
  - niedrige Bodenkontaktkräfte

Echte Sturzerfahrungen sind in der Praxis teuer und riskant. Deshalb werden die Policies zunächst in paralleler Simulation auf *GPU-Clustern* trainiert, die populärsten Anwendungen hierfür sind aktuell *NVIDIA Isaac Gym* und *MuJoCo*. Hunderte oder Tausende virtuelle Roboter lernen, wie in Abbildung 5 dargestellt, gleichzeitig in einer virtuell konzipierten Umgebung mit hochgenauer physikalischer Modellierung, was Trainingszeiten von Wochen auf Minuten oder Stunden verkürzt. In [11] wurde bspw. ein *Quadruped* in weniger als 20 Minuten Rechenzeit (aber mehreren Jahren simulierter Zeit) zu einer lauffähigen *Policy* trainiert.

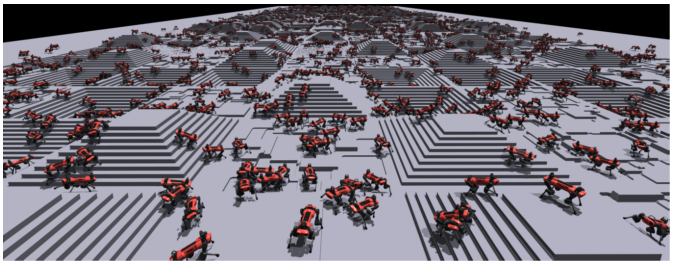


Abbildung 5: Tausende Roboter lernen parallel das Laufen in einer Simulationsumgebung in *NVIDIA Isaac Lab*. Quelle: [11].

In der Simulation werden gezielt zufällige Varianzen in physikalischen Parametern (*Domain Randomization*) erzeugt, um die Robustheit der entwickelten *Policies* zu erhöhen. Dadurch wird eine Anwendung in der echten Welt ermöglicht und die Abweichungen zwischen Simulationsdaten und Realdaten (*Sim-to-Real-Gap*) in bspw. Sensoren, Aktuatoren und physikalischer Modellierung verringert. Zudem ist in der Simulation eine systematische Untersuchung vieler Einflussfaktoren samt privilegierter Informationen möglich, die in der realen Welt nur schwer oder gar nicht eingesehen werden können. Durch den Zugang zu *Ground-Truth-Informationen* kann das *POMDP* Problem zum Teil in ein effizienter zu berechnendes *Supervised Learning* Problem überführt werden. Eine weitere Methode zur Verbesserung der erzielten Ergebnisse in der Simulation ist das *Curriculum Learning*, bei dem die Aufgabenschwierigkeit sowie die Varianz der physikalischen Parameter von einem niedrigen Niveau an schrittweise mit der Performance der *Policy* erhöht wird. [11, 9, 13, 14]

Nach der Entwicklung einer robusten *Policy* erfolgt der *Sim-to-Real-Transfer*, also die Übertragung der im Simulator entwickelten *Policy* auf den physischen Roboter. Faktisch geschieht dies durch die Implementierung der *Policy* auf dem *On-Board-Computer* des Roboters, sodass nun aus den aufgenommenen Sensordaten der Hardware des Roboters die entsprechenden Aktionen für die reale Welt generiert und mittels PD-Regler und Aktoren umgesetzt werden.

Ältere Arbeiten an humanoiden Robotern nutzen vorrangig rein propriozeptive Sensordaten, also die Wahrnehmung des eigenen Zustands als Eingangsdaten der *Policy* in der Beobachtung. Diese werden als *Blind Policies* bezeichnet und sind Verfahren ohne Beobachtung der Umgebung. Der aktuelle Trend geht zur Verwendung von propriozeptiven und exterozeptiven Sensordaten, also die Wahrnehmung des eigenen Zustands und die Wahrnehmung der Umgebung als Eingangsdaten der *Policy* in der Beobachtung zu nutzen. Diese werden als *Vision Policies* bezeichnet. Im Folgenden werden beide *Policy* Typen kurz erläutert.

4.2. Blind Policies

Blind Policies nutzen zur Erfassung von Sensordaten die *Encoder* der einzelnen Gelenke, um Informationen über die Gelenkwinkelstellungen zu erhalten. Zudem wird eine *IMU* (*Inertial-Measurement-Unit*) verwendet, um Beschleunigungen und Winkelgeschwindigkeiten zu erfassen, aus denen Lage und Ausrichtung des Roboters im dreidimensionalen Raum bestimmt werden. Beispielsweise wurde ein *World-Model-Reconstruction-Ansatz* vorgestellt, bei dem eine blinde *Policy* durch einen parallel trainierten Zustandsschätzer unterstützt wird, der den Zustand der Umgebung rekonstruiert [4]. Die beiden neuronalen Netze des Controllers und des Weltmodells werden dabei gemeinsam trainiert, aber der Informationsfluss dazwischen wird so begrenzt, dass der Controller letztlich blind arbeitet – er verlässt sich nur auf die vom Weltmodell geschätzten Zustände. Auf diese Weise lernte ein aktueller humanoider Roboter ohne Sensoren zur Beobachtung der Umgebung selbst auf vereistem und verschneitem Untergrund, siehe Abbildung 6, robust zu gehen und absolvierte erfolgreich eine 3,2 km lange Wanderung ohne jegliche menschliche Hilfe [4]. In [22] wurde ein ähnlicher Ansatz umgesetzt, während in [23] zusätzlich *Transformer-* und *LSTM-Netzwerke* genutzt wurden, um die Performance durch besseres Verständnis der zeitlichen Zusammenhänge zu steigern. In [24] wird eine verbesserte Balance der Ganzkörperbewegungen durch zusätzliche Schwerpunktbetrachtungen realisiert.

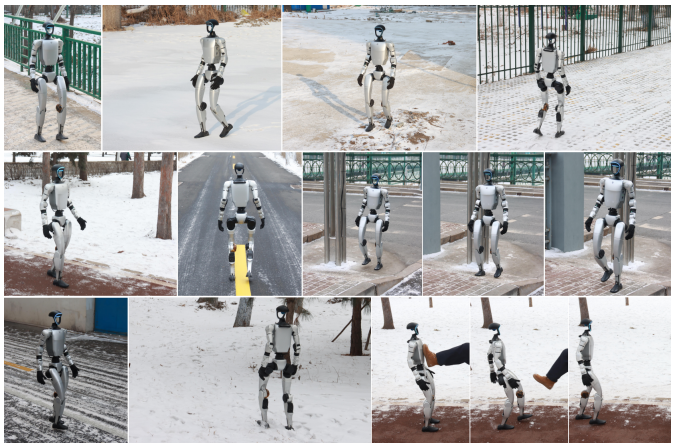


Abbildung 6: Einsatz im Außengelände: Tests im Außenbereich in von Eis und Schnee bedeckten Umgebungen. Dieses Modell ist in der Lage, erfolgreich verschiedene Geländetypen wie unebenes Terrain, Schotterflächen und verschneite Passagen zu durchqueren. Quelle: [4].

Diese *Blind Policies* gelten zwar als sehr robust gegenüber Sensorrauschen, müssen aber quasi tastend vorgehen und stoßen daher oft an Geschwindigkeitslimits [9]. In komplexen Umgebungen erreichen *Blind Policies* nur noch geringe Erfolgsraten und weisen eine stark erhöhte Sturzquote auf [9]. Diese komplexen Umgebungen erfordern daher zwingend eine Wahrnehmung der Umgebung, um auf Höhenunterschiede, wechselndes Terrain, bewegte Objekte oder andere äußere, nicht kontrollierbare Einflüsse wie Reibparameter reagieren zu können. Bei humanoiden Robotern ist perzeptives Feedback aus der Umgebung bedingt durch die Instabilität des Körperbaus im Gegensatz zu *Quadrupeden* für sichere *Locomotion* besonders relevant [13].

4.3. Vision Policies

Vision Policies nutzen zur Erfassung von Sensordaten zusätzlich zu *Blind Policies* noch Beobachtungen aus der Umgebung, meist durch Kamera-, Tiefenkamera- oder LiDAR-Daten zur Einschätzung der Entfernungen verschiedener Objekte. Diese erlauben selbstständiges Erkennen, Planen und Handeln in diversen Umgebungen und ermöglichen vorausschauende *Locomotion*. Gleichzeitig existiert bei *Vision Policies* die Herausforderung, dass das Sehvermögen durch z. B. schlechte Lichtverhältnisse, spiegelnde, transparente sowie wenig strukturierte Oberflächen oder Sensorfehler eingeschränkt werden kann. Infolge von fehlerhafter Wahrnehmung sind Fehlentscheidungen in der Bewegungsplanung möglich, die gefährliche Stürze bedeuten können. [9]

Forscher der Oregon State University entwickelten eine Strategie, bei der ein blinder RL-Laufcontroller durch ein visuelles Modul ergänzt wurde. Die Kamera des Roboters erfasst die vor ihm liegende Terrainstruktur. Diese wird mittels eines neuronalen Netzwerks aus Stereokamera-Daten in eine Höhenkarte umgewandelt und dient dem Roboter als Orientierung für den kommenden Schritt. Dank dieses Zusatzmoduls konnte der Roboter unebenes Terrain bewältigen. [17]

Im **VB-Com** Framework werden eine visuelle und eine blinde Policy parallel trainiert. Ein intelligenter Manager entscheidet dann in Echtzeit, welche der beiden Policies im aktuellen Roboterzustand bessere Ergebnisse verspricht. Bei guten Sichtbedingungen und zuverlässigen Daten wird die Vision Policy genutzt, bei Unsicherheit der Daten oder Sensor(teil)ausfällen übernimmt die Blind Policy. [9]

Im **BeamDojo** Projekt wurde besonders agile Fortbewegung selbst auf sporadischen Trittflächen, wie in Abbildung 7 demonstriert, die durch präzise Fußplatzierungen und Gleichgewichtskontrolle erlernt wurde. [14]

In BeamDojo wird einer der gebräuchlichsten Ansätze zur Verarbeitung der perzeptiven Sensordaten genutzt, eine roboterzentrierte Höhenkarte der direkten Umgebung. Der Ansatz basiert auf Arbeiten von [25, 26, 27] und liefert als finale Information ein 2D-Gitter mit fester Auflösung für den Roboter, in dem Höhenangaben des umliegenden Terrains enthalten sind. Diese Zwischenrepräsentationsform dient dem Roboter als einfach zu verarbeitende Informationsquelle zur Planung der nächsten Bewegungen und ist in Abbildung 8 dargestellt. In weiteren zentralen Arbeiten im Bereich Humanoid *Locomotion* und Humanoid Parkour werden vergleichbare Ansätze genutzt [13, 16, 18, 20, 21]. In [16] wurde durch die Einbeziehung der Beobachtung der Umgebung beispielsweise eine Reduzierung der Anzahl an Kollisionen um über 90% im Vergleich zu einer Blind Policy erreicht, was die Bedeutung der Umgebungswahrnehmung für sichere humanoide *Locomotion* unterstreicht.

Die rohe LiDAR-Punktwolke wird zunächst vorgefiltert, sodass alle Messpunkte, die außerhalb eines definierten Bereichs um den Roboter liegen oder keine Bodenpunkte sind, verworfen werden. Anschließend werden die Orientierungssignale der

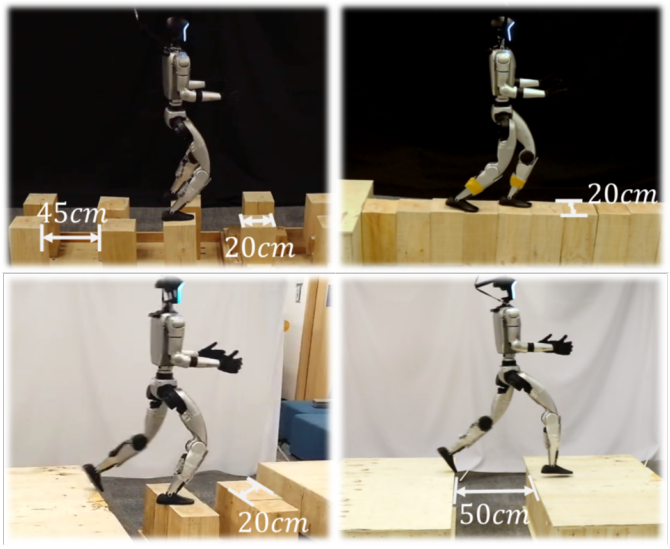


Abbildung 7: Experimente in der realen Welt an Hindernissen, die zuvor in der Simulation trainiert wurden. Die wenigen, engen Trittflächen erfordern eine hohe Präzision, Balance und Beweglichkeit des Roboters. Quelle: [14].

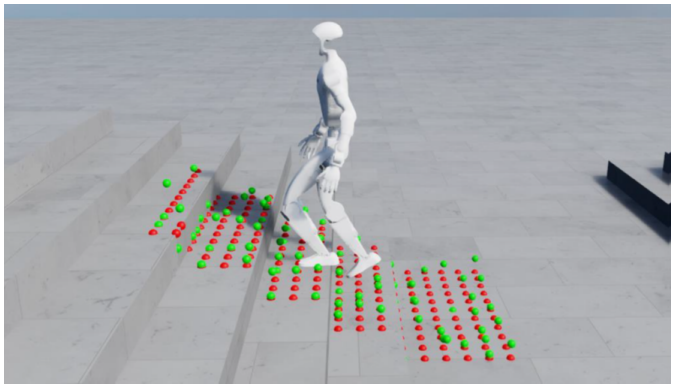


Abbildung 8: Darstellung der idealen Höheninformationen in rot und verrauschten Höheninformationen (i.d.R. *Policy-Input*) in grün. Quelle: [18].

IMU über z. B. den *FAST-LIO* Algorithmus [28, 29, 30] integriert, um die Punktwolke gegen Lageänderungen des Roboters zu stabilisieren. Alternativ wird aus Tiefenbildern einer Tiefenkamera durch ein neuronales Netz eine Höhenkarte geschätzt und verfeinert, welche anschließend im Training genutzt wird [17, 19, 21, 31]. In der Simulation kann effizientes *Supervised Learning* mit den *Ground Truth* Daten der Höhenkarte realisiert werden, um die Ergebnisse zu verbessern [13, 14]. In [32] wird ein *Voxel-Gitter*-basierter Ansatz erfolgreich genutzt, um aus verrauschten, unvollständigen Punktwolkendaten die Umgebung sehr detailliert zu rekonstruieren. Die Rekonstruktion aus originalen Messdaten wird beispielhaft in Abbildung 9 gezeigt.

4.4. Zwischenfazit Locomotion

Zusammenfassend zeichnet sich ab, dass RL-basierte Locomotion zunehmend von der reinen Propriozeption (Gelenk- und Inertialsensorik) hin zur perzeptiven (enthält Kamera und LiDAR) Fortbewegung geht. Anfangs war es ein Erfolg, dass ein

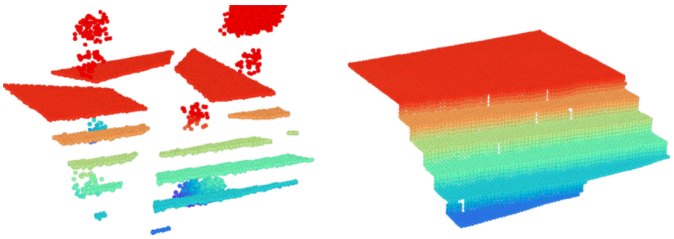


Abbildung 9: *Voxel*-Gitter-basierter Ansatz eingesetzt an einer Treppe. Links sind die originalen Messdaten dargestellt, rechts die Rekonstruktion der Umgebung. Quelle: [32].

Roboter überhaupt stabil laufen konnte, dies erreichte man am schnellsten ohne die Herausforderungen der visuellen Daten. Mittlerweile aber sind die Grundlagen vorhanden und Forscher arbeiten daran, den Robotern auch das Sehen und Vorausschauen beizubringen, ohne die Robustheit zu verlieren. Die wichtigsten aktuellen Arbeiten kombinieren deshalb Propriozeption und Vision, oft mittels weiterentwickelter neuronaler Netzwerke, die z. B. Höhenkarten schätzen, Zustände rekonstruieren oder zwischen verschiedenen *Policies* umschalten. Dadurch werden humanoide (und allgemein beinbetriebene) Roboter immer besser darin, komplexe Umgebungen zu meistern.

5. Anlernen von neuen Aufgaben

Aktuell ist das Programmieren von Robotern oft komplex oder wenig robust. Es muss entweder eine individuelle Lösung für eine spezifische Aufgabe entworfen werden, was mit viel Implementierungsaufwand verbunden ist oder ein genauer Bewegungsablauf wird z. B. über eine Handführung einprogrammiert, dies ist jedoch nicht robust gegenüber Veränderungen in der Umgebung. Ein aktives Forschungsfeld ist daher, dass Roboter sich anhand von Demonstrationen selbständig beibringen können Aufgaben robust auszuführen. Die Demonstration findet hierbei hauptsächlich über die Teleoperation statt, die im folgenden Subkapitel beschrieben wird. Danach folgen die Vorstellung von aktuellen Methoden zum Imitation Learning und Ansätze im Bereich der Visual Language Action Models, abschließend wird ein Zwischenfazit gezogen.

5.1. Teleoperation

Teleoperation bezeichnet das Steuern des Roboters durch das Kopieren der Bewegungen eines Menschen. Dies kann je nach Verfahren ortsgebunden oder ortsunabhängig, also theoretisch auch vom anderen Ende der Welt aus über das Internet, stattfinden. Die wichtigsten Schnittstellen, wie Bediener Befehle erteilen und Feedback bekommen können sind:

- **Traditionelle Eingabegeräte** wie Joysticks, Gamepads. Diese ermöglichen keine intuitive Bedienung von komplexen Robotern [33].
- **Manuelle Steuergeräte** oder spezialisierte Exoskelette ermöglichen eine präzise Steuerung des Roboters. Sie können bei komplexen Aufgaben ermüdend sein [33].

- **Haptische Schnittstellen:** Exoskelett-Handschuhe oder Force-Feedback-Joysticks übertragen taktile und Kraftinformationen an den Bediener, was die Geschicklichkeit verbessert. Kann anfällig für Kommunikationsverzögerungen sein [34].
- **Virtuelle Realität (VR):** VR-Brillen ermöglichen eine immersive Darstellung der Roboterumgebung und natürliche Tiefeinwahrnehmung bei Stereo-Video [35].
- **Kameras:** Markerlose Bewegungserfassung oder Tiefenkameras verfolgen Körper- und Handbewegungen und übertragen diese direkt auf Roboter [33].

Gegenüber der Verwendung von einfachen Fernbedienungen ermöglicht Teleoperation eine intuitive und feinfühlige Steuerung der Robotergelenke durch eigene Bewegungen. Dies ist bei humanoiden Robotern aufgrund der großen Anzahl an Freiheitsgraden (FG) von besonderem Vorteil. So besitzt beispielsweise der Unitree G1 in der Version mit robotischen Händen 43 FG [36]. Aus der Filmindustrie sind Motion-Capture-Anzüge und Trackingsysteme mit Infrarotkameras bekannt, um die Bewegungen eines Menschen zu erfassen. Aufgrund der großen Fortschritte im Bereich des Maschinellen Lernens, ist es nun auch möglich die Pose eines Menschen mit nur wenigen einfachen Kameras zu erfassen, was die Kosten deutlich senkt [33]. Die Fortschritte in der VR-Technik bieten zudem nun die Möglichkeit, für nur wenige hundert Euro VR-Brillen zu nutzen, um virtuell aus der Sicht des Roboters ihn zu teleoperieren. Wird eine bewegliche Kamera in dem Kopf des Roboters montiert, ist es möglich, dass sich der Mensch mit natürlichen Kopfbewegungen umgucken kann. Teleoperation mit Hilfe einer VR-Brille ist dargestellt in Abbildung 10. Gleichzeitig bietet, die VR-Brille den Vorteil, dass sie kamerabasiert die Position der Hände des Menschen sehr genau erfassen können, was sich gut zum Teleoperieren von Roboter eignet [35].

Da jeder Mensch etwas andere, zu dem humanoiden Roboter

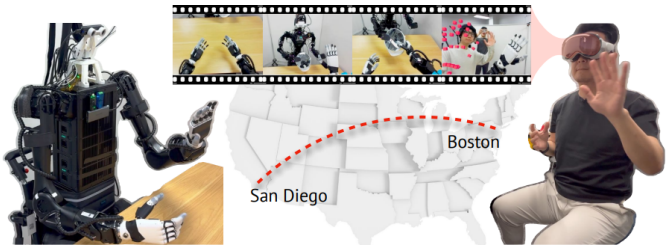


Abbildung 10: Teleoperation über große Distanzen mittels des Internets. Quelle: [35].

verschiedene Proportionen besitzt, muss die erfasste Pose in Roboterkoordinaten umgerechnet werden. Dies geschieht mit einer Kombination aus mathematischen Methoden (Geometrie), wie z. B. der inversen Kinematik und neuronalen Netzen. Aufgrund des höheren Rechenaufwandes wird dabei auf Methoden des maschinellen Lernens nur zurückgegriffen, wenn das Übertragen von der menschlichen Pose auf die Gelenkwinkel des Roboters besonders anspruchsvoll ist, wie z. B. bei der Übertragung von Handposen auf die Gelenkwinkel von robotischen Händen. [37].

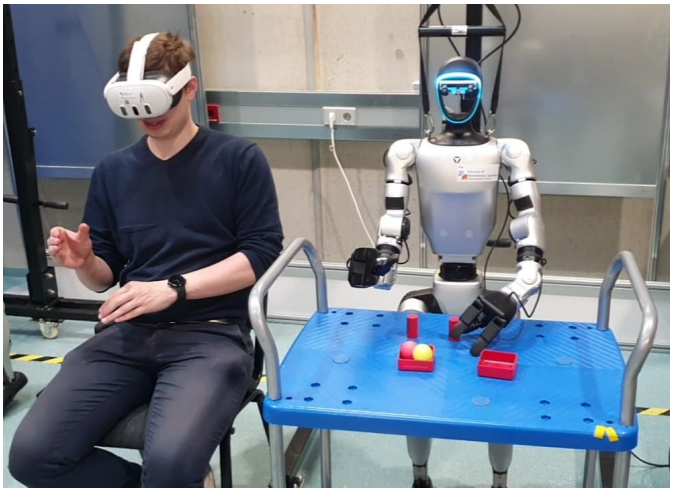


Abbildung 11: Teleoperation eines Unitree G1 mit einer Meta Quest 3 an der Leibniz Universität Hannover (IMES). Implementiert auf Basis von [35].

In Abbildung 11 ist die mit einer VR-Brille kamerabasierte Teleoperation eines humanoiden Roboters (Modell G1 von Unitree) dargestellt. Der Aufbau basiert auf OpenTeleVision [35].

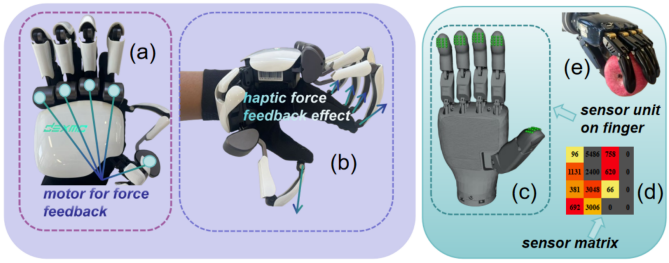


Abbildung 12: Funktion der haptischen Darstellung von Objekten mit FFGs.(a) Jeder Finger des Exoskelett-Handschuhs (FFG) ist mit einem Motor ausgestattet, der haptisches Feedback liefert. (b) Der menschliche Bediener spürt das haptische Kraftfeedback über den FFG. (c) Ein taktiler Sensor mit 16 Messeinheiten ist in die robotische Hand integriert. (d) Beispiel einer taktilen Sensormatrix, wenn ein Sensor ausgelöst wird. (e) Beispiel eines Greifvorgangs. Quelle: [34].

Der Nachteil der kamerabasierten Erfassung ist, dass der menschliche Bediener kein haptisches Feedback erhält. Aus der VR-Technik gibt es spezielle Handschuhe, die es, gepaart mit Drucksensoren an den Händen des Roboters, ermöglichen einen Widerstand zu fühlen, wenn der Roboter etwas greift. Solche Handschuhe werden als Force-Feedback-Gloves bezeichnet. Die Funktionsweise der Kraftdarstellung vom Roboter auf den Menschen ist in Abbildung 12 dargestellt. FFGs ermöglichen die Greifbewegungen vom Menschen auf den Roboter sehr genau zu übertragen. Der Preis pro FFG beträgt aktuell noch mehrere tausend Euro, so dass ein wirtschaftlicher Einsatz noch nicht für jeden Anwendungsfall möglich ist [34]. Da die Systeme sehr neu sind, ist zu erwarten, dass die Preise mit zunehmender Technologieverbreitung und dem Erreichen des Massenmarktes sinken werden.

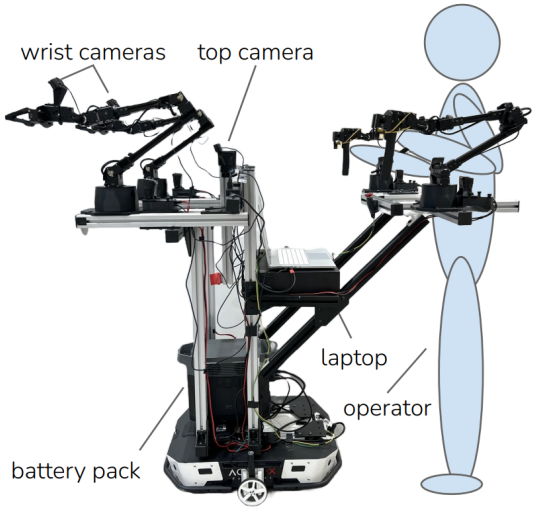


Abbildung 13: Übersicht des Mobile ALOHA teleoperation Aufbaus Quelle: [38]

Neben der kamerabasierten Posenerfassung gibt es auch die Möglichkeit, ein dem Roboter ähnliches Gestänge mit Positionssensoren in den Gelenken zu bauen oder Exoskelette zu verwenden, was eine sehr zuverlässige und präzise Steuerung der Roboterarme ermöglicht. Der Nachteil solcher Systeme ist, dass die Bedienung anstrengend und ermüdend sein kann. Ein solches System wird als manuelles Steuergerät bezeichnet und findet auch in Chirurgierobotern wie dem Da-Vinci-Roboter (Firma Intuitive Surgical) Verwendung. In Abbildung 13 ist ein Aufbau der Stanford University dargestellt. Die Arbeit [38] kommt zu dem Ergebnis, dass mit dem Einsatz von manuellen Steuergeräten sehr präzise Aufgaben durchgeführt werden können, wobei jedoch bislang nur einfache Greifer und keine robotischen Hände verwendet werden können. Manuelle Steuergeräte bieten gegenüber kamerabasierten Systemen weniger Flexibilität und benötigen mehr Platz.

Der Einsatz von Teleoperation bietet die Chance, dass gefährliche oder besonders anstrengende Tätigkeiten von ferngesteuerten Robotern übernommen werden, um die Gesundheit von Menschen zu schützen [39]. Ein Beispiel dafür stellen die ferngesteuerten Feuerwehr-Roboterhunde der chinesischen Firma Unitree da, welche für das Finden von Personen in brennenden Häusern eingesetzt werden [40]. Außerdem bietet Teleoperation die Grundlage, um Daten für das Imitation Learning zu sammeln, also dem Roboter basierend auf der Demonstration von menschlichen Experten neue Fähigkeiten beizubringen. Sollte der Roboter in bestimmten Situationen an die Grenzen seiner autonomen Fähigkeiten kommen, bietet Teleoperation die Möglichkeit, dass ein Mensch die Steuerung des Roboters übernimmt und die Situation löst. Danach kann der Mensch die Steuerung des Roboters wieder an die autonome Software übergeben. Dies ermöglicht es einem Menschen viele Roboter gleichzeitig zu überwachen.

5.2. Imitation Learning

Imitation Learning ermöglicht dem Roboter das Erlernen von autonomen Fähigkeiten, basierend auf beobachteten Demonstrationen von Experten. Verglichen mit VLAs, können mit IL neue Fähigkeiten mit deutlich weniger Rechenaufwand erlernt werden. Gleichzeitig ist Imitation Learning dafür aber auch weniger fähig als VLAs (siehe Abschnitt 5.3). In der humanoiden Robotik bietet sich die Verwendung von Teleoperation zur Datenerfassung besonders gut an, da so auf echten (nicht simulierten) Daten gelernt werden kann. Die grundlegenden Schritte des Imitation Learnings sind bei den Verfahren ähnlich und bestehen aus dem Beobachten der Demonstration eines Experten (Expert Demonstration), dem Aufzeichnen von Zustands-/Aktions-Paaren (State/Action Pairs) und dem anschließenden Lernen basierend auf den gesammelten Daten (Learning). Dieser Ablauf ist in Abbildung 14 dargestellt [41].

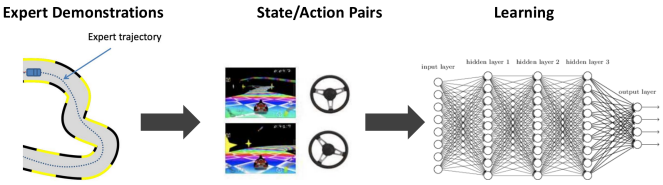


Abbildung 14: Die grundlegenden Schritte des Imitation Learnings; Quelle: [41]

Imitation Learning lässt sich in folgende Kategorien aufteilen:

- **Behavior Cloning (BC)**: Direkte Abbildung von beobachteten Zuständen auf Aktionen mittels überwachten Lernens.
- **Direct Policy Learning**: Direktes Lernen der Policy durch interaktive Demonstrationen.
- **Inverse Reinforcement Learning (IRL)**: Ermittlung der zugrunde liegenden Belohnungsfunktionen, die das Expertenverhalten steuern.

BC: Behaviour Cloning ist einer der ersten Ansätze, der beim Imitation Learning untersucht wurde. Beim Behaviour Cloning werden die vom Experten beobachteten Zustände direkt imitiert. Dieser Ansatz bietet den Vorteil, dass er intuitiv verständlich und effizient ist. BC weist jedoch mehrere Schwächen auf. Es findet kein *Reasoning* über die Zielzustände oder die Dynamik des Systems statt und die möglichen Intentionen der beobachteten Aktion werden nicht beachtet. Außerdem wird beim BC die Experten-Demonstration als optimal angenommen, was in der Praxis jedoch nicht immer zutrifft. Gleichzeitig kann es mehrere optimale Trajektorien geben. Führt der Experte die Aufgabe bei der Demonstration auf verschiedene optimale Weisen auf, ist es sehr schwierig für den Algorithmus eine Policy abzuleiten. Zudem neigt BC zu Instabilität, wenn die beobachteten Zustände von den demonstrierten Zuständen abweichen, da keine Policy bekannt ist, wie mit fehlerhaften Zuständen umgegangen werden soll. In diesem Fall verstärken sich die Fehler, so dass das System instabil werden kann [41].

Direct Policy Learning: Bei Direct Policy Learning via Interactive Expert Demonstrations wird das Problem des überwachten Lernens auf eine Sequenz solcher Supervised-Learning-Probleme reduziert. Das bedeutet, dass der Roboter Schritt für Schritt lernt, die richtige Aktion für eine bestimmte Beobachtung auszuführen, indem direkt aus den Demonstrationen eines Experten gelernt wird. Es gibt dabei zwei Ansätze:

- *Data Aggregation*: Hier wird der Trainingsdatensatz schrittweise erweitert. Der Roboter führt die aktuelle Policy aus, immer wenn Fehler oder neue Zustände auftreten, kann der Experte eingreifen und zusätzliche Demonstrationen liefern. Diese neuen Daten werden zum bestehenden Datensatz hinzugefügt und das Modell wird erneut trainiert [41].
- *Policy Aggregation*: Statt neue Daten zu sammeln, werden zusätzlich die bisherigen Policies miteinander kombiniert oder angepasst, um ein robusteres Verhalten zu erreichen [41].

Direct Policy Learning ermöglicht eine schnelle und direkte Verbesserung des Verhaltens, da das System unmittelbar aus Expertendemonstrationen lernt. Es erfordert keine komplexe *reward function* oder Umgebungsmodell und ist auch in dynamischen Umgebungen einsetzbar. Allerdings kann es bei unbekannten Situationen, also unbekannten States, zu Fehlern kommen (Distribution Shift). Oft sind zudem relativ viele qualitativ hochwertige Demonstrationen notwendig, um ein robustes Verhalten zu erreichen [41].
IRL: Ziel des Inverse Reinforcement Learning (IRL) ist es, anstelle einer den Experten direkt imitierenden Policy eine Belohnungsfunktion (Reward Function) zu lernen. Dabei wird angenommen, dass die Expertedemonstrationen optimal sind. Sobald die Reward-Funktion gelernt wurde, kann eine Policy für die zu lernende Aufgabe optimiert werden, die die Reward-Funktion maximiert. Oft gibt es mehrere Belohnungsfunktionen, welche das vom Experten beobachtete Verhalten erklären können. Dies wird als Mehrdeutigkeit der Belohnungsfunktion (Reward Ambiguity) bezeichnet. Es wird zwischen folgenden Ansätzen unterschieden:

- *Max-Margin Planning*: Hier wird die Reward Ambiguity gelöst, indem eine Belohnungsfunktion gelernt wird, bei der die Aktionen des Experten gegenüber anderen möglichen Aktionen einen möglichst großen Abstand (Margin) haben. Dadurch wird sichergestellt, dass die Expertenstrategie klar besser bewertet wird als andere Strategien.
- *Maximum Entropy IRL*: Bei diesem Ansatz wird die Reward Ambiguity gelöst, indem nach einer Belohnungsfunktion gesucht wird, die nicht nur die Demonstrationen erklärt, sondern gleichzeitig möglichst wenig zusätzliche Annahmen macht. Sie erlaubt viele mögliche Verhaltensweisen und führt dadurch zu robusterem Verhalten, indem Unsicherheiten in den Demonstrationen berücksichtigt werden. Es wird nach der Reward Function gesucht, bei der Experten-Demonstrationen eine maximale Entropie-Verteilung aufweisen.

IRL bietet den großen Vorteil, dass nicht nur ein bestimmtes Verhalten (policy) imitiert wird, sondern eine zugrundeliegende Reward-Funktion erlernt wird, was flexibles und anpassungsfähiges Verhalten in neuen Situationen ermöglicht. Dadurch kann die anschließend gelernte Policy oft besser generalisieren als bei direktem Imitieren der Experten-Demonstration. Allerdings ist IRL in der Praxis oft deutlich komplexer und rechenintensiver als direktes Imitation Learning. Die Belohnungsfunktion eindeutig zu rekonstruieren ist aufgrund der Reward-Ambiguity schwer [42].

Imitation Learning wird aktuell intensiv erforscht. Besonders vielversprechende Ansätze aus der aktuellen Forschung sind:

- **ACT**: Besonders einflussreich durch sehr gute Ergebnisse im Bereich feinfühliger, bimanualer Manipulation. Publikation durch [43] und [38].
- **Diffusion Policy**: neuartiger Ansatz, bei dem Aktionssequenzen mittels Diffusionsmodellen generiert werden. Publikation durch [44] und [45].
- **TD-MPC Policy**: Kombination von Model Predictive Control mit Reinforcement Learning über temporale Differenzmethoden. Publikation durch [46], [47] und [48].
- **VQ-BeT**: Verbindung von Imitation Learning und Quantisierungstechniken zur robusten Aktionsgenerierung. Publikation durch [49].

Von den aufgeführten Ansätzen aus dem Stand der Technik wird ACT nachfolgend erläutert, da dieser in vielen aktuellen Publikationen für Imitation Learning an humanoiden Robotern

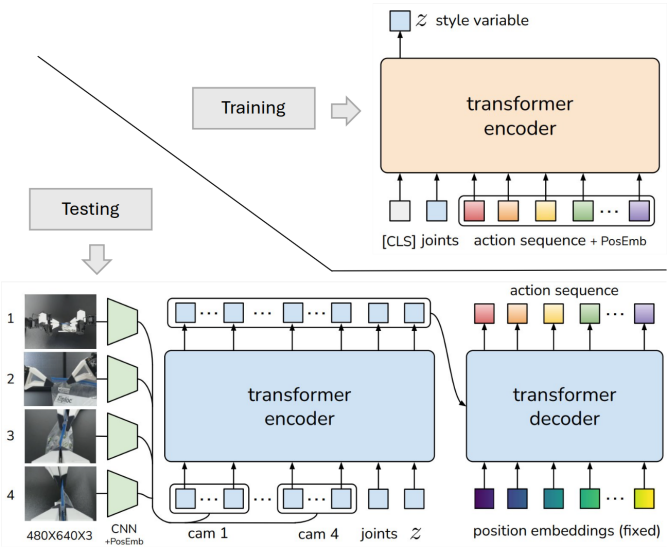


Abbildung 15: Architektur von Action Chunking with Transformers (ACT). Das Modell wird als *Conditional Variational Auto Encoder* trainiert und besteht aus einem Encoder und einem Decoder. **Oben (Training):** Der Encoder komprimiert die Trainingsdaten, bestehend aus Gelenkwinkeln und Aktions-Sequenzen. **Unten (Testing):** Der Decoder, bzw. die *Policy* von ACT kombiniert Bilder aus verschiedenen Blickwinkeln mit Gelenkwinkeln. Anschließend wird eine Aktionssequenz vorhergesagt. Quelle: [43].

verwendet wird, wie z. B. [33] und [35].

ACT: ACT (Action Chunking with Transformers) beschreibt einen Ansatz, bei dem komplexe Manipulationsaufgaben durch das Lernen von kurzen, zusammengefassten Aktions-Sequenzen (Chunks) dargestellt werden. In Abbildung 15 ist der Aufbau des ACT-Modells dargestellt. Das Modell ist wird ähnlich zu einem CVAE (Conditional Variational Auto Encoder) aufgebaut und wird wie ein CVAE trainiert [43].

- **Training**: Im Training komprimiert der Encoder die Trainingsdaten, welche die beobachteten Gelenkwinkel (joints) und Aktionssequenzen (action sequence) enthalten, in eine Stil-Variable (style variable) z [43].
- **Testing**: Im Testing werden aus den Kamerabildern mit Hilfe eines Convolutional Neural Networks (CNN) die wichtigsten Informationen extrahiert und in einer repräsentativen Matrix abgespeichert. Es können mehrere Kamerabilder aus verschiedenen Kamerawinkeln verwendet werden. Als Variable z wird im Testing der Mittelwert der vorherigen Action-Sequenz verwendet. Zusammen mit den Gelenkwinkeln und der Variable z, werden die vorverarbeiteten Kameradaten mit einem Transformer-Encoder in eine latente Repräsentation komprimiert. Anschließend werden die Daten mit Hilfe eines Transformer-Decoders in eine Action-Sequenz rekonstruiert [43].

ACT zeigt, dass es möglich ist mit kostengünstiger Hardware feine Manipulation auf hohem Niveau zu erlernen [38], [33], [35]. Dabei zeigt sich, dass die Qualität der Trainingsdaten einen großen Einfluss auf die erreichbaren Ergebnisse hat. Dabei sind insbesondere die Qualität der Demonstrationen und die Qualität der Kameradaten von Bedeutung, weshalb häufig mehrere Kameras aus verschiedenen Winkeln verwendet werden [43]. Beim Ansatz OpenTeleVision [35] wurde in Bezug auf humanoide Robotik herausgestellt, dass es von Vorteil ist bei Verwendung einer VR-Brille zur Teleoperation eine bewegliche Stereokamera mit engem Blickfeld zu verwenden, bei der sich die Kamera auf dem Kopf vom Roboter mitdreht, wenn der Bediener seinen Kopf dreht. Dies hat den Vorteil, dass die Kamera immer das anschaut, was für den Bediener gerade von Interesse ist. So bekommt der ACT-Algorithmus automatisch einen Fokus auf den Bereich, der gerade wichtig ist, was die Resultate deutlich verbessert [35]. In Abbildung 16 ist das Sichtfeld eines humanoiden H1-Roboters von Unitree bei einer statischen Weitwinkelkamera (Static Wide Angle) im Vergleich zu einer aktiv bewegten Kamera mit kleinem Blickfeld (Cropped Active) dargestellt. Dabei ist die Größe der Bilder proportional zu der Anzahl der Pixel dargestellt. Es ist zu beobachten, dass bei wesentlich weniger Pixeln, also einem deutlich kleinerem Bild, mehr relevante Aktionen pro Pixel erfasst werden, als bei der Weitwinkelkamera. Die Reduzierung der Pixel der Kameradaten führt zu einer Beschleunigung der Trainingszeit auf einer NVIDIA 4090 GPU um den Faktor 2. Die aufgezeichnete Aktion konnte in [35] nach Training mit ACT erfolgreich autonom ausgeführt werden.

Nachfolgend sind in Abbildung 17 und Abbildung 18 Demonstrationen für die mit ACT erreichbare feine Manipulation

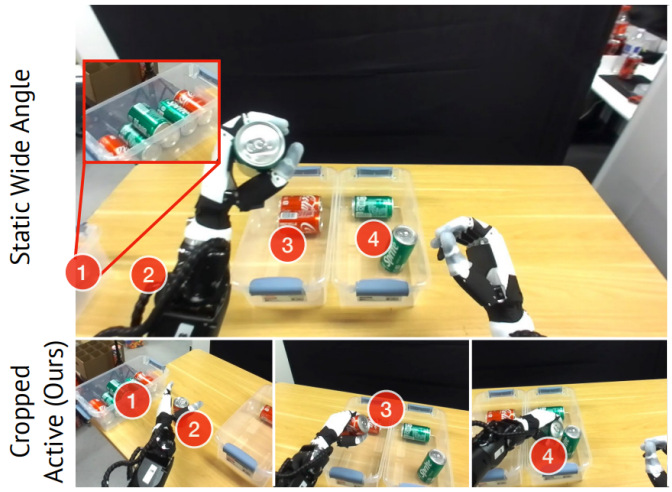


Abbildung 16: Vergleich zwischen Weitwinkelobjektiven und Ausschnittansichten. Aus den ursprünglichen Weitwinkelbildern werden Ausschnitte von relevanten Bereichen gebildet, die in der Abbildung mit roten Punkten gekennzeichnet sind. Die daraus folgende Reduzierung in der Anzahl der Pixel führt zu einer schnelleren Konvergenz des Trainings. Quelle: [35].

aufgeführt. Dabei ist zu beobachten, wie ein Roboter autonom eine Ziploc-Tüte öffnet und autonom Batterien in eine Fernbedienung einführt.

Abbildung 17: **Ziploc ACT:** Öffnen einer Ziploc-Tüte, welche aufrecht auf einem Tisch steht, mit Hilfe des ACT-Algorithmus. Die Tüte wird entlang einer 15 cm langen weißen Linie zufällig positioniert und aus etwa 5 cm Höhe fallengelassen, um eine zufällige Veränderung von Höhe und Aussehen der Tüte zu erzeugen. Der linke Arm greift zuerst den Körper der Tüte (Subtask #1: Greifen), danach klemmt der rechte Arm den Schieberegler der Tüte ein (Subtask #2: Einklemmen). Anschließend bewegt sich der rechte Arm nach rechts, um die Tüte zu öffnen (Subtask #3: Öffnen). Quelle: [43].

Populäre Frameworks: Software Frameworks für Machine Learning, Reinforcement Learning oder Imitation Learning an Robotern im Speziellen bieten die Möglichkeit durch viele vorimplementierte und gut dokumentierte Funktionen die Entwicklung von IL-Anwendungen deutlich zu vereinfachen. Solche Frameworks stellen teilweise die neusten Algorithmen vorimplementiert zusammen mit Beispiel-Datensätzen für Training und Tests und teilweise vortrainierten Modellen, die für die individuelle Anwendung angepasst werden können. Häufig laufen solche Frameworks, genauso wie aktuelle Forschung, unter Open-Source-Lizenzen, die kommerzielle Verwendung erlauben. Der Quellcode ist also frei verfügbar und kann für die individuellen Zwecke angepasst werden. Die Frameworks können oft in Python verwendet werden, einer leicht zu benutzenden und weit verbreiteten Programmiersprache. Im Hintergrund laufen die Frameworks häufig auf PyTorch, einer Open-Source-Software, die in der Entwicklung von KI-Anwendungen weit verbreitet ist. Besonders relevant für das Imitation Learning

Abbildung 18: **Batterie ACT:** Einsetzen einer Batterie in eine Fernbedienung. Die Fernbedienung wird entlang einer 15 cm langen weißen Linie zufällig positioniert. Die Rotationen der Batterien werden zufällig festgelegt. Der rechte Arm greift zuerst die Batterie (Subtask #1: Greifen) und setzt sie dann in den Batterieschacht ein (Subtask #2: Platzieren). Der linke Arm drückt auf die Fernbedienung, um ein Verrutschen zu verhindern, während der rechte Arm die Batterie vollständig hineinschiebt (Subtask #3: Einsetzen). Quelle: [43].

sind die Frameworks *Lerobot* und *Stable Baselines3*, die viele Funktionen zur Verfügung stellen, die das Training erleichtern [50], [51]. Wenn kein realer Roboter zur Verfügung steht können leistungsstarke Simulationsumgebungen wie *MuJoCo* oder *NVIDIA Isaac Lab* verwendet werden. *NVIDIA Isaac Lab* ist dabei nutzungsfreundlicher, besser dokumentiert und an sich quelloffen, basiert jedoch auf der proprietären, aber kostenlosen Physik-Engine PhysX und kann nur mit NVIDIA Grafikarten verwendet werden [52]. MuJoCo ist dafür vollständig Open Source, unterstützt mehrere Grafikkartenhersteller und auch die ausschließliche Verwendung der CPU und benötigt weniger Rechenressourcen [53, 54, 55].

Aktuelle Limitationen und Forschungsschwerpunkte: Beim Imitation Learning verbleiben, trotz der großen Fortschritten und vielversprechende Ergebnissen der letzten Jahren weiterhin Herausforderungen und Limitationen, die Gegenstand der aktuellen Forschung sind.

Eine Limitation ist die hohe Abhängigkeit von der Qualität und Vielfalt der Expertendemonstrationen. Sind die Trainingsdaten aus den Demonstrationen unvollständig, fehlerhaft oder nicht repräsentativ für alle Szenarien, verschlechtert sich die Generalisierungsfähigkeit der erlernten Policy. [56].

Weiterhin ist das Lernen von langfristigen Abhängigkeiten mit reinem Imitation Learning herausfordernd, da es schwierig ist in den Imitation-Learning-Algorithmen einen längeren Zeithorizont in Verbindung zu setzen [35], [43].

Zudem ist es schwierig mit ausschließlich kamerabasierten Systemen besonders feinfühlige Manipulationsaufgaben durchzuführen, da dem Experten-Demonstrator das haptische Feedback fehlt. Eine Lösung dafür könnte der Einsatz von Force-Feedback-Handschuhen (vgl. Abschn. 5.1) sein [35].

Abschließend lässt sich daher festhalten, dass Imitation Learning ein gutes Werkzeug darstellt, um dem Roboter neue Fähigkeiten kostengünstig anzulernen und für die High-Level-Strategie in ein größeres Modell oder Netzwerk eingebunden werden sollte, um dem Roboter ein hohes Maß an Autonomie zu ermöglichen.

5.3. Visual Language Action Model

Visual Language Action Models (VLA) basieren auf dem Zusammenspiel von Computer Vision, Natural Language Processing (NLP) und Reinforcement Learning, um Roboter in die Lage zu versetzen, visuelle Eindrücke und natürliche sprachliche Anweisungen zu verarbeiten und in präzise motorische Aktionen umzusetzen. Technisch beruhen diese VLA-Modelle auf einem mehrstufigen Verarbeitungsprozess, bei dem zunächst low-level visuelle Informationen extrahiert und zu abstrakten Repräsentationen transformiert werden. Parallel dazu wird die natürliche Sprache in handlungsrelevante Instruktionen übersetzt, welche von den Aktionsmodulen in präzise motorische Steuerbefehle umgesetzt werden, wie in Abbildung 19 dargestellt [57, 58].

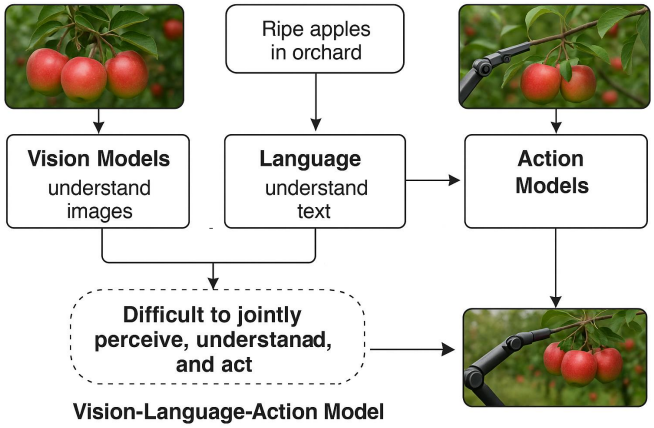


Abbildung 19: VLAs nutzen parallel Modelle zum Verarbeiten von aktuellen Bilddaten und natürlicher Sprache als Eingaben, um daraus eine Aktion für den nächsten Zeitschritt zu generieren; aus [58].

Im Vergleich zu klassischen Imitation-Learning-Ansätzen, bei denen Roboter ausschließlich durch die direkte Nachahmung menschlicher Aktionen trainiert werden, ermöglichen VLA-Systeme eine tiefgreifende Generalisierung. Der semantische Gehalt von Sprachbefehlen wird genutzt, um universell anwendbare Handlungsstrategien zu entwickeln. So zeigt das OpenVLA-Modell, das mit 970.000 Roboterdemonstrationen trainiert wurde, in einer Vielzahl von Manipulations- und Steuerungsaufgaben um bis zu 16,5% höhere Erfolgsraten als herkömmliche Modelle [57, 58]. Auch der NVIDIA Isaac GR00T N1, der Daten aus Videos sowie Simulationen verarbeitet, erreicht in Aufgaben der Objektmanipulation zwischen 70% und 80% und generalisiert auch auf neue Objekte, die nicht Teil des Trainingsdatensatzes waren [59]. Ergänzend dazu adressieren die Modelle $\pi 0$ und $\pi 0.5$ die Problematik der Generalisierung, indem sie den Bedarf an expliziten Demonstrationen minimieren und durch adaptive Sprach- und Kontextanalysen in Multi-Task-Umgebungen Erfolgsraten von etwa 50 bis 65% erzielen [60, 58].

Ein Nachteil der VLA-Ansätze ist der hohe Rechenaufwand, der zu einer langsamen Regelung führt. Ein Weg, die Steuerung mit VLAs zu beschleunigen, ist die Integration eines dualen Steuerungsansatzes, bei dem ein schnelles, reaktives Pfadgenerationsmodul (System 1) mit einem langsameren, strategischen Planungsmodul (System 2) gekoppelt wird, wie es mit Gemma vorgestellt wurde [61]. Diese Architektur ermöglicht es Gemma, kontextübergreifende Aufgaben wie die Manipulation komplexer Objekte mit Erfolgsraten von durchschnittlich 79% über unterschiedliche Aufgaben zu bewältigen. Einzelne Aufgaben werden dabei bereits jetzt fehlerfrei durchgeführt.

Trotz der erzielten Fortschritte bestehen weiterhin Herausforderungen: Die hohen Anforderungen an multimodale Trainingsdaten, der erhebliche Rechenaufwand für die Inferenz (Modellausführung) und die Optimierung der Fehlertoleranz bei mehrdeutigen Sprachbefehlen bleiben zentrale Problemfelder. Zukünftige Forschungsarbeiten zielen darauf ab, durch die Erweiterung von Sensornetzwerken, die Optimierung von Echtzeit-Inferenzstrategien sowie die Verbesserung der Generalisierung, Robustheit und Anpassungsfähigkeit dieser Systeme weiter zu steigern. Die VLAs sind dennoch die aktuell vielversprechendste Methode für die autonome Steuerung von Robotern in neuen Umgebungen [58].

5.4. Zwischenfazit zum anlernen neuer Aufgaben

In den letzten Jahren wurden erhebliche Fortschritte in der Art und Weise erzielt, wie Roboter neue Aufgaben erlernen können. Durch Ansätze wie das Imitation Learning sind sie mittlerweile in der Lage, bereits aus wenigen Demonstrationen zu lernen, Aufgaben autonom auszuführen und dabei robust gegenüber Umgebungsveränderungen zu bleiben, beispielsweise einer veränderten Positionierung von Objekten.

Ein weiterer vielversprechender Ansatz sind Visual Language Action Models, die es ermöglichen, Aufgaben ohne vorherige Demonstration allein anhand von Anweisungen in natürlicher Sprache auszuführen. Diese Modelle nutzen große multimodale Modelle, um aus textbasierten Anweisungen und Kamerabildern die entsprechenden Bewegungen für den nächsten Zeitschritt zu planen.

Beide Methoden erlauben es Robotern, neue Aufgaben schnell und effizient zu erlernen, ohne dass für das Training eine aufwendige manuelle Implementierung erforderlich ist. Insbesondere in der humanoiden Robotik stellen sie einen entscheidenden Fortschritt dar, da humanoide Roboter als vielseitige Systeme für unterschiedliche Aufgaben konzipiert sind und so der Integrationsaufwand deutlich gesenkt werden kann.

6. Bewertung verschiedener Einsatzmöglichkeiten

Nachdem in den vorangegangenen Kapiteln die aktuellen Forschungsrichtungen der humanoiden Robotik diskutiert wurden, beschäftigt sich dieser Abschnitt mit konkreten Anwendungsfeldern und diskutiert die Einsatzmöglichkeit von humanoiden Robotern.

6.1. Produzierendes Gewerbe

Im produzierenden Gewerbe bieten sich diverse Einsatzmöglichkeiten für humanoide Roboter. Hierbei können sowohl bimanuale Roboter, also Roboter mit zwei Armen, Roboter auf Rädern sowie humanoide Roboter mit Beinen, also bipedale Roboter, genutzt werden.

Das Nutzen von bipedalen Robotern ergibt dabei insbesondere Sinn, wenn es sich um komplexe Umgebungen handelt, bei denen z. B. Stufen oder andere Unebenheiten zu überwinden sind oder bei denen der Platz eingeschränkt ist. Für ebene Industriehallen kann der Einsatz von mobilen Roboterplattformen sinnvoller sein.

Beide lassen sich jedoch für ähnliche Anwendungen nutzen. Aufgaben lassen sich vereinfacht auf zwei Achsen auftragen, den Herausforderungen der Umgebungsbedingungen und der Aufgabenkomplexität. Beispielhaft wurde eine qualitative Einordnung verschiedener Aufgaben in Abbildung 20 vorgenommen. Aufgaben, die immer gleich bleiben, mit niedrigen Anforderungen an die Genauigkeit haben hierbei eine geringe Aufgabenkomplexität, während sich häufig ändernde Aufgaben oder Aufgaben, die eine hohe Präzision erfordern, eine hohe Aufgabenkomplexität aufweisen. Die Komplexität der Umgebungsbedingungen hängt von verschiedenen Umgebungsfaktoren ab, wie z. B. Feuchtigkeit oder Nässe, stark wechselnde Umgebungen, Hitze, Kälte oder elektromagnetische oder ionisierende Strahlung.

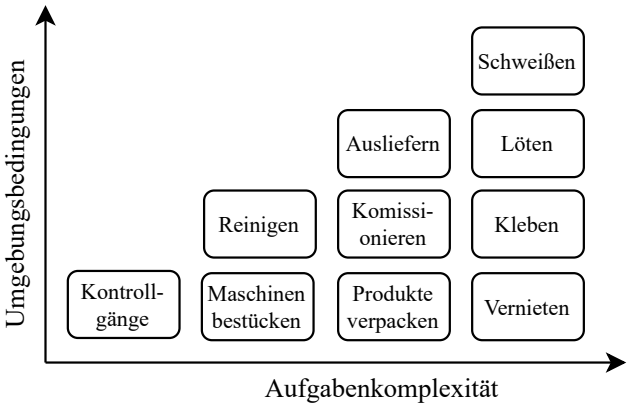


Abbildung 20: Qualitative Einordnung von verschiedenen Aufgaben über die Aufgabenkomplexität und die Umgebungsbedingungen.

Während die Automatisierbarkeit von Aufgaben über die Aufgabenkomplexität maßgeblich an den aktuellen Stand der Forschung gebunden ist, so ist die Automatisierbarkeit über die Umgebungsbedingungen maßgeblich von der robotischen Hardware limitiert. Da es bereits Roboter gibt, die selbst mit sehr rauen Umgebungen zurechtkommen, ist davon auszugehen, dass wenn die Nachfrage nach entsprechenden Systemen vorliegt, die Hersteller von humanoiden Robotern in Zukunft auch Roboter produzieren werden, die gegen diverse Umgebungseinflüsse geschützt sind.

Ein großer Vorteil, den solche generalistischen Systeme haben, ist dabei, dass sie ohne Umrüsten für verschiedene Aufgaben eingesetzt werden können. Sie können also auch über einen Tag für verschiedene Aufgaben dynamisch genutzt werden, was auch das Automatisieren von Aufgaben ermöglicht, die bis jetzt eine zu geringe Auslastung hatten, um sie zu automatisieren. Der Roboter kann so z. B. morgens einen Kontrollgang machen, über die Mittagszeit Maschinen bestücken und am Abend bereits Teile für den nächsten Tag kommissionieren.

6.2. Unterstützung im Haushalt

Auch als Unterstützung im Haushalt erfahren humanoide Roboter aktuell hohes Interesse. Hierbei sind insbesondere auch Roboter mit Beinen von Interesse, da sie sich in komplexen

häuslichen Umgebungen, in denen es ggf. Stufen oder Treppen gibt, besser zurechtfinden.

Auch bietet das häusliche Umfeld eine Großzahl unterschiedlicher Aufgaben, die aber jeweils nur verhältnismäßig selten anfallen, was eine Automatisierung nicht lohnenswert macht. So ist der Staubsaugerroboter aktuell einer der wenigen Roboter, die größere Adoption im privaten Bereich gefunden haben. Einen zusätzlichen Roboter anzuschaffen, der aufräumt, einen, der die Fenster putzt, einen, der den Müll rausbringt usw. würde sich nicht lohnen, da alle diese Aufgaben nur selten anfallen und die Roboter damit einen großen Teil der Zeit nur herumstehen würden. Dieses Umfeld ist ein beispielhaftes, in dem der Einsatz von generalistischen Robotern Sinn ergibt.

An unterschiedlichen Tagen und zu unterschiedlichen Zeiten am Tag können diese Roboter dann diverse unterschiedliche Aufgaben ausführen, um so ihre Auslastung zu erhöhen und entsprechend finanziell tragbar zu sein.

6.3. Medizinischer Sektor

Auch im medizinischen Sektor erfahren humanoide Roboter großes Interesse, insbesondere aufgrund des hohen Fachkräftemangels.

Hierbei könnten humanoide Roboter insbesondere für Hilfstätigkeiten eingesetzt werden. Das Bringen von Essen oder der Transport von Materialien oder Proben ist hierbei denkbar.

Die direkte Behandlung oder Pflege von Patienten hingegen erfordert noch eine große Menge technischer Innovationen sowie ein höheres Vertrauen und eine höhere Akzeptanz von Patienten und eine Automatisierung erscheint aus ethischen Gründen auch weniger erstrebenswert.

6.4. Unterhaltungssektor

Ein großer Treiber von Innovationen in der mobilen Robotik war in der Vergangenheit der Unterhaltungssektor. Insbesondere der Disney-Konzern setzt in seinen Vergnügungsparks auf moderne Robotik und Animatronik.

Roboter können als interaktive Charaktere in Freizeitparks, Theatern und Kinos eingesetzt werden, um Besuchern ein immersives Erlebnis zu bieten, indem sie in Echtzeit auf Fragen reagieren und sich dynamisch an die Interaktionen mit dem Publikum anpassen. Darüber hinaus können sie als Moderatoren von Live-Veranstaltungen oder Messen fungieren, bei denen sie Informationen präsentieren und individuell mit Teilnehmern interagieren. Schließlich könnten sie in Musik- und Kunstperformances eine tragende Rolle übernehmen, sei es durch das Spielen von Instrumenten, das Erzeugen choreografierter Tanzbewegungen oder das Interagieren mit Künstlern in kreativen Kollaborationen.

7. Aktuelle Limitationen

Trotz der vielen Innovationen auf der Software- und Hardwareseite gibt es zum aktuellen Zeitpunkt noch Limitationen, welche die größere Verbreitung von humanoiden Robotern zurückhalten.

Hierbei ist allem voran die Verfügbarkeit zu nennen. Kaum Systeme sind (Stand Mai 2025) zu kaufen und die, die es sind, erfordern in der Regel noch eine große Menge Entwicklungsarbeit auf Kundenseite, um produktiv Aufgaben durchzuführen. Es ist aber davon auszugehen, dass sich dies im aktuellen dynamischen Umfeld rasch ändern wird. Auch ist Robot as a Service (RaaS) als Modell denkbar, wobei Roboter inklusive der Software vermietet werden.

Auch hardwareseitig gibt es noch Einschränkungen. Die maximale Traglast der meisten robotischen Systeme ist unter der eines Menschen. Auch fehlen aktuell Roboter, die geschützt gegen Staub, Wasser und weitere Umgebungseinflüsse sind, wobei hier davon auszugehen ist, dass diese in folgenden Hardwaregenerationen folgen werden, wenn es einen entsprechenden Marktbedarf gibt.

Auch psychologische Faktoren spielen eine Rolle und können den Einsatz verlangsamen. Es ist unklar, ob und wie Menschen an bestehenden Arbeitsplätzen mit den Robotern gemeinsam arbeiten wollen. Hierbei ist es insbesondere wichtig das Personal mit einzubinden und die Technologie näher zu bringen.

Zuletzt spielt das Thema Sicherheit eine zentrale Rolle in der Verbreitung von humanoiden Robotern. Insbesondere bipedale Roboter, also Roboter mit zwei Beinen, müssen sich aktiv balancieren und haben das Risiko, bei Systemfehlern umzufallen und so Gegenstände und Personen um sie herum zu gefährden. Der Sicherheitsaspekt beschäftigt die Forschung sowie die Normung aktuell noch aktiv, ohne dass hier bereits ein finales Ergebnis erreicht worden ist. Mit der ISO/AWI 25785-1 ist aktuell die erste Normung, die sich mit der Sicherheit von bipedalen Robotern im industriellen Kontext beschäftigt, in Arbeit.

8. Zusammenfassung

Humanoide Robotik stellt eine große Chance in der aktuellen Transformation der Industrie und im Angesicht des demographischen Wandels dar.

Diese Studie geht auf aktuelle Trends im Bereich der Fortbewegung ein, insbesondere auf die Fragen, wie ein Roboter das Laufen erlernt. Daran angeschlossen wurden aktuelle Trends im Imitation Learning und der Visual Language Action Model (VLA) aufgezeigt. Beide erlauben den Robotern neue Aufgaben anhand weniger Demonstrationen robust zu erlernen. Insbesondere die VLA erlauben den Robotern sogar vollständig neue Aufgaben auszuführen, alleinig anhand einer Beschreibung der Aufgabe in natürlicher Sprache. Durch diese Techniken können humanoide Roboter in diversen Bereichen eingesetzt werden. Verschiedene Einsatzmöglichkeiten wurden hierbei diskutiert. Auch die aktuellen Limitationen der humanoiden Roboter wurden aufgezeigt. Viele davon sind auf die frühen Hardwaregenerationen zurückzuführen. Andere Bereiche wie die Sicherheit werden noch aktiv in der Forschung und Normung bearbeitet.

Abschließend bieten humanoide Roboter also ein großes Potential. Je nach Einsatzzweck kann ein zweiarmiger Roboter auf einer fahrenden Basis oder ein laufender Roboter sinnvoller

sein. Es ist davon auszugehen, dass Software und Daten, welche mit einem System aufgenommen worden sind, gut auf andere Systeme übertragbar sind. Solche generalistischen Systeme erlauben insbesondere die Automatisierung von häufig wechselnden oder nur selten anfallenden Aufgaben, für die robotische Lösungen bis jetzt nicht kostendeckend umsetzbar waren. Klassische ortsgebundene Systeme werden weiter dort von Vorteil sein, wo die gleiche Aufgabe schnell und oft über längere Zeiträume wiederholt wird. Die humanoiden Roboter werden die aktuellen Systeme also oftmals nicht ersetzen, sondern erlauben es, weitere Aufgaben zu automatisieren.

Förderhinweis

Diese Studie wurde durch das Transformationsnetzwerk neu/wagen finanziert. Die Initiative neu/wagen ist Teil des Förderprogramms „Transformationsstrategien für Regionen der Fahrzeug- und Zuliefererindustrie“ des Bundesministeriums für Wirtschaft und Energie (BMWE).

Erklärung

Die Autoren erklären, dass kein Interessenkonflikt vorliegt. Beim Erstellen dieser Studie wurde zum Teil generative KI zur stilistischen Überarbeitung genutzt. Alle Inhalte wurden auf ihre Richtigkeit geprüft.

Literatur

[1] A. Burstedde, J. Tiedemann, IW-Arbeitsmarktfortschreibung bis 2027 (2024).

[2] A. Burstedde, G. Kolev-Schaefer, Die Kosten des Fachkräftemangels (2024).

[3] E. Ahlers, V. Q. Villalobos, Fachkräftemangel in Deutschland? Befunde der WSI-Betriebs- und Personalrätebefragung 2021/22 (2022).

[4] W. Sun, L. Chen, Y. Su, B. Cao, Y. Liu, Z. Xie, Learning humanoid locomotion with world model reconstruction (2025). arXiv:2502.16230.
URL `https://arxiv.org/abs/2502.16230`

[5] Y. Tong, H. Liu, Z. Zhang, Advancements in humanoid robots: A comprehensive review and future prospects, IEEE/CAA Journal of Automatica Sinica 11 (2) (2024) 301–328.

[6] B. Singh, R. Kumar, V. P. Singh, Reinforcement learning in robotic applications: a comprehensive survey, Artificial Intelligence Review 55 (2) (2022) 945–990.

[7] D. Gaertner, Treppensteigen mit Laufrobotern: Untersuchung von Belohnungsfunktionen für Deep Reinforcement Learning, Master's thesis, Hochschule für Angewandte Wissenschaften Hamburg (2021).

[8] M. Ciupek, Humanoide Roboter: Teure Spielerei oder nützlicher Helfer?, Whitepaper, INGENIEUR.de, Düsseldorf, iNGENIEUR.de 11/2024 (nov 2024).

[9] J. Ren, T. Huang, H. Wang, Z. Wang, Q. Ben, J. Pang, P. Luo, VB-Com: Learning vision-blind composite humanoid locomotion against deficient perception (2025). arXiv:2502.14814.
URL `https://arxiv.org/abs/2502.14814`

[10] Y. Li, Deep reinforcement learning: An overview (2018). arXiv:1701.07274.
URL `https://arxiv.org/abs/1701.07274`

[11] N. Rudin, D. Hoeller, P. Reist, M. Hutter, Learning to walk in minutes using massively parallel deep reinforcement learning, in: A. Faust, D. Hsu, G. Neumann (Eds.), Proceedings of the 5th Conference on Robot Learning, Vol. 164 of Proceedings of Machine Learning Research, PMLR, 2022, pp. 91–100.
URL `https://proceedings.mlr.press/v164/rudin22a.html`

[12] R. S. Sutton, A. G. Barto, Reinforcement Learning: An Introduction, 2nd Edition, The MIT Press, 2018.
URL `http://incompleteideas.net/book/the-book-2nd.html`

[13] J. Long, J. Ren, M. Shi, Z. Wang, T. Huang, P. Luo, J. Pang, Learning humanoid locomotion with perceptive internal model (2024). arXiv:2411.14386.
URL `https://arxiv.org/abs/2411.14386`

[14] H. Wang, Z. Wang, J. Ren, Q. Ben, T. Huang, W. Zhang, J. Pang, BeamDojo: Learning agile humanoid locomotion on sparse footholds (2025). arXiv:2502.10363.
URL `https://arxiv.org/abs/2502.10363`

[15] T. Miki, J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, M. Hutter, Learning robust perceptive locomotion for quadrupedal robots in the wild, Science Robotics 7 (62) (jan 2022). doi:10.1126/scirobotics.abk2822.

[16] C. Ji, D. Liu, W. Gao, S. Zhang, Learning-based locomotion control fusing multimodal perception for a bipedal humanoid robot, Biomimetic Intelligence and Robotics 5 (1) (2025) 100213. doi:10.1016/j.birob.2025.100213.

[17] H. Duan, B. Pandit, M. S. Gadde, B. Van Marum, J. Dao, C. Kim, A. Fern, Learning vision-based bipedal locomotion for challenging terrain, in: 2024 IEEE International Conference on Robotics and Automation (ICRA), 2024, pp. 56–62. doi:10.1109/ICRA57147.2024.10611621.

[18] W. Sun, B. Cao, L. Chen, Y. Su, Y. Liu, Z. Xie, H. Liu, Learning perceptive humanoid locomotion over challenging terrain (2025). arXiv:2503.00692.
URL `https://arxiv.org/abs/2503.00692`

[19] X. Cheng, K. Shi, A. Agarwal, D. Pathak, Extreme parkour with legged robots, in: 2024 IEEE International Conference on Robotics and Automation (ICRA), 2024, pp. 11443–11450. doi:10.1109/ICRA57147.2024.10610200.

[20] Q. Zhang, G. Han, J. Sun, W. Zhao, C. Sun, J. Cao, J. Wang, Y. Guo, R. Xu, Distillation-PPO: A novel two-stage reinforcement learning framework for humanoid robot perceptive locomotion (2025). arXiv:2503.08299.
URL `https://arxiv.org/abs/2503.08299`

[21] Z. Zhuang, S. Yao, H. Zhao, Humanoid parkour learning (2024). arXiv:2406.10759.
URL `https://arxiv.org/abs/2406.10759`

[22] X. Gu, Y.-J. Wang, X. Zhu, C. Shi, Y. Guo, Y. Liu, J. Chen, Advancing humanoid locomotion: Mastering challenging terrains with denoising world model learning (2024). arXiv:2408.14472.
URL `https://arxiv.org/abs/2408.14472`

[23] I. Radosavovic, T. Xiao, B. Zhang, T. Darrell, J. Malik, K. Sreenath, Real-world humanoid locomotion with reinforcement learning, Science Robotics 9 (89) (apr 2024). doi:10.1126/scirobotics.adi9579.

[24] W. Xie, C. Bai, J. Shi, J. Yang, Y. Ge, W. Zhang, X. Li, Humanoid whole-body locomotion on narrow terrain via dynamic balance and reinforcement learning (2025). arXiv:2502.17219.
URL `https://arxiv.org/abs/2502.17219`

[25] T. Miki, L. Wellhausen, R. Grandia, F. Jenelten, T. Homberger, M. Hutter, Elevation mapping for locomotion and navigation using GPU, in: 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2022, pp. 2273–2280. doi:10.1109/IROS47612.2022.9981507.

[26] P. Fankhauser, M. Bloesch, M. Hutter, Probabilistic terrain mapping for mobile robots with uncertain localization, IEEE Robotics and Automation Letters 3 (4) (2018) 3019–3026. doi:10.1109/LRA.2018.2849506.

[27] P. Fankhauser, M. Bloesch, C. Gehring, M. Hutter, R. Siegwart, Robot-centric elevation mapping with uncertainty estimates, in: Mobile Service Robotics, WORLD SCIENTIFIC, 2014. doi:10.1142/9789814623353_0051.

[28] W. Xu, F. Zhang, FAST-LIO: A fast, robust LiDAR-inertial odometry package by tightly-coupled iterated kalman filter, IEEE Robotics and Automation Letters 6 (2) (2021) 3317–3324. doi:10.1109/LRA.2021.3064227.

[29] W. Xu, Y. Cai, D. He, J. Lin, F. Zhang, FAST-LIO2: Fast direct LiDAR-inertial odometry, IEEE Transactions on Robotics 38 (4) (2022) 2053–2073. doi:10.1109/TRO.2022.3141876.

[30] C. Bai, T. Xiao, Y. Chen, H. Wang, F. Zhang, X. Gao, Faster-LIO: Lightweight tightly coupled Lidar-inertial odometry using parallel sparse incremental voxels, IEEE Robotics and Automation Letters 7 (2) (2022) 4861–4868. doi:10.1109/LRA.2022.3152830.

[31] R. Yang, M. Zhang, N. Hansen, H. Xu, X. Wang, Learning vision-guided quadrupedal locomotion end-to-end with cross-modal transformers (2022). arXiv:2107.03996.
URL https://arxiv.org/abs/2107.03996

[32] D. Hoeller, N. Rudin, C. Choy, A. Anandkumar, M. Hutter, Neural scene representation for locomotion on structured terrain, IEEE Robotics and Automation Letters 7 (4) (2022) 8667–8674. doi:10.1109/LRA.2022.3184779.

[33] Z. Fu, Q. Zhao, Q. Wu, G. Wetzstein, C. Finn, HumanPlus: Humanoid shadowing and imitation from humans, in: Conference on Robot Learning (CoRL), 2024.
URL https://humanoid-ai.github.io/

[34] Z. Xu, H. Zhang, An immersive virtual reality bimanual telerobotic system with haptic feedback, arXiv (2025). arXiv:2501.00822.
URL https://arxiv.org/abs/2501.00822

[35] X. Cheng, J. Li, S. Yang, G. Yang, X. Wang, Open-TeleVision: Teleoperation with immersive active visual feedback, arXiv preprint arXiv:2407.01512 (2024).
URL https://robot-tv.github.io/

[36] U. Robotics, Unitree G1 developer documentation, https://support.unitree.com/home/en/G1_developer, accessed: 2025-04-23 (2024).

[37] Y. Qin, W. Yang, B. Huang, K. Van Wyk, H. Su, X. Wang, Y.-W. Chao, D. Fox, AnyTeleop: A general vision-based dexterous robot arm-hand teleoperation system, in: Robotics: Science and Systems (RSS), 2023.
URL https://yzqin.github.io/anyteleop/

[38] Z. Fu, T. Z. Zhao, C. Finn, Mobile ALOHA: Learning bimanual mobile manipulation with low-cost whole-body teleoperation, in: Conference on Robot Learning (CoRL), 2024.
URL https://mobile-aloha.github.io/

[39] F. R. Noreils, Humanoid robots at work: where are we? (2024). arXiv:2404.04249.
URL https://arxiv.org/abs/2404.04249

[40] Unitree Robotics, Fire control robots, accessed: 2025-04-26 (2024).
URL https://www.unitree.com/industry/fireControl

[41] Stanford University, Lecture 12: Imitation learning I — CS237B: Principles of robot autonomy ii, https://web.stanford.edu/class/cs237b/, accessed: 2025-04-26 (2025).

[42] Stanford University, Lecture 13: Imitation learning II — CS237B: Principles of robot autonomy ii, https://web.stanford.edu/class/cs237b/, accessed: 2025-04-28 (2025).

[43] T. Z. Zhao, V. Kumar, S. Levine, C. Finn, Learning fine-grained bimanual manipulation with low-cost hardware, arXiv preprint arXiv:2304.13705 (2023).

[44] C. Chi, Z. Xu, S. Feng, E. Cousineau, Y. Du, B. Burchfiel, R. Tedrake, S. Song, Diffusion policy: Visuomotor policy learning via action diffusion, The International Journal of Robotics Research (2023). doi:10.1177/02783649241273668.
URL https://diffusion-policy.cs.columbia.edu/

[45] C. Chi, Z. Xu, C. Pan, E. Cousineau, B. Burchfiel, S. Feng, R. Tedrake, S. Song, Universal manipulation interface: In-the-wild robot teaching without in-the-wild robots, in: Proceedings of Robotics: Science and Systems (RSS), 2024. doi:10.15607/RSS.2024.XX.045.
URL https://umi-gripper.github.io/

[46] N. Hansen, X. Wang, H. Su, Temporal difference learning for model predictive control (2022). arXiv:2203.04955.

[47] N. Hansen, H. Su, X. Wang, TD-MPC2: Scalable, robust world models for continuous control (2024).

[48] Y. Feng, N. Hansen, Z. Xiong, C. Rajagopalan, X. Wang, Finetuning offline world models in the real world, in: Proceedings of the 7th Conference on Robot Learning (CoRL), 2023.

[49] S. Lee, Y. Wang, H. Etukuru, H. J. Kim, N. M. M. Shafiullah, L. Pinto, Behavior generation with latent actions, arXiv preprint arXiv:2403.03181 (2024).

[50] R. Cadene, S. Alibert, A. Soare, Q. Gallouedec, A. Zouitine, T. Wolf, LeRobot: State-of-the-art machine learning for real-world robotics in Pytorch, https://github.com/huggingface/lerobot (2024).

[51] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, N. Dormann, Stable-Baselines3: Reliable reinforcement learning implementations, Journal of Machine Learning Research 22 (268) (2021) 1–8.
URL http://jmlr.org/papers/v22/20-1364.html

[52] M. Mittal, C. Yu, Q. Yu, J. Liu, N. Rudin, D. Hoeller, J. L. Yuan, R. Singh, Y. Guo, H. Mazhar, A. Mandlekar, B. Babich, G. State, M. Hutter, A. Garg, Orbit: A unified simulation framework for interactive robot learning environments, IEEE Robotics and Automation Letters 8 (6) (2023) 3740–3747. doi:10.1109/LRA.2023.3270034.

[53] J. Yoon, B. Son, D. Lee, Comparative study of physics engines for robot simulation with mechanical interaction, Applied Sciences 13 (2023) 680. doi:10.3390/app13020680.

[54] E. Todorov, T. Erez, Y. Tassa, MuJoCo: A physics engine for model-based control, in: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, 2012, pp. 5026–5033. doi:10.1109/IROS.2012.6386109.

[55] K. Zakka, B. Tabanpour, Q. Liao, M. Haiderbhai, S. Holt, J. Y. Luo, A. Allshire, E. Frey, K. Sreenath, L. A. Kahrs, C. Sferrazza, Y. Tassa, P. Abbeel, MuJoCo Playground (2025). arXiv:2502.08844.
URL https://arxiv.org/abs/2502.08844

[56] Stanford University, Lecture 14: Learning from human feedback — CS237B: Principles of robot autonomy II, https://web.stanford.edu/class/cs237b/, accessed: 2025-04-29 (2025).

[57] M. J. Kim, K. Pertsch, S. Karamcheti, T. Xiao, A. Balakrishna, S. Nair, R. Rafailov, E. Foster, G. Lam, P. Sanketi, Q. Vuong, T. Kollar, B. Burchfiel, R. Tedrake, D. Sadigh, S. Levine, P. Liang, C. Finn, OpenVLA: An open-source vision-language-action model, arXivArXiv:2406.09246, Version 3 (2024).
URL https://arxiv.org/abs/2406.09246

[58] R. Sapkota, Y. Cao, K. I. Roumeliotis, M. Karkee, Vision-language-action models: Concepts, progress, applications and challenges, arXivArXiv:2505.04769 (2025).
URL https://arxiv.org/abs/2505.04769

[59] Y. Zhu, L. J. Fan, das NVIDIA GEAR Team, NVIDIA Isaac GR00T N1: An open foundation model for humanoid robots, NVIDIA Research (2025).

[60] K. Black, N. Brown, J. Darpinian, K. Dhabalia, D. Driess, A. Esmail, M. Equi, C. Finn, N. Fusai, M. Y. Galliker, D. Ghosh, L. Groom, K. Hausman, B. Ichter, S. Jakubczak, T. Jones, L. Ke, D. LeBlanc, S. Levine, A. Li-Bell, M. Mothukuri, S. Nair, K. Pertsch, A. Z. Ren, L. X. Shi, L. Smith, J. T. Springenberg, K. Stachowicz, J. Tanner, Q. Vuong, H. Walke, A. Walling, H. Wang, L. Yu, U. Zhilinsky, π0.5: a vision-language-action model with open-world generalization, Physical Intelligence (2025).

[61] G. D. Gemini Robotics Team, Gemini Robotics: Bringing AI into the physical world, arXiv 2503.20020 (2025).
URL https://arxiv.org/pdf/2503.20020