

AI Robots and Humanoid AI: Review, Perspectives and Directions

LONGBING CAO, Macquarie University, Australia

In the approximately century-long journey of robotics, humanoid robots made their debut around six decades ago. The rapid advancements in generative AI, large language models (LLMs), and large multimodal models (LMMs) have reignited interest in humanoids, steering them towards real-time, interactive, and multimodal designs and applications. This resurgence unveils boundless opportunities for AI robotics and novel applications, paving the way for automated, real-time and humane interactions with humanoid advisers, educators, medical professionals, caregivers, and receptionists. However, while current humanoid robots boast human-like appearances, they have yet to embody true humaneness, remaining distant from achieving human-like intelligence. In our comprehensive review, we delve into the intricate landscape of AI robotics and AI humanoid robots in particular, exploring the challenges, perspectives and directions in transitioning from human-looking to humane humanoids and fostering human-like robotics. This endeavour synergizes the advancements in LLMs, LMMs, generative AI, and human-level AI with humanoid robotics, omniverse, and decentralized AI, ushering in the era of *AI humanoids* and *humanoid AI*.

CCS Concepts: • **Computing methodologies** → **Artificial intelligence**.

Additional Key Words and Phrases: AI, AI-powered Robotics, AI Robot, Humanoid Robot, AI Humanoid, Humanoid AI, Large Language Models, Large Multimodal Models, Generative AI, Decentralized AI, Metaverse, Humanlevel AI, Humane Humanoid, Humanlike Humanoid, Robotics

ACM Reference Format:

Longbing Cao. 2024. AI Robots and Humanoid AI: Review, Perspectives and Directions. *ACM Comput. Surv.* 0, 0, Article 0 (2024), 37 pages. <https://doi.org/XXXXXXXX.XXXXXXX>

1 INTRODUCTION

In the rich history of robot development spanning a century [17], humanoid robots emerged approximately 60 years ago as a groundbreaking advancement characterized by anthropomorphic forms and appearance [58]. It wasn't until about 30 years ago that humanoid robots began to exhibit notable human-like features such as human-looking structures, senses, behaviors, functions, interactions, and reasoning [25, 51].

In the past decade, fueled by rapid advancements in deep learning and generative AI (GAI), particularly in large language models (LLMs) and large multimodal models (LMMs) [32], humanoid robotics has witnessed remarkable progress, showcasing exceptional performance across various aspects and applications [57]. This convergence of generative and human-level AI with humanoid robotics marks the dawn of the era of humanoid AI.

Humanoid AI has emerged into a human-AI-robotics-web-integrative ecosystem, as illustrated in Fig. 1. Humanoid AI is poised to revolutionize the landscape of the intelligent digital economy, societies, and cultures. Projections suggest a substantial growth trajectory, with the market size of humanoid robots anticipated to reach USD\$13.8 billion by 2028, boasting a staggering compound annual growth rate of 50% per year according to MarketsandMarkets. Goldman Sachs

Author's address: Longbing Cao, Macquarie University, Sydney, NSW, Australia.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

Manuscript submitted to ACM

Manuscript submitted to ACM

1

predicts an even more substantial market valuation of USD\$38 billion by 2035. Meanwhile, generative AI commands an even larger market value, estimated at USD\$2.6-4.4 trillion annually, with a projected 7% increase in global GDP [19].

However, despite these promising projections, only a limited number of humanoids are currently empowered by large language models (LLMs) or driven by generative AI. This discrepancy highlights significant untapped potential, as well as gaps and opportunities within the realm of humanoid AI, particularly in the domain of AI-empowered humanoid robotics.

Humanoid robotics, propelled by recent breakthroughs in data science, machine learning, and AI, such as generative AI (GAI), large language models (LLMs), and large multimodal models (LMMs), has ushered in a new era of revolutionary advancements and possibilities:

- Transitioning from traditional task-specific hand engineering and programming methodologies to semi-task-specific, task-agnostic, or even open-task applications; fostering greater versatility and flexibility.
- Evolving from predefined and rule-driven behaviors to online, real-time, and learning-driven task execution; continuously improving performance, adaptability, real-world feedback, and experiences.
- Moving away from individual robot-centric approaches towards multi-robot, multi-task, multi-party, and process-oriented operations, control, planning, and task execution; enhancing efficiency and scalability in complex environments.

These remarkable advancements pave the way for the creation of a new breed of robot AI - humanoid AI, and a new generation of robots - humanlike robotics. They seamlessly integrate human appearances, behaviors, and emotions into robotic systems, forming AI robots and AI humanoids. With the emergence of humanlike robots, discussions and speculation about their future trajectory have intensified various aspects below.

- 1) How will generative AI propel the evolution of robots?
- 2) In what ways will human-level AI redefine the capabilities of humanoid robots?
- 3) How can human cognitive and social features be seamlessly incorporated into physically humanlike robots to enable humane and humanlike humanoids?
- 4) What sets apart humanlooking robots from humanlevel robots?
- 5) What challenges lie ahead, and what does the future hold for humanlike robots?

The rapid evolution of human-level AI synergizing with humanoid robots, alongside the emergence of future generations of humanlike robots, is reshaping the landscape at an unprecedented pace. Humanoid and humanlike robots are emerging as unique and promising platforms and ecosystems for deploying diverse, multilevel, and multipurpose intelligent systems and applications, spanning from artificial narrow intelligence (ANI) to artificial general intelligence (AGI), and even potentially artificial superintelligence (ASI). Real-time, interactive humanoids epitomize a metasynthetic AI milestone and accomplishment [10]. However, the current emphasis on constructing humanoid robots with humanlike

Manuscript submitted to ACM

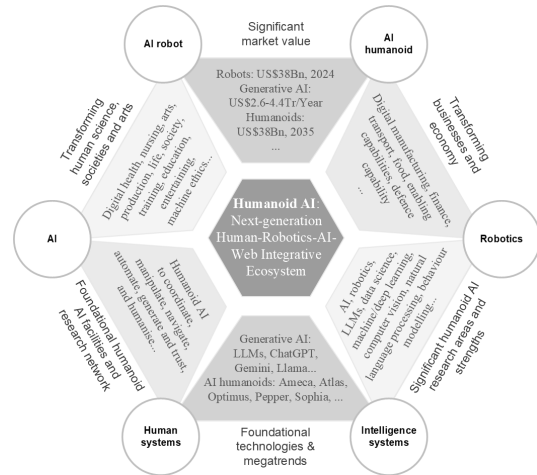


Fig. 1. Humanoid AI: A human-AI-robotics-web-integrative ecosystem.

appearances falls short of achieving AGI. To bridge this gap, the integration, simulation, and implementation of human-level AI and humanlike intelligence into humanoid robots will usher in new eras of humane, humanlike, and ultimately humanlevel robotics. Yet, the development of humanlike robots must not only replicate human physical, cognitive, and social attributes but also address ethical, legal, and social concerns and challenges.

In this article, we paint a comprehensive picture of AI-powered humanoids and humanoid AI by examining the synergies among large language models (LLMs), large multimodal models (LMMs), generative AI, and human-level AI with humanoid robotics. We begin by delineating the current state-of-the-art of humanoid robots, which possess humanlike appearances but are limited to artificial narrow intelligence (ANI) functions and intelligence levels. These robots have yet to attain humane, humanlike, or humanlevel capabilities, including artificial general intelligence (AGI) and artificial superintelligence (ASI). We then delve into the taxonomy, potential evolution and futures of transitioning humanlooking humanoid robots into humane, humanlike, and ultimately humanlevel entities, empowered by advancements in ANI, AGI, and ASI for robotics. Our discussion centers on the transformative journey toward developing humanlike to humanlevel humanoids. These advanced humanoids will embody omni-intelligent systems, seamlessly integrating human-level AI into robotics and extending the capabilities of humanoid robots with new paradigms of intelligence, ultimately progressing toward an integrated AGI and ASI robotic ecosystem. This evolution also signifies the emergence of a new generation of AI robots and humanoid AI. Furthermore, we explore various functional and nonfunctional requirements for humanlike to humanlevel robots, as well as techniques and future prospects for deep mind modeling of humanoids. We emphasize the importance of enabling and managing humanity in humanoid robots as they interact with other objects and humans in dynamic, real-time, and interactive environments. Our approach focuses not on mechanical, electronic, or biological aspects but rather treats humanoid robots as humanlike intelligent systems, with a special emphasis on soft AI-powered robotics and humanoid AI.

2 EVOLUTION AND CATEGORIZATION OF HUMANOID ROBOTS

There are about 30 types of humanoid robots available in the market and literature. Table 1 provides an overview and comparison of 29 typical AI-empowered humanoid robots endowed with humanlike features and associated AI and technical functions. The evolution of humanoid robotics has seen the paradigm shift from humanlooking humanoids to humane, humanlike and humanlevel humanoids. The family of humanoid robotics, in particular, AI humanoids, or humanoid AI, is the synergy and metasyntesis of robotics and AI and of human systems and intelligence systems.

2.1 Humanoid Evolution: From Humanlooking to Humane and Humanlike Paradigm Shift

Humanoid robots possess human structures with human appearance through assembling human-size and humanlooking body parts, which undertake human senses, human behaviors, human functions, human humanity, and human intelligence. The evolving landscape and spectrum of humanoid robotic development can be summarized in Fig. 2. The evolution of humanlooking humanoids to humane, humanlike and humanlevel humanoids can be categorized into six aspects and stages, i.e., from humanlooking structures to humanlike senses, humanlike behaviors, humanlike functions, humanlike humanity, and humanlevel intelligence.

Evolution Spectrum of Humanoid Robots. The first generation of humanoids focuses on replicating *human structures* and makes robot appearance with *humanlike structures*. Typically, a full-body humanlooking humanoid robot is built with human body structures such as human head, neck, chest, torso, arms, hands, legs and foot. A half-body humanoid robot may only have the upper part of humanlooking bodies, such as head, neck, shoulder and chest. Existing

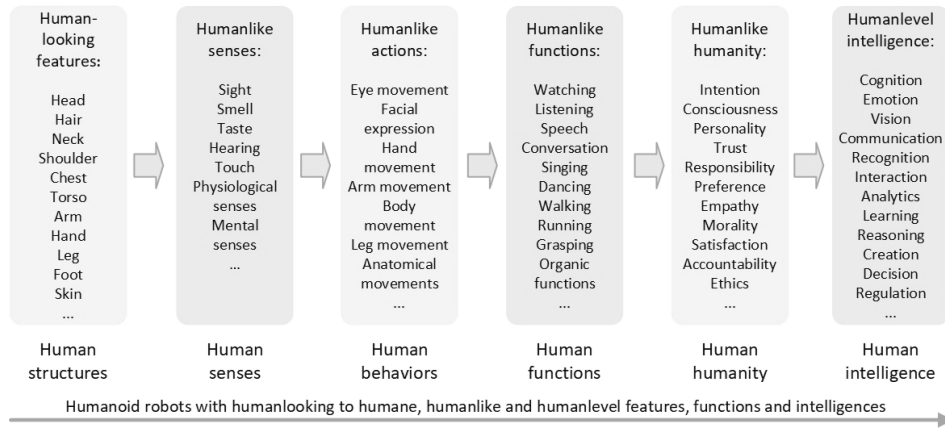


Fig. 2. Evolution landscape of humanlooking to humane, humanlike and humanlevel humanoid robots.

humanoid robotics incorporates either non-expressive robot heads into robots with actuated stereo vision, neck or hearing or expressive android heads with facial traits, artificial skins, and hair etc.

Second, humanoids aim to replicate *human senses* and imitate *humanlike senses* in humanoids. Accordingly, a humanoid is built with world perception and sense models, including a visual system for sight, an auditory system for audition, a gustatory system for taste, an olfactory system for smell, an external surface for feel, and cognitive components such as a brain with a vestibular system for balance, orientation, position, depth, and navigation.

Third, *human behaviors* are mounted to humanoid body parts to undertake or generate *humanlike actions*. Humanoids are then with human behaviors, such as eye movement, facial expression, and body postures with different hand or leg movements.

Further, *humanlike functions* are implemented into humanoid structures and senses to simulate *human functions*, such as speech, conversation, singing, dancing, walking, running, grasping, and watching, etc.

In addition, *humanlike humanity* has to be incorporated into humanoids to make them humane and develop the humanity of robots. The humanlike robot humanity may implement robot intention, consciousness, personality, trust, responsibility, preference, empathy, morality, satisfaction, accountability, and ethics.

The ultimate goal of humanoid robots are to implement *human intelligence* into robots toward humanlevel humanoid intelligence. They include cognition, vision, emotion, communication, recognition, interaction, collaboration, analytics, reasoning, learning, decision-making, and regulation.

From Humanlooking to Humanlevel Robotics. Humanoid robots is experiencing a fast humanization progression and evolution from humanlooking robots with human body parts to humanlike and humanlevel robots with humanlike structures, senses, behaviors, functions, and intelligences¹. Humanized humanoid robots are equipped with humanlike features and functions, such as by adding representative human traits and functions, e.g., eye contact, humanlike voice, social interactions, and emotional interpretation. This makes humanoid robots not only to avoid the ‘uncanny valley’ dilemma [34] but also to function humanly or like human beings, leading to *humanlike robotics*.

A humanlike robot looks and acts like human beings by replicating *human structures*, in particular, head structures such as humanlooking eyes, nose, mouth, ears, skin and hair, and other body parts such as hands and feet. Advanced

¹https://en.wikipedia.org/wiki/Humanoid_robot

humanlike robots may also simulate *human behaviors*, such as facial emotions and body posture and actions. High-end humanlike robots pursue *humane functions*, such as conversation, speech, singing, dancing, walking, running, jumping, grasping, and human organs such as electronic and AI-empowered brain, hands and legs. Rarely but essentially, humane robots would have *subjective human features*, i.e., humanity, such as personality, trust, empathy and preferences, etc. In general, humanlike robots need to replicate and be empowered with weak to strong *human intelligence*, such as cognition, vision, communication, hearing, interaction, emotion, and decision-making.

The above robot and humanoid progression pathway indicates a significant thinking and paradigm shift in defining, designing, manufacturing, operating and managing humanoid robots, promoting humanlooking to humane, humanlike and humanlevel robotics and robot AI.

Humanoid and Humanlike Robot Applications. Humanoid robots gain increasingly widespread, complex and commercial applications. Depending on their application situations and environments, humanoid robots can be customized and professionalized into service humanoids, convention humanoids, caregiving humanoids, advising humanoids, entertaining humanoids, home humanoids, factory humanoids, online humanoids, arts humanoids, military humanoids, social humanoids, and other task-specific humanoids.

Accordingly, we list a few typical applications of humanoids [25]. Examples include home, personal and elderly assistance and caregiving; entertainment such as for game design and competition; social services, demonstration and exhibition of robotics and AI advances and other specific functions; education such as for primary education, inquiries, question/answering, and online learning; marketing, hospitality and customer services such as for general inquiries, reception, chatbot, campaign, and activity organization and management; digital arts such as for creating digital songs, dances and artworks; health and medical research and services such as prosthesis, orthosis, personalized healthcare aids, eldercare, companionship and nursing support, and supporting people with disabilities; businesses, manufacturing and industries such as mining, assembly, sorting, painting, material handling, pick-and-place, and delivery, and operating specially designed equipment and vehicles, such as stuntronics and animatronic devices; cognition and neuroscience research such as human brain modeling and cognitive functions on neurobots; military, astronomy and space exploration; disaster management and performing hazardous jobs such as nuclear and earthquake rescue; and large-scale social activities such as for theme parks and exhibitions.





2.2 Taxonomy of AI Humanoid Robots

The taxonomy of AI humanoid robots reflects the interaction and synergy between robotics and AI and between intelligence systems and human systems. Fig. 3 shows a taxonomy of AI robots and AI humanoid robots.

First, on the high and technical level, AI robotics and AI humanoid robotics are the synergy of robotics and AI. Robotics comprises a family of robotic functions, tasks, and techniques, including robot interaction, robot collaboration, robot planning, robot navigation, robot manipulation, robot perception, robot learning, robot control, robot cognition, robot emotion, and robot ethics. AI functions, tasks and techniques make intelligent robotics, including by cognitive science, computer vision, NLP, speech recognition, signal recognition, pattern recognition, machine and deep learning, data analytics, computational intelligence, knowledge representation, human-computer interaction, AI and machine ethics and humanity.

Second, on the methodological level, AI humanoids integrate human systems and intelligence systems. Human systems consist of important traits and merits including human structures, human senses, human behaviors, human functions, human humanity, and human intelligence, which make humanoids humanlike and humanized. Intelligence systems implement various paradigms of intelligences, i.e., X-intelligences [12], from traditional intelligence paradigms

Table 1. Human-oriented features and issues of AI and LLMs-enabled humanoid robots.

Name & Company	Picture	Humanlike structures	Humanlike behaviors & functions	AI & technical functions	Applications	Shortage
Ameca ² , Engineered Arts, UK		Human look, face, head, neck, torso, arms, hands, fingers, legs, and feet, etc	Facial expressions, emotions, posture, eye movement, conversation, hand movement, interaction, collaboration, automation, etc	Speech recognition, image recognition, NLP, multimodal modeling, online connection to LLMs and cloud services, emotion detection, human-robot interaction and collaboration; Tritium, Python, cloud-based	R&D, human-robot conversation, exhibition, reception, product testing, museum, entertainment, training, and education	Limited practical or industrial applications, weak planning and navigation
RoboThespian ⁴ , Engineered Arts, UK		Head, face, neck, torso, arms, hands, fingers, legs, and feet, etc	Facial expression, gesture, singing, conversation, interaction, automation	SHORE object recognition ⁵ , recognition of speech, face and image, emotion detection, human-robot interaction, etc.; Tritium, Python, cloud-based	Delivery, demonstration, STEM education, entertainment, research	Limited mobility, physical interaction capabilities, and practical or industrial applications
Apollo ⁶ , Apptronik, US		Face, head, neck, torso, arms, hands, fingers, legs, and feet	Mobility, interaction, delivery, safety, general purpose	Path planning, navigation, object detection, general-purpose human-robot collaboration, autonomous	Commercial humanoid, retail, manufacturing, logistics, construction, oil and gas, electronics production, retail, home delivery, elder care, etc	No facial expressions and NLP-based conversation and communication, no advanced AI capabilities
ARMAR-III ⁷ , Karlsruhe Institute of Technology, Japan		Head, eyes/cameras, torso, arms, fingers	Conversation, interaction, use tools designed for humans, such as power drill and hammer ⁸	Learning, action and intention recognition, inference, human-robot collaboration, sensor-actuator-controller automation; ArmarX	Using tools	No face, legs and feet, no advanced AI capabilities

Continued on next page

² <https://www.engineeredarts.co.uk/robot/ameca/>

³ Images of the robots were identified online.

⁴ <https://wiki.engineeredarts.co.uk/RoboThespian>





⁵ <https://wiki.engineeredarts.co.uk/Sensors#SHORE>

⁶ <https://apptronik.com/apollo>

⁷ <https://www.sts.kit.edu/downloads/data-sheet-amar-6.pdf>

⁸ <https://h2t.iar.kit.edu/english/397.php>

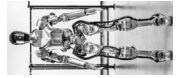



Table 1 continued from previous page

Name & Company	Picture	Humanlike features	Humanlike functions	AI & technical functions	Applications	Shortage
Atlas ⁹ , Boston Dynamics, US		Non-humanlike head, torso, arms, legs, and feet	Whole-body mobility, bimanual manipulation, navigation, dexterity, walking, dancing ¹⁰ , running, jumping etc. ¹¹	Athletic intelligence, model predictive control, real-time perception, dynamic manipulation, agility, adaptability, and reliability ¹²	R&D robot, dynamic robot, warehouse operations, athletics, delivery, emergency services in search and rescue	No humanlike head, face, and emotion, weak advanced AI capabilities
Beoami ¹³ , Beyond Imagination, US		Head, face, neck, torso, arms, hands, fingers, wheel	Navigation, gesture, performing physical skilled and unskilled labor, safety, planning, scheduling, multi-tasking, quality control	Learning to perform tasks, object recognition, sensing, haptic communication, learning by observation, NLP, motion planning	Manufacturing, construction, hospitality, nursing, healthcare, agriculture, dangerous work, manufacturing, logistics, retail ¹⁴	No bipedal, legs, and no facial expressions
Digit ¹⁵ , Agility Robotics, US		Non-humanlike head, neck, torso, arms, legs, feet	Navigation, walking, movement operations, interaction, safety	Human-robot collaboration, speech and object recognition, perception, manipulation, compatible with third-party software	warehousing, manufacturing, logistics	No human face and emotions, weak advanced AI capabilities
EVE ¹⁶ , IX, Europe		Non-humanlike head, torso, arms, hands	Navigation in unstructured space, performing tasks like opening the door, packing goods, patrolling, interaction, safety	Perception, path planning, navigation, object detection, learning by observation, human-robot collaboration, autonomous, IXOS	Retail, manufacturing	No bipedal, no humanlike head and emotions, weak advanced AI capabilities

Continued on next page

⁹<https://bostondynamics.com/atlas/>
¹⁰<https://www.youtube.com/watch?v=fn3KWN1kuAw&t=2s>
¹¹<https://bostondynamics.com/blog/flipping-the-script-with-atlas/>
¹²<https://bostondynamics.com/blog/picking-up-momentum/>
¹³<https://www.beoami.ai/>
¹⁴<https://www.beoami.ai/>
¹⁵<https://agilityrobotics.com/products/digit>
¹⁶<https://www.ix.tech/androids/eye>





Table 1 continued from previous page

Name & Company	Picture	Humanlike features	Humanlike functions	AI & technical functions	Applications	Shortage
Figure 01 ¹⁷ , Figure AI, OpenAI, US		Non-humanlike head, neck, torso, arms, hands, fingers, legs, feet	Walking, conversation, dexterous manipulation, interaction, navigation, payload	Speech-to-speech reasoning, path planning, navigation, object detection and recognition, decision explanation, learning, autonomous	General-purpose and commercial humanoid, manufacturing, logistics, warehousing, and retail	No facial expressions
Forerunner KI ¹⁸ , Kepler, China		Non-humanlike head, neck, torso, arms, hands, fingers, legs, feet	Walking, navigation, object detection, heavy-duty tasks, high-risk operations, autonomous	Visual recognition, navigation, online development, open collaboration, Kepler OS	General-purpose humanoid, research, education, inspection, production, warehousing, logistics, outdoor tasks	No facial expressions, limited AI capabilities
Geminoid ¹⁹ , Osaka University, Japan		Anthropomorphic head, hair, face, skin, torso, arms, fingers, legs, feet	Facial expression, talking, eye movement, head and neck movements, interaction, some psychological aspects	Teleoperated android, human-robot interaction	Simulating humanlike robot appearance, perception and movement of a real person, social robotics studies, telepresence applications	Restricted mobility, weak AI capabilities
GR-1 ²⁰ , Fourier Intelligence, China		Non-humanlike head, neck, torso, arms, hands, fingers, legs, feet	Walking, obstacle evasion, slope navigation, disturbance response, simple manipulation, collaboration	Path planning, navigation, object detection, human-robot collaboration, autonomous	General-purpose humanoid, industry, healthcare, rehabilitation, home, and research	No facial expressions, limited advanced AI capabilities

Continued on next page

¹⁷<https://www.figure.ai/>¹⁸<https://www.gotokepler.com/home>¹⁹<http://www.geminoid.jp/projects/kibans/Data/Geminoid2.pdf>²⁰<https://robots.fourierintelligence.com/>

Table 1 continued from previous page

Name & Company	Picture	Humanlike features	Humanlike functions	AI & technical functions	Applications	Shortage
Hi ²¹ , Unitree, China		Partial head, neck, torso, arms, legs, feet	Walking, running, jumping, climbing, manipulation, balancing, delivery	360-degree depth perception, path planning, navigation, object detection, advanced powertrain, autonomous	General-purpose humanoid	No head and hands, limited AI capabilities
Jia Jia ²² , University of Science and Technology in China, China		Anthropomorphic head, hair, skin, face, neck, torso, arms, fingers, legs, feet	Facial expressions, talking, interaction	Cognitive modeling, semantic understanding, automated reasoning and planning, knowledge acquisition, human-robot interaction, kinematics and cloud robotics	Public engagement	Limited mobility, predefined functions, limited facial expressions and movement, limited advanced AI capabilities
Junko Chihira ²³ , Toshiba, Japan		Anthropomorphic head, hair, face, skin, neck, torso, arms, fingers, legs, feet	Talking, gesture, eye and head movement	Scripted and pre-set speech and movement	Reception, customer service	Limited mobility, no AI capabilities, no emotion
Kime ²⁴ , Maceo Robotics, Spain		Non-humanlike head, torso, arms	Arm movement, object grasping, manipulation	Object detection and manipulation	Serve food and beverage products untrumped	No legs and feet, no emotion, limited AI capabilities
Nadine ²⁵ , Nanyang Technological University, Singapore		Anthropomorphic head, face, hair, skin, neck, torso, arms, fingers, legs, feet	Perception, talking, greeting, facial expressions, eye contact, upper body movements, conversation, memorization	Recognizing objects, face, emotion and gesture, communication	Receptionist, customer service, and personal coach	Limited physical mobility, limited AI capabilities

Continued on next page

²¹<https://www.unitree.com/en/h1/>





²²<https://www.nature.com/articles/d42473-018-00091-3>

²³<https://www.global.toshiba/ww/news/corporate/2015/10/tp1901.html>

²⁴<https://www.maceorobotics.com/en/robot-camarero-kime>

²⁵https://en.wikipedia.org/wiki/Nadine_Social_Robot





Table 1 continued from previous page

Name & Company	Picture	Humanlike features	Humanlike functions	AI & technical functions	Applications	Shortage
NAO ²⁶ , Aldebaran, Europe		Non-humanlike head, neck, torso, arms, legs, feet	Movement, perception, dialogue, gesture, interaction, navigation	Object and speech recognition; NAOqi OS, C++, Python, JavaScript and ROS interface	Research, education, healthcare, entertainment, RoboCup	No emotion, limited physical strength, limited AI capabilities
OceanOneK ²⁷ , Stanford Robotics Lab, US		Non-humanlike head, eyes, neck, torso, arms, hands	Diving, perception, stereo vision, touch/haptic communication, feeling, navigation, interaction	Object detection, path planning, navigation, human-robot haptic interaction, whole-body control, autonomous	Underwater/deep-sea exploration, deep dive, hmanual manipulation tasks	Not bipedal, limited senses, no facial expressions, predefined automation
Optimus ²⁸ , Tesla, US		Non-humanlike head, neck, torso, arms, fingers, legs, feet	Walking, dancing, hand movement, dextr manipulation of objects	Path planning, navigation, object detection, fine manipulation, full self-driving, full body control, autonomous	Demonstration, manufacturing, logistics	No facial expressions, limited advanced AI capabilities
Pepper ²⁹ , SoftBank Robotics, Japan		Non-humanlike head, neck, torso, arms, hands, fingers, legs and wheel base	Conversation, gesture, friendly and engaging interaction, autonomous navigation, eye and body language	Real-time data collection, speech, facial and object recognition, ChatGPT connection; NAOqi OS, C++, Python, JavaScript and ROS interface	Retail, healthcare, hospitality, education, life science	No emotion, not bipedal, limited advanced AI capabilities

Continued on next page

²⁶<https://www.aldebaran.com/en/nao>²⁷<https://khatib.stanford.edu/ocean-one-k.html>²⁸<https://electrek.co/2023/12/12/tesla-unveils-optimus-gen-2-next-generation-humanoid-robot/>, <https://newatlas.com/robotics/tesla-optimus-gen-2-robot/>²⁹<https://us.softbankrobotics.com/pepper>, <https://unitedrobotics.group/en/robots/pepper>

Table 1 continued from previous page

Name & Company	Picture	Humanlike features	Humanlike functions	AI & technical functions	Applications	Shortage
Phoenix ³⁰ , Sanctuary AI, Canada		Non-humanlike head, neck, torso, arms, hands, fingers, legs, feet	Mobility, walking, fine manipulation and dexterity, teleoperation, telepresence, touch/haptic communication, grasping, autonomous	General-purpose robot, path planning, navigation, object detection, LLMs for language-to-action	Manufacturing, retail	No facial expressions
Promobot ³¹ , Promobot, Europe		Non-humanlike head, neck, torso, arms, hands, movement platform	Conversation, body movement, interaction, navigation	Face and speech recognition, path planning, navigation, object detection, interaction, autonomous; Ubuntu, compatible with third-party software	Service robot, consultant, promoter, concierge, tour guide, assistant	Not bipedal, no facial expressions, limited advanced AI capabilities
RT2 ³² , Google, US		Non-humanlike structure	Perception, learning, reasoning, control	Human recognition, LLMs, vision-language-vision learning from web and robot, learning to control, multi-task demonstration, reasoning, symbol understanding	Research, grasping, retail, open-world manipulation	No human features
Sophiar ³³ , Hanson Robotics, China		Head, anthropomorphic face, skin, neck, torso, arms, hands, fingers, legs, feet	Conversation, gesture, facial expression, interaction, walking, emotion	Human-robot interaction, object detection, symbolic AI, perception, NLP, expert systems, adaptive motor control, generation	Research, demonstration, medicine education, service, entertainment, human-robot interaction	Scripted interaction, limited physical mobility

Continued on next page





³⁰<https://sanctuary.ai/resources/news/sanctuary-ai-unveils-phoenix-a-humanoid-general-purpose-robot-designed-for-work/>

³¹<https://promo-bot.ai/robots/promobot-v4/>

³²<https://deepmind.google/discover/blog/rt-2-new-model-translates-vision-and-language-into-action/>, although RT2 is not a humanoid, its LLM-enabled general-purpose design and functions are easily transferable to humanoids

³³<https://www.hansonrobotics.com/sophiar/>

Table 1 continued from previous page

Name & Company	Picture	Humanlike features	Humanlike functions	AI & technical functions	Applications	Shortage
Surena IV ³⁴ , University of Tehran, Iran		Non-humanlike head, neck, torso, arms, hands, legs, feet	Walking, speaking, object manipulation, interaction	Speech, face, object and activity detection, state monitoring, speech generation, motion planning, action imitation, robot interaction, ROS, C++, Python, Lisp, Java, JavaScript and Ruby	Academic research	No facial expressions, weak advanced AI capabilities
THRC ³⁵ , Toyota, Japan		Non-humanlike head, neck, torso, arms, hands, legs, feet	Natural walking, mobility, wearable control, foot and hand movement, joint control, motion planning, whole-body coordination	Flexible control, planning, remote maneuvering system mirroring user movements to the robot	Home, medical surgery, nursing, avatar, construction, caregiving	No humanlike head and emotions, preprogrammed
Valkyrie ³⁶ , NASA, General Motors, US		Non-humanlike head, neck, torso, arms, legs, feet	Space walking, movement, interaction, recognition, manipulation, gripping and building things	Perception, path planning, navigation, object detection, manipulation, remote operations and control	Dexterous robot, manned missions, space exploration, autonomous	No facial expressions, limited AI capabilities
Walker X ³⁷ , UBYTECH Robotics, China		Non-humanlike head, neck, torso, arms, hands, fingers, legs, feet	Perception, walking, navigation, grasping, manipulation, hand-eye coordination, self-balancing	Human-robot interaction, path planning, navigation, object detection, force/position control, multimodal interaction, programmable autonomy; Ubuntu, ROSA	Academic research, training, service	No face and emotion, programmed

³⁴<https://surenahumanoid.com/surenaIV.html>, [https://en.wikipedia.org/wiki/Surena_\(robot\)#](https://en.wikipedia.org/wiki/Surena_(robot)#)

³⁵<https://global.toyota/en/detail/19666346>

³⁶<https://www.nasa.gov/podcasts/houston-we-have-a-podcast/valkyrie/>

³⁷<https://www.robotlab.com/higher-ed-robots/store/walker-humanoid-service-robot-for-research>

such as symbolic intelligence, connectionist intelligence, natural intelligence, social intelligence, domain intelligence, computational intelligence, and networking intelligence, to emerging paradigms including data intelligence, algorithmic intelligence, learning intelligence, generative intelligence, behavioral intelligence, emotion intelligence, and cognitive intelligence.

As a result, AI humanoids replicate, synthesize and meta-synthesize [10] human systems and intelligence systems. This expands, upscales and transforms robotics toward a spectrum of humanoid functions. A family of AI humanoid functions emerge, such as humanoid interaction, humanoid collaboration, humanoid planning, humanoid navigation, humanoid manipulation, humanoid perception, humanoid learning, humanoid control, humanoid cognition, humanoid emotion, and humanoid humanity.

Consequently, the applications of AI humanoids generate and satisfy various purposes and produce diversified types of humanoids. Examples include social humanoids, walking humanoids, conversational humanoids, interactive humanoids, expressive humanoids, generative humanoids, manipulative humanoids, teleoperated humanoids, service humanoids, entertaining humanoids, cognitive humanoids, and imitative humanoids.

3 HUMANE AND HUMANLIKE HUMANOIDS: FUNCTIONAL AND NONFUNCTIONAL REQUIREMENTS

Here, we summarize the functional and nonfunctional requirements and challenges in humane and humanlike humanoids, as illustrated in Fig. 4.

3.1 Humanoid's Functional Requirements and Challenges

Functional requirements consist of main robotic functions, including humanoid cognition, perception, communication, interaction, analytics and learning, planning, adaptation, behaviors and dynamics, anthropomorphic and mimetic features, mechanical, electrical, biological and social designs, manipulation and control.

3.1.1 Humanoid Cognition. *Humanoid cognition* simulates and implements human cognitive thinking, traits, and capabilities. Accordingly, robotic cognition builds humanlike brain and mental models with functions and capabilities including reasoning, inference and decision-making mechanisms, and enabling robotic goal, intention and emotion. Robotic mental modeling [49] attributes the theory of mind with thoughts, desires and intentions etc. and formalizes them as shared mental representations between robots and between humans and robots to simulate the awareness, roles, responsibilities, knowledge structures, cognitive mechanisms and processes of human cognition, reasoning, teaming, interaction, communication and collaboration etc for human-like problem-solving and decision-making.

Humanoid intention and emotion imitate human facial expressions in robots and embodies and learns robotic intention and emotion. This represents a new stage of humanlike, humanlevel and humane robotics. *Robotic emotions* [39] may be presented by hand-coded, automated, and learning-based designs and methods. *Robotic affection* combines personality trait, attitude, emotion, mood and interpersonal stance [31]. The generation, recognition, learning, alignment and adjustment of automatic, real-time and complex robotic emotion, affection and intention guided by goals and tasks, in the wild, during performing activities, mixed with other robotic modalities, or during interaction and collaboration with humans represent some future directions.

3.1.2 Humanoid Perception. Humanoid perception applies humanlike senses to represent and understand information and environment. Humanoid robots are equipped with proprioceptive and exteroceptive sensors to perceive, understand and model robot's internal states and behaviors, target objects and their states and behaviors, interactions between robots and objects, and the environment. Robot perception [41] captures internal states of humanoid robots, including

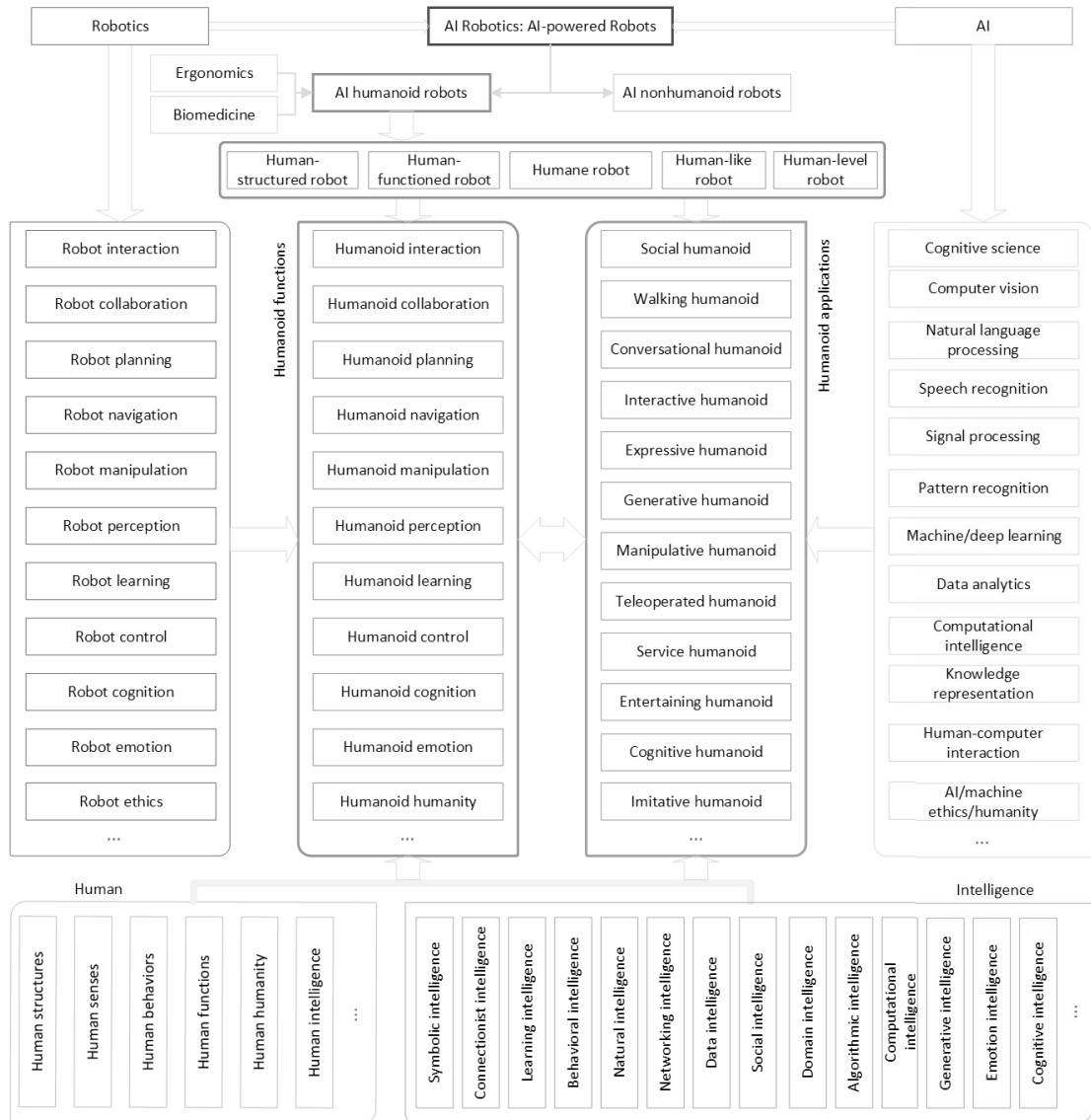


Fig. 3. The taxonomy of AI humanoid robots: Interactions and synergy between robotics, AI, intelligence and human systems.

information about their locomotion, position, velocity, depth, orientation, behaviors such as attention and gaze movement, and emotion. A humanoid may also identify the target objects and their layouts, states, behaviors, and activities, etc., e.g., human pose and facial features. In the cases where there are robot-object interactions and human-robot interactions, the perception will further identify the interaction modes, activities, and types, etc. The environment may include the context and surroundings of a robot and their objects, their materials and textiles, where robot perception identifies environmental states, and conditions (e.g., lightening conditions), and layout, etc.

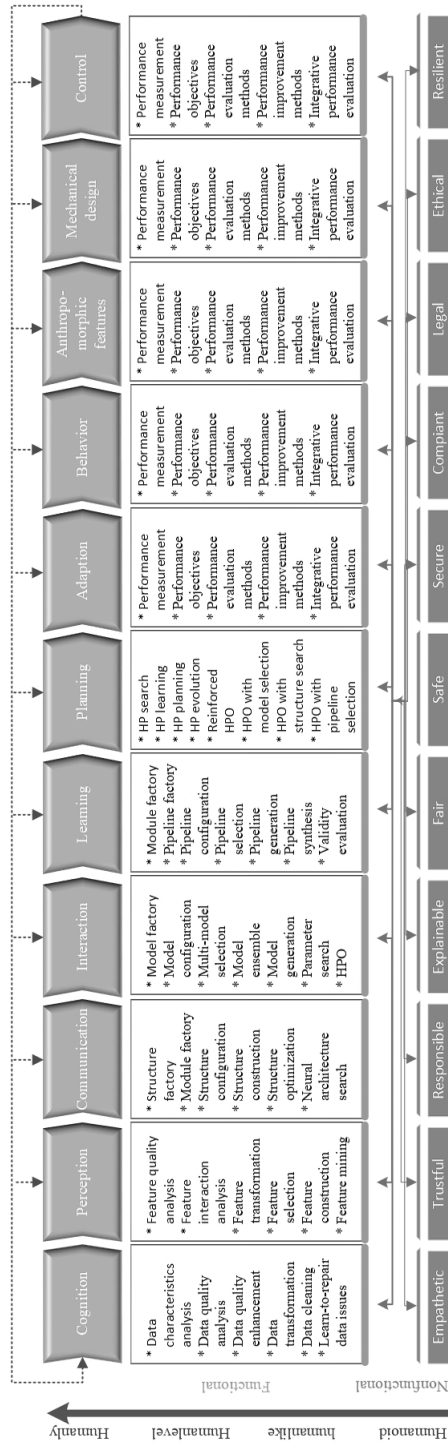


Fig. 4. Functional and nonfunctional objectives, tasks and intelligences of humanoid, humanlike, humanelevel and humane humanoid Robots.

Humanoid vision [40] captures and enables robotic gaze, gesture, pose, action, motion, expression, movement, collaboration, learning, and communication; and supports specific tasks such as object handover. To enable perception, sensors such as actuated cameras, force, microphone, inertial measurement units, encoders, sensitive resistors, sonar, rotary, tactile, torque, laser scanner, lidar, joint and inclinometer may be placed in a humanoid's head, neck, chest, torso, hand, etc body parts. Sensors or multi-sensor fusion may capture internal and external states, signals or videos relating to vision, audio, olfactory, tactile, shape, curve, and posture of human body parts; humanly features such as emotion and facial expression; physiological signals of human body, such as heart rate, blood pressure, body temperature, brain activity, and muscle activation; and environmental range and layout, or contextual representation and memory, etc.

Perception may serve high-level functions such as tracking, detection, recognition, identification, classification, control, localization, navigation, planning, mapping, manipulation, and grasping. Perception supports specific tasks relating to human-robot interactions, object manipulation, gait planning, proprioceptive state estimation, action prediction, and emotion recognition. In performing perception, humanoids recognize internal states such as stability, dynamics, safety, efficiency, and experience; and external aspects such as identifying, recognizing and locating external stimuli, layout and relationships with objects; and interactive states such as robustness and adaptability to dynamic and unforeseen environmental changes.

3.1.3 Humanoid Communication. Humanoid communication imparts, transports or exchanges information within a humanoid, between humanoids, between a humanoid and its environment, or between humanoids and other objects like humans. Accordingly, communication involves participants, such as other humanoid robots, humans, or objects; media, such as in imagery, video-based, acoustic, textual, or emotional forms; content such as speech, navigation and actions; and context such as topic area and background. Communications may involve a single party, multiple parties, or a team, where multi-party humanoid or human-humanoid communication is essential. Communications may be unidirectional or directional e.g. human-to-humanoid communication and humanoid-to-human communication, or nondirectional such as in peer-to-peer or group communications.

Communications take place within humanoid robots in various ways, e.g., through coordination between head and hands or between eye and hands by passing and processing sensory information, and signal transformation such as from perception to conversation; between humanoid robots or between humans and robots, e.g., through cooperation or negotiation between robots for collective tasks; and between robots and their environment, e.g., through tracking and navigation; etc.

Communications may take forms of linguistic communications, such as by speech, message-passing, dialogue, query-answering, or retrieval; and non-linguistic communications, such by eye contact, emotional response, and gestural and graphical movement. Accordingly, communications may be expressive such as by emotional expressions and exchange or non-expressive such as through sensory information consumption and sharing. In some cases, communications may take the form of translation and transformation, such as text-to-speech translation, speech-to-text translation, image-to-text transformation, text-to-image transformation, vision-to-language transformation, and language-to-vision transformation, etc. These translation and transformation tasks undertake cross-modality information transport.

3.1.4 Humanoid Interaction. Humanoid interaction connects, associates or relates humanoids to each other or to other objects like humans. Interactions take place between humanoid robots, humans and environments and within robot communities, e.g., forming human-robot interactions [23] and specifically human-humanoid-environment interactions for humanoid robots. Interactions may be directional, physical, cognitive, perceptual, behavioral, responsive, social,

contextual, or environmental. Interactions between humanoid robots and humans may take specific forms of coordination, cooperation, negotiation, collaboration, or even conflicting or competition; and general forms of association, connection, recognition, identification, detection or tracking. Interactions are associated with, during or for tasks, such as grasping, assembly, communications or collaboration. Robot collaboration or human-robot collaboration [5, 43] fulfills a common task based on defined team formation, goal, plan, actions, strategies, constraints, and conflict resolution, etc. Humanoid-human collaboration is built on hand-crafted knowledge base and protocols, large data pretraining and real-time finetuning and learning, or even mixed reality and metaverse.

Social humanoids engage social interactions [27] in a robot society or a human-robot coexisting community. Social humanoids follow social and cultural norms and express emotions and sentiments during the engagement and in fulfilling social skills and solving social problems. Accordingly, robot interaction may be associated with intention, emotion and affection. Affective and emotional robot interaction [35] may be collaborative for collaboration, assistive for assistance, mimicry for imitation, coordination or cooperation, and general or multi-purpose for different communications.

3.1.5 Humanoid Analytics and Learning. In general, AI-powered humanoids are capable of analytics and learning from data, environment, and other objects. Humanoid robots are increasingly learning driven and enabled by learning systems, in particular the recent advances in reinforcement learning, learning by demonstration, learning from imitation, robot analytics, and recently generative AI and LLMs and LMMs for robots. First, *robot reinforcement learning* optimizes goals, tasking performance or behaviors of robots, including by traditional and deep reinforcement methods with on-policy or off-policy settings. *Robot programming, or learning by demonstration* [1], such as learning from training, examples, simulation, teleoperation or shadowing, is a typical robot learning approach for unsupervised, semi-supervised, and few-shot supervised learning tasks. During demonstration, domain experts or end users can teach robots to conduct specific tasks, robots then distill knowledge or skills from these demonstrations without pre-programmed knowledge to execute new tasks. Another approach is *learning by animation*, such as from sensors on teacher or external observations. Learning from augmented reality or learning from metaverse may be new opportunities to design, develop, train and evaluate novel robot skills and applications at a cost-effective way through robot-metaverse integration.

Humanoid analytics identify, detect or predict patterns or exceptions of and insights into robot signals, behaviors, movements and trajectories, and their external, contextual and environmental factors and changes. Robot analytics may be conducted to support any functional and nonfunctional objectives and different paradigms of robotic intelligences, as shown in Fig. 4. Analytics may be on historical, current, future, new, cold-start and evolving data and behaviors.

Deep learning, LLMs and GAI for humanoids In recent years, deep learning advances in particular LLMs and GAI have revolutionized humanoid robots toward implementing humanlike and humanlevel robotic intelligence. They acquire and analyze any sources, modalities and formats of large scale and mixed inputs to robots through pretraining and then refinetuning. *Humanoid pretraining* may be performed on signals, videos from sensors of body parts, the activities and behaviors of robot body or parts; interactions, communications and coordination between robots and between robots and their environments; and transitional processes such as from vision to language and translations such as from speech to image. *Robot learning* has experienced fast progression from vision-language modeling (VLM) to vision-language-action (VLA) modeling. In Section 4, we will further discuss robot analytics and directions for LLMs, GAI and deep learning into robotics, and specifically on opportunities including omnimodal perception-to-behavior modeling for mindful, cognitive and actionable humanoids in Section 4.2.

3.1.6 Humanoid Planning and Search. Humanoid planning decides what, how, when or where a humanoid behaves. *Robot planning* widely explores goals of navigating robots, moving body parts, taking actions or performing tasks,

etc., which has been customized into areas including path planning, motion planning [50], and task planning, etc. Robot planning may be undertaken in known or unknown clutter environments with known or unknown obstacles. Robot aims to optimize target motion, movement, path, or task-specific objectives, and to minimize concerning issues, such as collisions, time or energy consumption. Classic robot planning relies on robot search such as heuristic search, stochastic search to find, seek or retrieve interested information or target. Further, humanoid applies techniques including probabilistic roadmap, evolutionary computing, geometric modeling, reinforcement learning, and machine learning methods to planning.

With humanoid robots enabled by LLMs and LMMs, *learn-to-plan*, or *planning by learning*, becomes a new fashion of planning in complex contexts. For example, a robot may manipulate open world unseen objects [44] enabled by a pretrained VLM, which extracts object category-specific features from image and text inputs. Further, the robot can undertake actions or complete instruction with policies informed or constrained by the categorized object, the image, and the textual instruction. For example, SayCan [24] can *plan-to-act* based on the input language. A recent VLM-based learn-to-plan advance is the Google RT-2, which builds on a VLM backbone, plans from both image and text commands, and conducts visually grounded planning. The *chain-of-thought* reasoning can enable learning long-horizon planning and low-level skills. It is claimed that RT-2 can achieve highly-improved robotic policies by instantiating VLAs based on PaLM-E and PaLI-X. RT-2 can significantly improve generalisation performance and emergent capabilities using the power of web-scale *vision-language pre-training* (VLP).

3.1.7 Humanoid Adaptation. Humanoid adaptation adjusts a robot to fit its goals, conditions, environments and their dynamics. Humanoid robots are often challenged by live dynamics, states or environments and changes of robotic functions and tasks, which are particularly common for humanoids with real-time automated capabilities but challenge their design. Accordingly, real-time humanoids should have the adaptability and provide corresponding support to plan change, action adjustment, strategy tuning, and performance measures. This requires adaptive designs and capabilities, such as mechanical design with adaptive sensormotors, adaptive control techniques, online and active perception and recognition, evolving human-robot interaction, evolving learning, dynamic finetuning, and proactive risk and compliance.

Humans are adaptive and smart enough to make adjustment per dynamics or changes of terms and conditions. How to make humanlike robots to act, reason and interact like a human being in real-world scenarios posits humanoids toward humanlevel and humane but poses a fundamental challenge to humanoid robotics. Typically, shared *mental models* are enforced for all robots, which cannot capture novel or unseen scenarios. Then, predefined flexibility may be incorporated into design, such as bipedal locomotion and dexterous manipulation for evolving environments, which enables adaptation to predictable dynamics and changes. For novel and uncertain dynamics, *learning to adapt*, or *adaptation by learning*, represents a new approach, particularly suitable and essential for creating humanlevel robots. Humanoid adaptation by learning may be enabled by predictive architectures, data-driven adaptive learning capabilities, neuro-inspired control, and evolutionary learning. In complex or even extreme real-world scenarios, more sophisticated adaptive designs and human-supervised adaptation [55] may be essential to understand humanoid intention, learn their behaviors, and sense their emotions over time sequentially, and predict their next-moment actions.

In a robot society and a human-robot integrative society, morphological computation and evolutionary learning may be useful for adaptive evolutionary humanoids. Adaptive evolutionary humanoids may amalgamate biomimetic design, morphological computation and evolutionary learning, to enable adaptive mutation, crossover, selection and optimization of actions, behaviors, and states of humanoids iteratively and sequentially.

3.1.8 Humanoid Behaviors and Dynamics. Human dynamics and human intelligence dynamics are comprehensive and evolving, embodied through human cognitive dynamics, behavior dynamics, and system dynamics, etc. Here, we discuss humanoid behaviors and dynamics.

Humanoid behaviors refer to how humanoids behave, act, and react, etc. *Robot behaviors* take various forms for different purposes under specific scenarios, such as robot operations of actions, gesture, movement and navigation, interactions with other robots or humans [4], and teleoperations in a remote environment. Humanoid robot gestures [14] comprise human gesture-simulated appearances, behaviors and states, such as facial expression, eye gaze, communicative pointing and signs, and manipulative motions, etc. Humanoid robot movement further embodies actions such as eye movement, hand movement, and leg movement. *Robot navigation* [33] takes actions and follows trajectories per goals and intention. Navigation may be passive and guided by visual, vocal or other types of perception, a map, demonstrations, simulation or pretraining. In other cases, robots navigate through active and mapless exploration or learning. Robot gesture, movement, navigation and action generate and drive robot dynamics [46], which links to kinematics, contact mechanics, and centroidal dynamics, e.g., enabling robot mobility by bipedal or wheeled devices.

Humanoid teleoperations The telepresence and teleexistence of humanoid robots in remote environments, such as for telehealth and telenursing, are associated with robot teleoperations [18]. This involves local behaviors and telebehaviors, such as telocalization, teleperception, retargeting, mapping, planning and control. By integrating with metaverse etc simulation platforms, humanoid teleoperations may be further conducted in a digital twin, where initial physical scenes (such as medical test) could be captured and analyzed in the physical world, then converted into the virtual world (e.g., input into 3D simulators) for further manipulation and analysis, optional actions (e.g., medical treatments) could be tested in the virtual world, those verified by professionals (e.g., a doctor) could then be deployed to humanoids for actions (e.g., guiding a patient's teletreatment).

Robots may also be equipped with mechanisms and systems to enable behavior cloning or imitation. Robots conduct their behaviors in the lifecycle of robot tasking, generating sequential and full-cycle state spaces with behavior attributes, processes, consequences and impacts [9]. Robot behavior analytics and computing is thus useful to understand, monitor, predict and manage behavior effects and improve robot objectives.

3.1.9 Humanoid's Anthropomorphic and Mimetic Features. Humanoid robots are embodied with anthropomorphic and humanly mimetic, morphological, zoomorphic or caricatured features, including artificial organs such as electronic skin (e-skin), bionic and sensory tactile skin, prosthesis, wearables and 3D surfaces; anthropomorphic activities such as eye movement, skin sense, tactile sense and hand gesture; cognitive, sentimental and emotional states and features, such as facial expressions. Anthropomorphic body parts are embedded with humanlike functions, senses and traits such as touch, pressure, force, contact, hardness and texture, or even humanly physiological features such as temperature, pain, pulse rate, blood pressure, and bodily fluids such as sweat and tears. *Human mimetic mechanisms* [2] and *anthropomorphism* [58] for humanoids imitate human body structure and functions, such as skeletal structures and functions, human body proportions, link length, mass balance, muscle arrangement, insertion points, and joint performance with joint range and output power. Implementing human features in robots also develops *humanoid aesthetics*. In addition, social humanoids may hold humanized features and traits, such as personality as embodied by extroversion, agreeableness, conscientiousness, neuroticism and openness [22].

3.1.10 Humanoid's Mechanical, Electrical, Biological and Social Designs. Robot design [42] concerns hardware and software development in mechanical, electronic, electrical, social and software manners. These include mechanical

structures, kinematics, processing units, sensors, electronic and electrical devices, power supply, social forms and activities, and software architectures and implementation.

Humanoid's mechanical design creates and enables humanoid robot body and parts, such as robot head, neck, torso, hand, and leg; robot operations and control, such as locomotion mechanisms, force and strength manipulation, legged or wheeled systems, and transduction mechanisms and functions; robot functions such as grasping and assembly; robot communications such as coordination; robot behaviors such as navigation and tracking; robot aesthetics such as enabling humane and humanlooking features; and robot control and assurance such as robot safety and slippage. Mechanical designs, mechanisms and functions are further integrated with sensory and electronic mechanisms and functions, such as resistive, capacitive, optical, piezoelectric and magnetic mechanisms and devices.

Humanoid's biological design [6] is bio-inspired and simulates biological and biomedical mechanisms, structures, materials and control into humanoids for more natural, humanlike and rational structures and functions and more intricate tasking and performance. The design of humanoid body and parts involves various biomedical, ergonomic and neural mechanisms, parameters, and settings, such as degrees of freedom, size, weight, strength, load, volume, mass, power supply, energy and constraints.

Humanoid social design enables social activities of robots in the robot team or society, such as interactions and collaborations with other robots; supports humanoid-human interactions, teaming and collaborations; enforces social robot norms, such as avoiding collision and attacking humans; and incorporates ethics into social humanoid robots, such as performing safety and privacy check before actioning and enforcing accountability.

3.1.11 Humanoid Manipulation and Control. Robot manipulation Humanoids may manipulate their body parts such as eye movements, facial expression, or postures, or external objects such as grasping objects. Robot manipulation serves various tasking purposes and involves diversified robot parts. Robot manipulation may be classified into typical types including system-centric manipulation, object-centric manipulation, action or state-centric manipulation, transition-centric manipulation, and process-centric manipulation. System-centric robot manipulation operates a robot part, e.g., manipulating the locomotion. Object-centric robot manipulation instructs, controls and manipulates robots toward specific objects under object background. Action-centric and state-centric robot manipulation guide, control and operate robots toward taking specific actions or approaching target states, e.g., rearranging an object or reorienting a robot hand. Transition-centric robot manipulation focuses on robot state or behavior transition from one to another, e.g., raising a hand and then positing it toward a target object. Process-centric robot manipulation involves multiple steps of action or state progression or change, e.g., from perceiving the environment to moving a hand and then grasping an object.

Robot control Humanoid control represents a strong type of manipulation of humanoids or external objects, where a humanoid influences, directs or manages its behaviors, actions or their processes. The objectives and tasks of robot control are varied [38]. A robot may control its parts such as head, eyes, emotion, gait, legs and hands. A robot may perform control to implement its functions such as trajectory, horizon, gesture, posture, collaboration, interaction, behavior position, and speed. Robot control can execute specific tasks such as bipedal locomotion, omnidirectional walking or satisfy nonfunctional requirements such as safety. Humanoid robots may be controlled by different mechanisms. Functionally, humanoid control relies on the corresponding mechanical and electrical humanoid designs to implement specific objectives and tasks, such as optimization methods, dynamic models, bionic methods, reinforcement learning, demonstration learning, imitation learning, and prediction. Controls are incorporated into humanoid body parts, such as by a central nervous system, a motor controller, stability and balance measures such as zero moment point, capture

point and centre of pressure, and data- and model-driven analytical and learning techniques. Control environment may be open, closed, semi-open or semi-closed in terms of interactions with environments. Control modes may be autonomous, semi-autonomous or hand-engineered in a static, dynamic or adaptive manner. Automated control may be driven by policies through real-time learning or rules predefined or hand-coded.

Learning-to-control and data-driven and behavior-based control represent new control modes and scenarios, where analytical and learning systems such as anomaly detection, predictive learning and generative AI are incorporated into robot engine for monitoring, detecting, predicting and preventing control failures, damages or accidents [45]. In AI-empowered humanoids, a typical technique of learning-to-control is deep reinforcement learning, which obtains optimal actions or state-action combinations, i.e., policies and strategies, to decide and govern humanoid behaviors and actions. Humanoid learning-to-control may also apply learning by demonstration, imitation, prediction, and recommendation etc techniques to obtain control capabilities.

Humanoid robots also involve new control problems or approaches, which may relate to humanized features, functions and tasks. For instance, facial expression control, eye and mouth movement control, human gesture and posture control, hand and leg movement control, and human-humanoid interaction can be guided by hand-engineered rules or policies or learning-driven findings. These require not only ergonomic design but also control, in particular, real-time, interactive and online control from multitask, multimodal and omnimodal VLA modeling and cognition-to-action translation from web-scale data, humanoid data, and external environments.

3.2 Humanoid's Nonfunctional Requirements and Challenges

With the fast growth and advancement of robotic techniques and functions, humanoid robots are expected safe and secure in their operations and action-taking; trustful and responsible for their actions; transparent and explainable in their outputs; empathetic and rational during teaming, collaboration and interaction; and compliant, legal and ethical for integrity and regulation.

Accordingly, the functions and quality of robots need to be ensured by various nonfunctional requirements, objectives and constraints. Nonfunctional objectives, requirements and criteria of humanoid robots cover many aspects, categorized into traditional requirements on the robotic quality of services, and new control problems relating to humanized requirements.

3.2.1 Humanoid's Quality of Services. Robotic quality of services (RQoS) involves aspects of robot maturity, safety, stability, balance, robustness, reliability, generalization, accuracy, safety, privacy, computing-efficiency, energy-efficiency, adaptability, versatility, agility, dexterity, transparency, explainability, and reproducibility. Each of them emphasizes specific quality and performance, which could be further quantified by QoS measures.

Humanoid capability maturity measures the functions, capabilities, states, level and strength etc of robots being mature, which is further quantified in terms of robot capability maturity assessment specifications and functional and nonfunctional criteria and measures. A checklist of robot capability maturity may be created to check aspects and concerns such as: Whether the functions are mature? How robot risk, safety, security and trust etc are self-assessable, self-regulated, and extra-assessable? What commonly trustful evaluation, regulation, compliance, risk and safety assessment protocols, specifications and authority are available?

Humanoid evaluation verifies and validates functional and nonfunctional requirements and performance. The evaluation criteria and measures are both objective and subjective, and both technical and nontechnical. The perspectives

of evaluation relate to robotic objectives, designs and performance. These may span aspects from state to behavior, from domestic to social, from individual to collective, from physical to virtual, and from implicit to explicit, etc.

Humanoid safety [28] is commonly concerned, which consists of Physical safety and psychological safety. *Physical safety* refers to no unintentional or unwanted contact between robots and humans, or comfortable physical contact with force below thresholds and without physical harm. *Psychological safety* refers to indirect psychological harm, discomfort or stress such as human trust, unexpected intention, and negative emotions. Robot safety can be ensured by careful programming, ensuring safety criteria and constraints such as biomechanical limits, injury prevention, and control strategies, and safety assurance mechanisms such as detecting, predicting, preventing and intervening with force, collision, falling, crash, risky and abnormal activities. In open settings, safety monitoring, detection, prediction, prevention and intervention need to be enforced for novel, cold-start, evolving and changing scenarios and accidents.

The *transparency*, or opacity, of robotic designs, behaviors and outcomes affects human confidence in robots and their applications and services. Two mechanisms to ensure humanoid transparency are explainability and reproducibility. *Explainability* provides mechanisms, tools and interfaces to understand and interpret robot designs, working mechanisms, behaviors, outputs and consequences, etc. *Reproducibility* ensures the repeatability, consistency and integrity of robot designs and outputs under the same circumstances. Explainability and reproducibility become increasingly essential for complex robot designs and tasking. Transparency, accountability, auditability, explainability and reproducibility are essential in implementing robot automation, on-the-fly and in-the-wild design, operation and decision-making, and open settings.

3.2.2 Nonfunctional Humanized Requirements. Nonfunctional humanized requirements articulate subjective human traits such as robot personality, and societal, moral and legal considerations and requirements such as robot responsibility, trust and ethics. Nonfunctional humanized requirements are increasingly recognized for humanoid robots driven by humanlike intelligence.

Humanoid ethics Robot ethics³⁸ [20] concerns about the privacy, manipulation, opacity and bias of robotic systems and the equilibrium effect and consequence of robotic designs, mechanisms (e.g., autonomy and predictive decision-making) and behaviors during operations, decision-making and human-robot interaction, etc. *Ethical robots* are built with transparent, fair and unbiased design and decision-making; proactive compliance, due process and auditing to potential effects, influences, vulnerability, deception, misinformation, synthetic issues, failures, changes and uncertainty; and accountable compliance, interference and intervention measures and mechanisms. Humanoid robots may further involve humane aspects such as confirmation bias, deception, malicious design, attack and manipulation. Depending on the level of ethical enforcement, a humanoid robot may be explicitly ethical, implicitly ethical, or fully ethical.

Humanoid personality, intentionality and empathy In humanizing and personifying robots toward humanlike, humanlevel and humane intelligent systems, humanoid robots are further humanized with properties, traits and issues such as humanlike aggressiveness, morality, consciousness and personality [30]. For example, a humanoid robot may present extroverted, agreeable, conscientious, neurotic, open or aggressive personality, forming *robot personality* in articulating robot features and functions and in robot tasking and employment. A humanoid may be rational with intention and consciousness for being or doing good, with moral and empathetic actions and behaviors during tasking or human-humanoid interaction, teaming and collaboration, which creates *robot intentionality* and *robot rationality*. In addition, robot personality, intentionality and empathy require bidirectional design and support. On one hand, a robot can access a human or other robot's cognitive, emotional and affective states, and then behaves and interacts with them

³⁸Ethics of Artificial Intelligence and Robotics: <https://plato.stanford.edu/entries/ethics-ai/>

with empathy and morality and meeting human comfort. On the other hand, humans can fully access and control a robot's intentional stance and personality, robot behaviors can be managed and interpreted by a human being.

Humanoid trust and responsibility Robot trust refers to unidirectional, reciprocal and bidirectional human-humanoid reliability and confidence, i.e., by humans in a robot's roles, responsibilities, capabilities, actions, functions, decisions, and performance, and vice versa. Factors influence human trust in robots may be various, e.g., lack of emotional intelligence, limited ability to adapt to new situations, unclear decision-making processes, privacy concerns, negative media portrayal, unclear code of ethics, or regulatory oversight. Accordingly, the trust and responsibilities of humanoid robots may be embodied by various functional and nonfunctional mechanisms and measures, such as reliability, adaptability, security measures, programmed ethical guidelines, intention transparency, ethical behaviors, ability to self-control and self-regulation, trustful past track record, and subjectivity such as empathy and responsibility. These robot trust factors and measures need to be embodied by the trustfulness in robot design, behaviors, effect and performance, such as automation, live recognition, and real-time prediction and decision-making; the predictability and transparency of risk, safety, security and consequence; the detection, prevention and resolution of system issues such as faults, unsuccessful task completions, and irreparable mistakes.

As subjective properties increasingly affect the humanization of robots, enforcing the morality of humanoid robots essentially enables what moral liability, accountability, rights and responsibilities a humanoid should hold and how such attributes should be distributed in multi-party settings. However, robot empathy would not serve as the final guard rail. Robot trust may further require robot legal rights, such that a robot must not incur human injuries and must obey orders. This triggers the study on the rights of humanoids and artificial consciousness, e.g., whether a humanoid is a legal entity, and to what extent a robot can be given what legal rights comparable to humans. Robot legal rights are crucial for humanoid robot services as law enforcement, a robo-adviser, an elderly companion, or a child caregiver.

Humanoid risk and compliance A humanoid robot may incur or be associated with various potential risky scenarios or actual risks and dangers. *Humanoid risks* relate to methodological and technical risks, such as fragile generalization, faulty designs, and incomplete compliance; malicious manipulation such as misuses, misinformation, malicious attacks, and manipulated behaviors and automation; ethical issues such as deception and privacy concerns; and unexpected exceptions, such as hallucination of automated generation, and singularity and superintelligence of self-improvement and intelligence emergence, when humanoids gain control and rights beyond human perception and acceptance. As real-time, interactive, decentralized and personalized automated tasks and operations of humanoid robots become widespread, *robot surveillance* not only requires automated, real-time, tailored and predictive risk and compliance monitoring, diagnosis and management, but also human surveillance i.e. keeping human in the compliance loop. In this regard, significant challenges exist in real-time and automated scenario-oriented human-humanoid regulation, e.g., in accountable human-robot role allocation and responsibility sharing, efficient regulation in complex settings, privacy-preserving surveillance, risk-aware human-humanoid cooperation, and essential human-centered control and mitigation.

4 ENABLING HUMANE AND HUMANLEVEL HUMANOIDS

There are many techniques required to enable humane and humanlevel robots, such as mechanical, material, biomedical, electrical and anthropomorphic designs. Here, we focus on intelligent techniques to enable humane and humanlevel robots, humanizing robots toward humane and humanlevel features, structures, functions and moral traits; digitizing human features in robotics; and intelligentizing robots with human intelligence in complex decentralized, distributed, or even virtualized applications and environments. These include essential studies on building *mind-to-action* mindful and

actionable humanoids, supporting omnimodal *perception-to-behavior* modeling, advancing humanoid with humanlevel AI, hybridizing humanoid with metaverse and mixed reality, and hybridizing humanoid with decentralized AI, to name a few.

4.1 Mind-to-Action Mindful and Actionable Humanoids

The concepts of AI mind, mindful AI with AI mindfulness, and AI action, actionable AI and AI actionability [11] apply to humanoid robots as well, aiming for producing mind-to-action embodied mindful and actionable humanoids. In addition, humanoid mind models make cognitive humanoids possible, where cognition-to-action translation and transformation would be possible and essential.

On one hand, making humanoids humanlike and humane requires to build humanoid mind and develop mindful humanoids. A *mindful humanoid* with AI mindfulness simulates and amalgamates human mind and mindfulness to develop *humanoid mind models* and *humanoid mindfulness*. Mindful humanoids may fulfill relevant human thinking mechanisms, cognitive and psychological traits, sensation, reasoning, and critical analysis capabilities, etc. They are also mindful of their goal, plan, activity and effect as a result of applying certain personality, intention, emotion or action.

On the other hand, humanoids undertake various actions relating to their mind, goals, tasks and environments. *Actionable humanoids* not only conduct actions but also make them meaningful and valuable technically and practically, i.e., making humanoids actionable. The *actionability* [13] of humanoids, including their designs, actions, and performance, satisfies various performance measures, e.g., interpretable to end users, convertible to business value and impact, and satisfying business, social and economic measures.

Further, a humanoid robot as a closed-form independent AI system needs the transformation from mind to action. *Mind-to-action* translation requires a real-time translation and transfer of intention and emotions etc subjective mindfulness to actionable operations and activities; the reflection and feedback on actions to mind; and the iterative processes between mind and actions over a task period. This mind-to-action translation raises many research challenges and opportunities, such as mind representation (decomposed to intention representation, sentiment and emotion representation, etc.), building mind models for humanoids, dynamic and interactive alignment between mind and actions, and coordination between multiple thoughts and their corresponding actions.

Mindful humanoids are cognitive, which replicates and tailors human cognitive intelligence, traits and properties into humanoid cognitive models. Humanoid cognition may be characterized through representing humanoid's intention, affection, emotion, speech, gesture, actions, haptic signals, and physiological signals, etc. The mind-to-action translation demands *cognition-to-action* modeling and translation, which integrates the relevant modeling tasks, including personality modeling, intention learning, goal learning, emotion learning, action learning, and behavior learning, and translates learned knowledge on humanoid cognition to robot actions.

4.2 Omnimodal Transitional Humanoid Modeling

To enable mind-to-action humanoids, a critical task and challenge is the *perception-to-behavior* transition, translation and transformation, which involve multiple to omniversal modalities from visual, acoustic and textual to behavioral aspects and the translation and transformation from perception to behavior undertaking where these modalities may be involved.

Accordingly, this omnimodal perception-to-behavior translation and transformation trigger various challenges, such as multimodal to omnimodal perception fusion by synergizing visual, vocal, textual and behavioral representations; point-based, sequential, spatial or function-oriented fitting, alignment and matching between image (or video), sound

and text for the same task or object; and temporal, spatial or spatiotemporal developments of multimodal perceptions and behaviors.

In recent years, increasing attention has been paid to multimodal deep learning and specifically multimodal large models, such as visual query-answering (VQA), VLMs, VLP, image-language pretraining, image-to-text generation, and text-to-image generation, etc. Building on Google Robotic Transformer 1 (RT-1)³⁹ on multitask modeling [7], the new Google Robotic Transformer 2 (RT-2)⁴⁰ [8] further supports high-capacity VLA modeling on web data and robotic vision data, and translates the learned knowledge into generalised instructions to control robotic actions. RT-2 uses the pathways language and image model (PaLI-X) and pathways language model embodied (PaLM-E) as the backbone.

With the quantification of robot personality, intention, sentiment and emotion etc subjectivity and development of humanoid mind models, further studies will be inclined to the fitting, alignment, reasoning and inference, matching and reasoning of omnimodal mind-to-action translation. This requires further studies on transforming and translating from vision-language modeling and vision-to-action modeling into mind-to-action modeling and perception-to-behavior modeling.

4.3 Humanlike Humanoids with Humanlevel AI

The fast-paced development and applications of generative AI towards humanlevel AI is transforming and revolutionizing humanoid design paradigms. It has never been so promising to advance humanoids with humanlevel AI and create humanlevel humanoid intelligence.

The advances and futures of generative-to-humanlevel AI landscape deepen and widen the spread and expansion of humanlike humanoid studies to new-form and new-path problems, research methods, and functional and practical means. Examples are Transformer [52], vision transformer [29] and Internet data enabled generative pretrained LLMs, VLMs [44], VLA models [8], reinforcement learning with human feedback (RLHF) supported prompt engineering and chain-of-thought prompting transforming multi-step into intermediate reasoning [54], and vision-language translation for real-time multi-lingual conversations, multi-type tasking, multimodal interactions, and pathway-oriented processing. The synergy between these advances and humanoids creates the new generation of humanoids: humanlevel humanoids.

Humanlevel humanoids copes with varied specific purposes, functions, tasks, and applications, we highlight a few: multi-task and multi-party settings; reinforcement from live feedback, refinement and adaptation; and omnimodal vision-language-action modeling and mind-to-action translation.

Multi-task and multi-party settings: Humanlevel humanoids may handle multiple tasks concurrently, recurrently or sequentially in an individual, team or adversarial mode. Examples of multi-task settings are hybridizing language parsing, vision tasks with control; dialogue-based tasking with coordination through exchanging information; and translating and reasoning tasks from perception or through conversations in natural language. Multi-party settings involve multiple humanoid robots, humanoids with humans, or with third parties, where they coordinate, collaborate or compete with each other to undertake tasks. These multi-task and multi-party settings require corresponding design, coordination, communication, and control, e.g., identifying, recognizing and differentiating visual, vocal and textual information from different humans who are involved in a multi-party humanoid-human conversation, and responding correspondingly to the right party in their interest and preferred sentiment.

Reinforcement from live feedback, refinement and adaptation: In live scenarios, acquiring live feedback from humans to humanoids, supporting real-time, edge-based or on-humanoid fine-tuning and making them adaptive to the onsite

³⁹Google Robotic Transformer 1 (RT-1): <https://blog.research.google/2022/12/rt-1-robotics-transformer-for-real.html>

⁴⁰Google Robotic Transformer 2 (RT-2): <https://deepmind.google/discover/blog/rt-2-new-model-translates-vision-and-language-into-action/>

and live interaction, coordination, communication, tasking and teaming scenes, workflows, dynamics and changes form essential functions as well as challenges. Examples are feedback-based humanoid tasking and coordination through receiving, processing and transforming human feedback on task performance, validation of responses and iterative refinement with feedback. To enable live feedback-based refinement and adaptation, RLHF needs to represent, incorporate live response, feedback and interaction into humanoid models, optimize objectives, refresh functions, and update responses accordingly.

Omnimodal vision-language-action modeling and mind-to-action translation: as discussed in Sections 4.1 and 4.2, these manage the requirements and functions for enabling humanoids to real-time, interactive multimodal fusion, translation, and transformation of multiple modalities of input signals or between them. For example, upgrading LLMs and VLP such as language-image pretraining and language-video pretraining to mind-to-action and perception-to-behavior modeling is essential to understand, translate and reason about multimodal information fusion, translating input recognition to decision actions, etc. In addition, humanoids may also perceive and understand environmental constraints and task semantics to activate, adjust, optimize and operate humanoid actions and task performance.

4.4 Humanoid Animation, Imitation, Digital Twins and Metaverse

Hybridizing humanoid robotics with metaverse and mixed reality will create novel humanoid animation, humanoid-human or humanoid-society digital twins and metaverse continuum and promote humanoid virtuality-to-reality. Metaverse, digital twin, and virtual, augmented and mixed reality techniques can assist human-humanoid interaction, interfacing and collaboration in creating virtuality-reality coexisting humanoid-human design, development, workplace and tasking settings, and expand humanoid's capabilities and capacity [3, 15, 48, 53]. Animation, metaverse and mixed reality etc can form on-robot, on-body or on-environment integration into virtuality-reality-combined robot augmentation, collaborative programming, teleoperation, projected environment and background, trajectory visualization, navigation, augmenting perception and telepresence, and increasing expressiveness.

For example, by creating a 3D presentation of a humanoid and its workplace, a robot can be operated, monitored or tutored for specific tasks in its working environment. Instructors with multi-modal virtual, augmented and mixed reality devices can operate and control humanoid robots and conduct various robotic tasks, such as imposing robot movement restrictions, trajectory teaching, changing robot joint angles, and creating task-specific programs. For human-unfriendly environments, humanoids can be teleoperated to undertake specific tasks remotely using projection mapping to facilitate robotic tasks.

4.4.1 Humanoid Virtuality-to-Reality. Humanoid virtuality-to-reality can enable humanoid operation and manipulation by simulation, imitation, animation, demonstration, and generation.

Humanoid Animation- and Simulation-to-Operation Large-scale robot manipulation and experiments etc may be costly, risky, insecure, unsafe or unactionable. *Robot animation, simulation and programming* provide theories and tools to mimic, evaluate, adjust, optimize and program robots in a virtualized cost-effective platform before their confirmation, production and deployment. Robotics animators and simulators may provide interfaces, methods and functions to describe, select, connect, code, expand and manage robot goals, types, parts, coordinates, objects, environments, functions, tools, workflow, paths and behaviors, etc. in a 2D or 3D animation or simulation platform. They can simulate physical systems, activities and processes, and generate, compile and deploy the animated, simulated or optimized programs into physical robots for further experiment, prototyping, production or deployment. Humanoid animation and simulation may clone, virtualize, imitate or replicate human body parts, attributes, behaviors, emotions,

conversations and expressions, generate humane robotic parts, properties, emotions, and actions for more humanlike appearance, presentation, and mechanisms. Further, humanoid learning, planning, control and interaction by animation and simulation emerge as essential mechanisms for robotics at scale.

Robot animation, simulation and programming may face so-called *simulation-to-reality gaps* (sim-to-real gaps for short) from real-life robot appearance, design, manipulation and experiment. Gaps may be related to misalignment between simulation and reality, and the unrealisticness and anthropomorphic quality of simulation assumptions and systems. These may be caused by simulation constraints on robot settings, simulation platforms, essential and sufficient toolsets, simulator behavior settings, programming tools, simulator capabilities and capacity, and rendering quality, etc.

Humanoid Imitation-to-Operation *Imitation learning* is another type of simulation or demonstration. It can serve as an alternative to reinforcement learning. Robots learn by demonstrations or are taught by a supervisor, demonstrator or examples, such as from virtual simulations or a filmed physical scenario (a video), and then clone, imitate or adjust the demonstrations or examples. Robots may also learn the underlying optimal reward function or policies from teachers, adapt to the taught examples, develop or generate new behaviors. Robot imitation thus supports behavior cloning, intent learning, behavior or scene prediction, recommendation or generation. By integrating techniques like DeepFake, animation, LLMs and digital twins, humanoid imitation could learn from sequential, hierarchical, multimodal and live human activities, and social activities; processes, mechanisms and strategies for handling complex tasks; general and specific skills for problem-solving; and intention, morality and humanity exhibited during human actions and interaction. Various extensions include learning by imitation, learning to imitate [47], generation by imitation, domain adaptation or domain transfer. Humanoid imitation also face issues like imitation-to-reality gaps, in particular, for unseen, new and open world objects, scenarios and tasks, and efficient and effective in-distribution, in-domain and in-context imitation and imitation at scale.

Humanoid Digital Twins and Metaverse While the integration of humanoids with digital twins and metaverse is an open area yet explored, integrating humanoids with digital twins and metaverse may create humanoid digital twins and humanoid metaverse useful for designing, developing and managing specific applications. Humanoid digital twins and humanoid metaverse may be built using a collection of mixed techniques, including humanoid robotics, LLMs and LMMs, digital twins, metaverse, decentralized AI techniques, Web3, 3D design, and operating and software systems to enable these varied techniques.

Examples are humanoid digital twins or metaverse for digital health, digital finance, digital marketing, digital customer services, and digital disaster management. For example, a *humanoid general practitioner* (HGP) may offer online telehealth services as follows. Once receiving the appointment request by a patient, the HGP may launch its virtual digital twin or metaverse platform, where the HGP can use a virtual digital human to illustrate, inquire and discuss health or medical concerns and symptoms with the patient. The patient can tell her problems, upload her medical test results to the metaverse, HGP can then extract, process and analyze the uploaded materials, animate, illustrate and explain the patient's problems through the digital twin mapped to the patient per her demographics and medical conditions, and make assessment of the patient's complaints. Further diagnosis and treatment suggestions may be made and explained to the patient using the digital human. All discussion, materials and HGP diagnosis and analysis can be sent to a human doctor for verification, adjustment or confirmation. This HGP with health digital twin could save time, extend coverage, and augment service depth and health education etc of healthcare and telehealth. Another example would be humanoid robo-advising supported by metaverse for digital finance.

4.4.2 Humanoid Demonstration-to-Generation. Generative humanoid intelligence can be learned, augmented and generated by demonstrations, advanced models such as LLMs, VLM and LMMs, and broadly by generative-to-humanlevel AI.

Humanoid Demonstration-to-Operation *Robot demonstration* illustrates robots with their expected functions, tasks, scenarios, behaviors, actions, and processes, etc. Robots can then learn from single, few or sequential demonstrations to operate, plan, navigate or evolve over existing, new, unseen, open or lifelong tasks or environments. Humanoids could learn from human demonstrations, such as positioning, posture, behaving, emotional, vocal and facial expressions, and ethical responses, to improve, evolve or develop their world knowledge, humanlike actions and responses. Further, robots may clone, optimize and expand the learned knowledge, activities or intelligence from demonstrations, and even forecast and generate new actions, knowledge or intelligence through generative AI.

Robot learning by demonstration leverages robot simulation and mitigates some of sim-to-real gaps by enhancing robot’s familiarity and realisticness of realistic and open worlds, and building robot experience, memory and reasoning. Similar to robot simulation, robot demonstration faces *demonstration-to-reality gaps* (or Demo-to-real gaps for short). Demo-to-real gaps may be caused by demonstration biases, distribution shift, misalignment, or the limitation and insufficiency of demonstration number (e.g., from zero shot, one shot to few shots), settings, capabilities, and capacity, etc. Robot learning by demonstration still faces significant challenges for unstructured, ill-structured, unseen-structured, unknown-structured, transitional, lifelong, nonstationary settings and environments. Both robot simulation and demonstration are also challenged by zero, new, evolving, unseen, unknown and open objects, contexts, scenarios or tasks.

Humanoid Generation-to-Operation With the fast-pace development of generative AI including LLMs, LMMs, VQA, VLPs, VLMs and VLAs, *automated generation* becomes a feasible and essential intelligent means to simulate and produce AI and realize more complex humanlevel intelligence, which could substantially improve the intelligence of robots. *Humanoid intelligence generation* may present new opportunities and challenges than existing LLMs, such as VQA, VLP and VLM, e.g., for image-to-text, video-to-text, text-to-image, and text-to-video, making the generation suitable for humanoids. Examples are 1) single-modal generation, e.g., generating humanoid behaviors, actions, responses, and emotions; and 2) cross-modal generation, such as image-to-speech, image-to-conversation, image-to-emotion, animation-to-speech, animation-to-conversation, animation-to-singing, animation-to-dancing, text-to-speech, text-to-conversation, text-to-action, and text-to-emotion.

Humanoid generation could be built on humanoid animation, simulation, imitation and demonstration, where animation, simulation, imitation and demonstration serve as supervision, providing one to few shots for robot learning and further generation. Zero-to-few shot learning, weakly-supervised learning, transfer learning, contrastive learning, and Knowledge distillation etc would be useful to augment humanoid learning and generation by LLMs or LMMs.

4.5 Decentralized Humanoids: On-robot, Edge and Cloud Humanoids

Decentralized Humanoids Decentralized AI builds on decentralized autonomous organizations (DAOs), blockchain technologies, Web3, decentralized finance, cryptocurrencies, decentralized science (DeSci), and other decentralized technologies to enable, develop and support AI tasks, behaviors and systems over the web, cloud or edge or on device. Hybridizing humanoids with decentralized AI could explore various applications, e.g., 1) connecting humanoids to cloud facilities and learning systems, enabled by cloud AI such as ChatGPT; 2) connecting and enabling humanoid edge networks by connecting humanoids in a community and implementing data-sharing, model-sharing, multi-task, and capability-sharing humanoid edge intelligence; 3) enabling on-humanoid AI such as on-robot live perception, action

and response and peer-to-peer on-robot AI tasks and activities; and 4) robot-robot edge and cloud hybrid humanoid networks and intelligent tasks and systems.

Decentralizing humanoids also requires an edge network, which may develop humanoid edge intelligence, such as humanoid interfacing, interaction, conversation, communication, cooperation, collaboration, sharing, storage, memorization, action and intervention over humanoid edge networks. These further involve techniques for privacy-preserving, secure, trustful, responsible, and efficient operations, control, communication, and collaborations. Among the humanoid edge network, decomposing, splitting, dispatching, distributing, pipelining, orchestrating, aggregating or balancing roles, responsibilities, tasks, data, and resource are also important issues to address.

On-Humanoid, Edge and Cloud Humanoids On-humanoid, edge and cloud Humanoids will create a new era of humanoid robotics and bring about enormous new challenges and opportunities of robotics, humanoid and decentralized AI. Currently, humanoids enabled by LLMs are cloud humanoids. Cloud humanoids only have limited on-robot capabilities and capacity, they depend on live high-speed connection and communication to humanoid cloud platforms and enabling services, such as ChatGPT for performing LLMs. This makes cloud humanoids dependent on specific cloud infrastructure, computing and services, making it costly and inflexible for operations and development. It also makes humanoids restrictive to wifi or telco network connection, making humanoids incapable of on-device automation and causing issues including latency, unreliability and risk in real-time interactions, coordination and communications. Cloud humanoids also raise concerns on sharing privacy, safety and security of data and humanoids to cloud service providers, and making regulation difficult to be accountable instantly. Accordingly, on-humanoid and edge humanoid development emerge for more independent, automated decentralized robotic and humanoid intelligence.

On-humanoid intelligence, intelligent applications and services require energy-efficient, information-compressed, real-time, personalized, interactive and privacy-preserving platforms, applications, and communications. A humanoid is an independent automated, self-organizing intelligent agent. It can operate and manage to sense, interact, navigate, judge, plan, act, control and decide, hold its own goals, undertake actions, and assess the live environment independently in real time. For this, on-humanoid research will become an essential area to enable on-humanoid perception, recognition, conversation, interaction, regulation, etc. For example, LLMs and VLAs trained on cloud will have to be converted to humanoid-compatible LLMs and VLAs, such as by selecting mobile apps-discriminate parameters through selective parameter-efficient fine tuning. For *edge humanoid* intelligence and systems connecting to humanoid robots, the edge network needs to handle energy-efficient, privacy-preserving, humanoid-neutral and -specific data acquisition, recognition, analytics, learning, reasoning, planning and communication, etc. Edge humanoid network hosts techniques and functions to support both the networking, communications and shared functions of humanoid network and client humanoids.

Fig. 5 illustrates a framework of decentralized humanoid AI systems, where humans interact humanoids in multimodal channels and manipulate humanoids through humanoid edge desktops. Humanoids connect to their cloud server and communicate the perceived information with humans to the server, compute and learn the information, retrieve, request or execute Internet-based third-party cloud services such as ChatGPT and GitHub-based open source APIs or applications. The processed response will then be deployed by humanoid cloud servers to humanoids for executing actions.

5 OPEN ISSUES AND DIRECTIONS

The above discussion on new-generation humanoids, such as developing functional and nonfunctional requirements, enabling mindful and actionable humanoids, humanoid digital twins and metaverse, and decentralized humanoids open

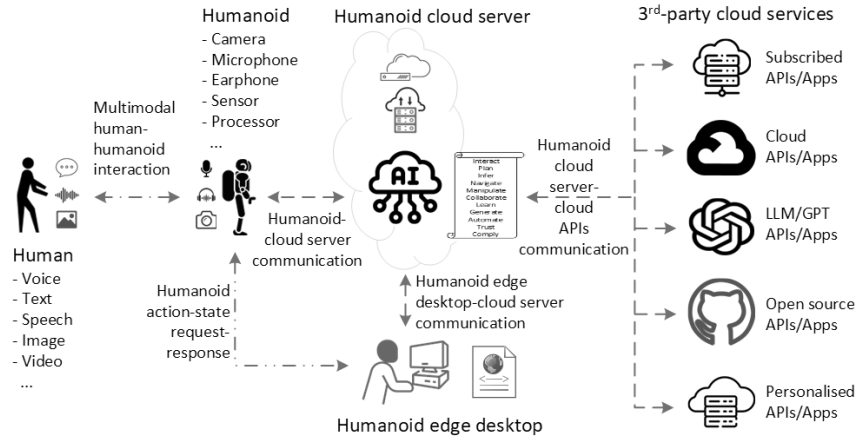


Fig. 5. Decentralized humanoids: On-humanoid, edge and cloud humanoid AI framework, synergizing humans, humanoids, edge and cloud devices, algorithms and services including LLMs.

a new age of humanoid robotics, proposing many open issues and directions beyond the existing humanoid studies and developments.

5.1 Challenges in Humane and Humanlevel Humanoids

While generative AI advances including LLMs and LMMs has the potential to revolutionize humanoid robots toward embracing humanlike AI, making humane and humanlevel humanoids is still an ambitious task with a suite of challenges, including the maturity and applicability of existing generative AI to humanoids, the constraints on humanoid systems, and the challenges in enabling humanlevel humanoid robotics.

There are still many open issues and challenges, such as hallucinations, in existing generative AI, including LLMs and LMMs. Issues and challenges may come from or be associated with data, modeling, evaluation, ethics, societal and human elements. Examples are bias and fairness, distribution shift, contextual understanding, multimodal capabilities, personalized sparse satisfaction, adversarial attacks, misinformation and disinformation, manipulation and deception, observation and data overreliance, overfitting or overmanipulation, lack of thinking and humanity, and potential adoption of technical, security, privacy, ethical, economic, psychological, sociocultural or sociopolitical risks.

Second, humanoid robots pose specific constraints and challenges in developing humane and humanlevel robotic intelligence. These include limitations on robotic locomotion, mechanics, sensing, responding capabilities, memory, degree of freedom, dimensionality, modality, and computational capabilities, which cannot support LLMs and LMMs on devices and edge networks. More fundamental limitations lie in the existing designs and mechanisms of enabling robotic manipulation, intelligence such as vision, perception, learning, memorization, control, planning, and actions. Simulating and replicating human intelligence into humanoids requires substantial theoretical and implementation advances in humanoid and robotic cognition, humanity, biomechanics, and behavioral studies, etc. To avoid the uncanny valley phenomena of humanoids and the uncanny valley risks and hallucinations of LLMs into LLM-enabled humanoids, humanoid robotics expects not only humanlooking machines but also humane expression, behaving, response, interaction, conservation, and collaboration, etc.

Accordingly, humanlevel humanoid robotics requires new-generation LLMs and LMMs and develops humanlevel humanoid AI. This may aim for developing humanlike robotic intelligence before achieving any superintelligence or superhuman intelligence. Examples are to develop 1) on-humanoid LLMs, VLP, VLM and LMMs; 2) humanoid-edge (local server or multiple humanoid networks) LLMs and LMMs; 3) humanoid humanity; 4) humanlike humanoid intelligent capabilities such as inspiration from children learning procedure, human personality, intention, subjectivity and empathy under positive, negative and neutral scenarios and activities; 5) humanlike social humanoid intelligent capabilities such as team coordination, cooperation, competition, and conflict resolution during social activities.

In the following sections, we further discuss issues and opportunities relating to humanoid humanity and subjectivity, quality and service assurance, live humanoid tasking in complex environments, and open humanoid tasking in open worlds.

5.2 Humanoid Humanity and Subjectivity

The humanity and subjectivity of humanoids characterize the humane and subjective traits, attributes, qualities and states of humanoids of being humankind, humanlike to humanlevel. Humanoid humanities are embodied through humane attributes such as personality, morality, intention, compassion, sentiment and emotion, enabling and managing responsible, trustworthy, sympathetic and ethical individual and collective disposition, behaviors and decisions.

Existing work on embodying these humane and subjective attributes, traits and qualities into humanoids is very limited and relies on facial expression and understanding of perception and observations. Examples are human mood modeling, such as visual facial expression recognition, textual semantics such as by topic modeling and sentiment analysis, and human body state modeling such as posture recognition [36, 37]. They are observational, embodying and modeling humanoid humanity and subjectivity may require substantial new thinking and theories. This goes beyond the existing humanoid design enabled by generative AI and large multimodal modeling, which cannot support humankind live and real-time interaction and collaboration.

More generally, it is essential to build a theory of humanoid humanity and subjectivity, e.g., by designing humanoid humanity models, quantifying humanoid subjectivity, structuring humanoid trustfulness, simulating humanoid subjectivity, benchmarking humanoid subjectivity, and developing rules of law for humanoid humanity and subjectivity.

- Characterizing humanoid humanity: characterizing and learning humanoid robot's personality, intentionality, consciousness, empathy and trust in real time and live scenarios driven by generative AI and large multimodal models; defining, imitating and simulating attributes e.g. extroversion, conscientiousness, neuroticism, aggressivity, emotion and intention; and quantifying them in terms of multimodal vision-to-language, perception-to-action tasking, or self-motivated empathy etc settings.
- Structuring humanoid trust and responsibilities: defining, specifying and simulating the trustfulness and responsibilities of humanoid robots by humans and humanoid perception and confidence in humans during humanoid and human interaction and collaboration; modeling trust [56] and responsibility in context of perception-to-action pretraining and finetuning, e.g., quantifying humanoid's LLM and perception-to-action performance with performance measures and criteria such as robustness, reliability, accuracy, and user satisfaction; formulating humanoid's self-regulation and self-assessment capability and mechanisms, e.g., accountability, interpretability, and exception resolution; exploring synergy between external regulatory agents and humanoids for humanoid ethical assurance, and functional and nonfunctional regulation and governance.

- Humanoid humanity simulation: creating use cases, strategies and simulation testbeds for mimicking, generating, implementing, monitoring and adjusting humanoid personality, intentionality, empathy and trust; simulating humanity dynamics and evolution over interactions and collaborations; simulating and controlling subjective expectations under scenarios such as self-improving emergence, immoral and irrational actions, free of risk and compliance regulation.
- Benchmarking humanoid humanity and subjectivity: developing benchmarkable database, use cases, testbed and checklist for experimentation and case studies of humanoid personality, intentionality, consciousness, empathy and trust; exploring open platforms for uncertain, diversified and large-scale scenarios such as self-improving emergence and hallucination, sampling and algorithmic biases, and irreparable mistakes.
- Rules of law for humanoid humanity management: discovering and summarizing general principles, rules of law for monitoring, detecting, predicting, intervening and managing personality, intentionality, consciousness, empathy and trust of real-time, interactive and multimodal humanoid robots; identifying verifiable rules and code of conduct for different use cases, such as humanoid tasking, humanoid-human teaming, interactions, and collaborations.

5.3 Live Humanoid Tasking in Complex Environments

In principle, the goal of humanoids is to make them actionable in live settings for live tasking. This would make humanoids expressive with multiple modalities and aspects of capabilities and intelligences, and be capable for complex robotic tasking. Examples are humanoids with diversified multiple goals and tasks as humans for multimodal tasks, multi-party tasks, and real-time tasks in real and live environments. This raises many interesting but challenging opportunities, below, we highlight a few.

Humanoid tasking in complex live scenes aims to enable humanoid operations in real-world settings and environments, such as with multiple robot sensors and parties, real-time actions and processes, low-quality backgrounds, and immediate response requirements. This requires fundamental facilities such as sensor and signal coordination and fusion, instant multimodal response generation and alignment, multi-party coordination, and human-humanoid-shared mental modeling.

Data characteristics and complexities resilient humanoid tasking enables humanoids to work in poor quality environments and be agile and resilient to complex data and behaviors during tasking. Data includes historical and background information such as from the Internet, live data and behaviors during interactions and processes, external information required during operations. Making humanoid functions and operations responsive to these sources of data and their data changes in real time poses significant challenges and opportunities for efficient live robot modeling and computing in real-time settings.

Multimodal humanoid tasking aims to engage and integrate multiple modalities of humanoid robots, such as visual, vocal and behavioral functions. Hybrid and integrative robot tasks may engage and integrate these modalities, e.g., synergizing gesture, speech and action to make robots more expressive, multimodal and multi-aspect.

Multi-party humanoid tasking facilitate tasks with multiple robots, robots and humans, and robots and other agents or systems (e.g., metaverse, attackers or defenders). Moderating robot-to-robot, human-robot, and human-robot-agent teaming, communication, interaction, collaboration, or competition etc becomes increasingly demanding for real-world robotic applications. This requires technical tools to enable multi-party humanoid-human dialogue or debate, cooperation between humanoids and human collaborators playing different roles.

Real-time and interactive humanoid tasking further synergizes multi-modalities and multi-parties for real-time, interactive tasking [26]. Facilitating live interactions, collaborations or defence needs to acquire multisources of data and information, fuse multimodal signals, and enable multi-party conversations. This poses many challenges, e.g., real-time multimodal and multi-sensor information fusion, alignment between humans and humanoids, etc.

Bilateral human-humanoid modeling. Existing humanoid studies often focus on one-sided manipulation, i.e., empowering humanoids toward intelligent capabilities and tasking. To make humanoid humane and humanlike, bilateral human-humanoid modeling would be essential, where humanoids are treated as pseudo human beings for more natural, humanlike interactions with humans, while humans can also engage humanoids in human ways. Accordingly, we need to build theories and tools for bilateral understanding, interaction, conversation, identification, recognition, collaboration and management both from human to humanoid and from humanoid to human. This means both humans and humanoids 1) have the abilities to identify, select and engage interesting partners; 2) can actively and proactively initiate, control and manage conversations, tasks and engagement with the counterparts on-demand; and 3) hold the conditions, rights and criteria to determine interactions and collaborations.

5.4 Open Humanoid Tasking in Open Worlds

Live robotic tasking further raises the requirements for open tasking in open settings and their corresponding challenges. In fact, real-time interactive humanoids and human-humanoid systems in live environments are open complex intelligent systems [10], live in open environments, undertake open tasks, and act under open settings.

Open humanoid systems. A live humanoid is open, demonstrating various types and aspects of partial to systemic openesses. Examples include internal interactions within and between objects and subsystems of a humanoid; external interactions between a humanoid and its community; dynamics of humanoid parts, subsystems and systems; and uncertainties of humanoid part and system's states, state transition, and performance. *Humanoid open structuring* may be required for automated humanoids in complex environments, which require evolving, on-demand and self-organizing structure configuration, modularization, architecture search and discovery, memory customization, and modular and functional activation and deactivation. These require corresponding designs and mechanisms to characterize, define, enable, organize and manage open humanoid systems.

Open humanoid environments. In an open environment, a humanoid may present openness in terms of 1) uncertain, changing, unseen and unknown contexts; 2) dynamic and uncertain information, energy and communication flow between humanoids and their environments; 3) changing constituents and structures of a humanoid; 4) evolving tasking requirements and capabilities of a humanoid; and 5) potential nonstationary quality, service and performance of a humanoid. These scenarios and issues require dynamic, online, active, evolving and self-organizing humanoid capabilities, i.e., *open capabilities*, for (re)alignment, recognition, (re)arrangement, (re)manipulation, adjustment, (re)adaptation and transfer.

Open humanoid settings may broadly refer to 1) open input scenarios, such as open data, open domain, open vocabulary, open context and open environment, which are dynamic and evolving with change, drift and shift over time, space, frequency, structure/distribution or semantics; 2) open tasking requirements, as discussed below on open humanoid tasking; and 3) open output requirements, such as open set (with new and unseen labels), open-structured outputs (e.g., changing, new and unknown forms, formats, layouts, distributions), open evaluation (e.g., with evolving criteria and objectives), and open feedback (e.g., changing reviews, opinions, result preference, and quality expectation on outputs). These require humanoids to be sufficiently flexible, adaptive, evolving and self-organizing.

Open humanoid tasking. Live humanoid tasking may be open, catering for open humanoid systems, environments and settings. In addition, open tasking may also include tasks that are open, e.g., open data processing (e.g., processing data with evolving, uncertain, unknown and unseen nature); open goal settings (e.g., without firm and predefined goals and objectives); open perception (such as perceiving open worlds and enabling evolving, reconfigurable and self-developing perception tools and capabilities); open representation (e.g., with embedding capabilities of coping with open inputs); open learning (e.g., under open settings, with evolving learning objectives, changing learning structures, or open output requirements); open reasoning (e.g., reasoning over uncertain actions, without constraints, or under open conditions); open actioning (e.g., enabling new, changing and beyond-design actions and behaviors); open control (e.g., enabling control in open environment and evolving tasking); open planning (e.g., pursuing plans in open contexts, planning by learning from open data and objectives); open evaluation (e.g., changing or adaptive evaluation criteria, methods or measures over input conditions and human feedback change); open optimization (e.g., without predefined optimization objectives or criteria, evolving objectives over input condition change); and open domain transfer (e.g., the target domains are evolving, uncertain, new or unseen); etc.

5.5 Humanoid Quality and Service Assurance

In Section 3.2, we have discussed various issues relating to the quality of services of humanoids from functional and nonfunctional perspectives. Here, we further expand the discussion on HQoS to humanoid design quality, implementation quality, service quality, output trust. These are important for making humanoids mindful, actionable and ethical.

Addressing biases, fairness and trust in humanoids. GAI and LLMs may involve various types of biases [21], including sampling biases, modeling biases, and evaluation biases. *Sampling biases* may further generate biases and fairness issues, such as sample set biases, distributional biases, where samples may be associated with demographic, confirmation, memory, linguistic, ethnic/cultural, political or ideological biases, etc. *Modeling biases* may be caused by hypothesis biases, methodological biases, architectural biases, input-model or model-task matching and fitting biases, etc. *Evaluation biases* appear in biased or unfair evaluation methodologies, objectives, objective functions, evaluation measures, and evaluation processes, etc. *Humanoid trust* needs to address these biases and unfairnesses, foster trustworthiness of humanoid operations and actions from various aspects, such as faith, fear, confidence, feeling, valence, control and reciprocation of humanoids, in order to create humanoid norms, regulations, and laws [16].

Humanoid design quality assurance. LLMs and GAI-enabled ChatGPT and other models have shown their limitations and gaps such as hallucination, biased answers, irrelevant outputs, too general responses, non-personalized responses, low actionability of results, ethical concerns, security and privacy challenges. Humanoids connecting to and enabled by such LLMs and GAI are also constrained by these issues, making their outputs vulnerable, untrustworthy and nonactionable by users. Humanoid capability maturity has to address these issues, either benefiting from the improvement of their dependent LLMs and variants or addressing LLM issues within humanoid systems. Here are a few directions: conducting energy-efficient and parameter-efficient finetuning of pretrained LLMs or their results, incorporating LLM results as prior knowledge and training vertical robotic models, algorithmic debiasing, mitigating biases in training sampling, enabling active and online refinement of pretrained models, and human-in-the-loop refinement and decision-making. Humanoid systems need to address biases and trust, which requires new modeling, learning, optimization, evaluation and feedback theories and techniques to address aspects of such as complex biased scenarios, design generalization, bias and error propagation, unanticipated capability or result emergence, and alignment with human feedback for humanoids.

Humanoid implementation quality assurance. To improve humanoid quality of services and satisfy humanoid non-functional requirements, quality assurance is essential for implementing humanoids. We highlight the following aspects: quality assurance of data and processing, quality assurance of hardware and software, and quality assurance of interfacing and outcomes, aiming for assurance the quality of input, implementation, output and outcome.

First, *humanoid quality assurance of data and processing*: Both collecting historical, present and future inputs and real-time data streaming and recordings during humanoid operations and human-humanoid interactions, and their manipulation and processing require data and processing quality assurance. Data and processing quality assurance addresses data quality issues such as biases, inconsistencies, irregularity, and noises, and ensure fair and unbiased data sampling, cleaning, transformation, augmentation, denoising, and debiasing, etc. to make data ready.

Second, *humanoid quality assurance of hardware and software*: humanoid hardware and software are specially fabricated to fulfill functions and performance measures in Section 3. Accordingly, their implementation, quality and performance needs assurance, which may involve many aspects, for example, sufficient matching between hardware and software, systematic integrity, function cohesion, coding biases, and privacy and security assurance.

Last but not least, *humanoid quality assurance of interfacing and outcomes*: this includes fair and unbiased usability of user interaction, interfacing and support; interpretability and explainability of workflow, factors and outputs; auditability and accountability of operational processes, actions and results; monitoring, evaluating and intervening potential consequences, risk, safety, security or trust concerns of outputs; and overseeing, mitigating and regulating business, societal or human effects and impact.

6 CONCLUDING REMARKS

Real-time, interactive, and realistic humanoid robots epitomize the pinnacle of systematic AI development, transcending traditional robotics and intelligent systems. These cutting-edge humanoids represent a new generation of robots that seamlessly integrate mechanical, electrical, biological, psychological, physiological, ergonomic, aesthetic, and sociocultural technologies and achievements. Leveraging recent AI advancements, particularly in large language models (LLMs) and generative AI, real-time, interactive, and realistic humanoid robots demonstrate unprecedented potential in embodying humanlike features, senses, behaviors, functions, and intelligences.

Although state-of-the-art humanoids still fall short of attaining humanlike intelligence, they showcase increasingly impressive AI applications. AI-empowered humanoids are propelling the evolution of humanlooking robotics towards humanlike and humane elements, features, and functions. Prior to reaching superintelligence or humanlevel intelligence, comprehensive technical requirements, challenges, issues, and directions pave the path for the development of humanlike and humane humanoids.

ACKNOWLEDGMENTS

This work is partially sponsored by the Australian Research Council Discovery grant DP190101079, DP240102050, ARC LIEF grant LE240100131, and the ARC Future Fellowship grant FT190100734.

REFERENCES

- [1] Brenna D. Argall, Sonia Chernova, Manuela M. Veloso, and Brett Browning. 2009. A survey of robot learning from demonstration. *Robotics Auton. Syst.* 57, 5 (2009), 469–483.
- [2] Yuki Asano, Kei Okada, and Masayuki Inaba. 2017. Design principles of a human mimetic humanoid: Humanoid platform to study human intelligence and internal body system. *Science Robotics* 2, 13 (2017), eaaq0899.

- [3] Alessio Baratta, Antonio Cimino, Francesco Longo, and Letizia Nicoletti. 2024. Digital twin for human-robot collaboration enhancement in manufacturing systems: Literature review and direction for future developments. *Comput. Ind. Eng.* 187 (2024), 109764.
- [4] Leonard Bärnmann, Rainer Kartmann, Fabian Peller-Konrad, Alex Waibel, and Tamim Asfour. 2023. Incremental Learning of Humanoid Robot Behavior from Natural Interaction and Large Language Models. *CoRR* abs/2309.04316 (2023).
- [5] Andrea Maria Bauer, Dirk Wollherr, and Martin Buss. 2008. Human-Robot Collaboration: a Survey. *Int. J. Humanoid Robotics* 5, 1 (2008), 47–66.
- [6] Douglas Blackiston, Sam Kriegman, Josh C. Bongard, and Michael Levin. 2023. Biological Robots: Perspectives on an Emerging Interdisciplinary Field. *Soft robotics* 10 (2023), 674–686. Issue 4.
- [7] Anthony Brohan, Noah Brown, Justice Carbajal, and et al. 2023. RT-1: Robotics Transformer for Real-World Control at Scale. In *Robotics: Science and Systems XIX, Daegu, Republic of Korea, July 10-14, 2023*, Kostas E. Bekris, Kris Hauser, Sylvia L. Herbert, and Jingjin Yu (Eds.). <https://doi.org/10.15607/RSS.2023.XIX.025>
- [8] Anthony Brohan, Noah Brown, Justice Carbajal, and et al. 2023. RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control. *CoRR* abs/2307.15818 (2023). <https://doi.org/10.48550/ARXIV.2307.15818>
- [9] Longbing Cao. 2010. In-depth behavior understanding and use: The behavior informatics approach. *Inf. Sci.* 180, 17 (2010), 3067–3085.
- [10] Longbing Cao. 2015. *Metasynthetic Computing and Engineering of Complex Systems*. Springer.
- [11] Longbing Cao. 2022. Beyond AutoML: Mindful and Actionable AI and AutoAI with Mind and Action. *IEEE Intell. Syst.* 37, 5 (2022), 6–18.
- [12] Longbing Cao. 2022. A New Age of AI: Features and Futures. *IEEE Intell. Syst.* 37, 1 (2022), 25–37.
- [13] Longbing Cao, Dan Luo, and Chengqi Zhang. 2007. Knowledge actionability: satisfying technical and business interestingness. *IJBIDM* 2, 4 (2007), 496–514.
- [14] Alessandro Carfi and Fulvio Mastrogiovanni. 2023. Gesture-Based Human-Machine Interaction: Taxonomy, Problem Definition, and Analysis. *IEEE Trans. Cybern.* 53, 1 (2023), 497–513.
- [15] Christine T. Chang and Bradley Hayes. 2020. A Survey of Augmented Reality for Human-Robot Collaboration. *arXiv* (2020), 1–38.
- [16] Jin-Hee Cho, Kevin S. Chan, and Sibel Adali. 2015. A Survey on Trust Modeling. *ACM Comput. Surv.* 48, 2 (2015), 28:1–28:40.
- [17] Paolo Dario and Guang-Zhong Yang. 2017. Humanoid robotics - History, current state of the art, and challenges. *Sci. Robotics* 2, 13 (2017).
- [18] Kourosh Darvish, Luigi Penco, Joao Ramos, Rafael Cisneros, Jerry E. Pratt, Eiichi Yoshida, Serena Ivaldi, and Daniele Pucci. 2023. Teleoperation of Humanoid Robots: A Survey. *IEEE Trans. Robotics* 39, 3 (2023), 1706–1727.
- [19] Databricks. 2023. The great acceleration: CIO perspectives on generative AI.
- [20] Paul Dumouchel. 2023. Ethics & Robotics, Embodiment and Vulnerability. *Int. J. Soc. Robotics* 15, 12 (2023), 2087–2099.
- [21] Emilio Ferrara. 2023. Should ChatGPT be Biased? Challenges and Risks of Bias in Large Language Models. *CoRR* abs/2304.03738 (2023).
- [22] Terrence Fong, Illah R. Nourbakhsh, and Kerstin Dautenhahn. 2003. A survey of socially interactive robots. *Robotics Auton. Syst.* 42, 3-4 (2003), 143–166.
- [23] Michael A. Goodrich and Alan C. Schultz. 2007. Human-Robot Interaction: A Survey. *Found. Trends Hum. Comput. Interact.* 1, 3 (2007), 203–275.
- [24] Brian Ichter, Anthony Brohan, Yevgen Chebotar, and et al. 2022. Do As I Can, Not As I Say: Grounding Language in Robotic Affordances. In *CoRL 2022 (Proceedings of Machine Learning Research, Vol. 205)*. PMLR, 287–318.
- [25] Shuuji Kajita, Hirohisa Hirukawa, Kensuke Harada, and Kazuhito Yokoi. 2014. *Introduction to Humanoid Robotics*. Springer Tracts in Advanced Robotics, Vol. 101. Springer.
- [26] Takayuki Kanda, Hiroshi Ishiguro, Michita Imai, and Tetsuo Ono. 2004. Development and evaluation of interactive humanoid robots. *Proc. IEEE* 92, 11 (2004), 1839–1850.
- [27] Oliver Korn. 2019. *Social Robots: Technological, Societal and Ethical Aspects of Human-Robot Interaction (Ed.)* (1. ed.). Springer International Publishing.
- [28] Przemyslaw A. Lasota, Terrence Fong, and Julie A. Shah. 2017. A Survey of Methods for Safe Human-Robot Interaction. *Found. Trends Robotics* 5, 4 (2017), 261–349.
- [29] Yanghao Li, Chao-Yuan Wu, Haoqi Fan, Karttikeya Mangalam, Bo Xiong, Jitendra Malik, and Christoph Feichtenhofer. 2022. MViTv2: Improved Multiscale Vision Transformers for Classification and Detection. In *CVPR 2022*. 4794–4804.
- [30] Liangyi Luo, Kohei Ogawa, Graham Peebles, and Hiroshi Ishiguro. 2022. Towards a Personality AI for Robots: Potential Colony Capacity of a Goal-Shaped Generative Personality Model When Used for Expressing Personalities via Non-Verbal Behaviour of Humanoid Robots. *Frontiers Robotics AI* 9 (2022), 728776.
- [31] Derek McColl, Alexander Hong, Naoaki Hatakeyama, Goldie Nejat, and Beno Benhabib. 2016. A Survey of Autonomous Human Affect Detection Methods for Social Robots Engaged in Natural HRI. *J. Intell. Robot. Syst.* 82, 1 (2016), 101–133.
- [32] Bonan Min, Hayley Ross, Elior Sulem, Amir Pouran Ben Veysseh, Thien Huu Nguyen, Oscar Sainz, Eneko Agirre, Ilana Heintz, and Dan Roth. 2024. Recent Advances in Natural Language Processing via Large Pre-trained Language Models: A Survey. *ACM Comput. Surv.* 56, 2 (2024), 30:1–30:40.
- [33] Ronja Möller, Antonino Furnari, Sebastiano Battiato, Aki Härmä, and Giovanni Maria Farinella. 2021. A survey on human-aware robot navigation. *Robotics Auton. Syst.* 145 (2021), 103837.
- [34] Masahiro Mori, Karl F. MacDorman, and Norri Kageki. 2012. The Uncanny Valley [From the Field]. *IEEE Robotics Autom. Mag.* 19, 2 (2012), 98–100.
- [35] Temitayo A. Olugbade, Liang He, Perla Maiolino, Dirk Heylen, and Nadia Bianchi-Berthouze. 2023. Touch Technology in Affective Human-, Robot-, and Virtual-Human Interactions: A Survey. *Proc. IEEE* 111, 10 (2023), 1333–1354.
- [36] Ziggy O'Reilly, Davide Ghiglino, Nicolas Spatola, and Agnieszka Wykowska. 2021. Modulating the Intentional Stance: Humanoid Robots, Narrative and Autistic Traits. In *ICSR 2021 (Lecture Notes in Computer Science, Vol. 13086)*. Springer, 697–706.

- [37] Jairo Pérez-Osorio and Agnieszka Wykowska. 2019. Adopting the Intentional Stance Towards Humanoid Robots. In *Wording Robotics - Discourses and Representations on Robotics*, Jean-Paul Laumond, Emmanuelle Danblon, and Céline Pieters (Eds.). Springer Tracts in Advanced Robotics, Vol. 130. Springer, 119–136.
- [38] Silvia Proia, Raffaele Carli, Graziana Cavone, and Mariagrazia Dotoli. 2022. Control Techniques for Safe, Ergonomic, and Efficient Human-Robot Collaboration in the Digital Industry: A Survey. *IEEE Trans Autom. Sci. Eng.* 19, 3 (2022), 1798–1819.
- [39] Niyati Rawal and Ruth Maria Stock-Homburg. 2022. Facial Emotion Expressions in Human-Robot Interaction: A Survey. *Int. J. Soc. Robotics* 14, 7 (2022), 1583–1604.
- [40] Nicole L. Robinson, Brendan Tidd, Dylan Campbell, Dana Kulic, and Peter Corke. 2023. Robotic Vision for Human-Robot Interaction and Collaboration: A Survey and Systematic Review. *ACM Transactions on Human-Robot Interaction* 12, 1 (2023), 1–66.
- [41] Arindam Roychoudhury, Shahram Khorshidi, Subham Agrawal, and Maren Bennewitz. 2023. Perception for Humanoid Robots. *Curr Robot Rep* 4 (2023), 127–140.
- [42] Saeed Saavedvand, Masoumeh Jafari, Hadi S. Aghdasi, and Jacky Baltes. 2019. A comprehensive survey on humanoid robot development. *Knowl. Eng. Rev.* 34 (2019), e20.
- [43] Francesco Semeraro, Alexander Griffiths, and Angelo Cangelosi. 2023. Human-robot collaboration and machine learning: A systematic review of recent research. *Robotics Comput. Integr. Manuf.* 79 (2023), 102432.
- [44] Austin Stone, Ted Xiao, Yao Lu, Keerthana Gopalakrishnan, Kuang-Huei Lee, Quan Vuong, Paul Wohlhart, Brianna Zitkovich, Fei Xia, Chelsea Finn, and Karol Hausman. 2023. Open-World Object Manipulation using Pre-trained Vision-Language Models. (2023).
- [45] Rajesh Subburaman, Dimitrios Kanoulas, Nikos G. Tsagarakis, and Jinoh Lee. 2023. A survey on control of humanoid fall over. *Robotics Auton. Syst.* 166 (2023), 104443.
- [46] Tomomichi Sugihara and Mitsuharu Morisawa. 2020. A survey: dynamics of humanoid robots. *Adv. Robotics* 34, 21-22 (2020), 1338–1352.
- [47] Jingkai Sun, Qiang Zhang, Yiqun Duan, Xiaoyang Jiang, Chong Cheng, and Renjing Xu. 2023. Prompt, Plan, Perform: LLM-based Humanoid Control via Quantized Imitation Learning. *CoRR* abs/2309.11359 (2023).
- [48] Ryo Suzuki, Adnan Karim, Tian Xia, Hooman Hedayati, and Nicolai Marquardt. 2022. Augmented Reality and Robotics: A Survey and Taxonomy for AR-enhanced Human-Robot Interaction and Robotic Interfaces. In *CHI '22*. 553:1–553:33.
- [49] Aaquib Tabrez, Matthew B. Luebbers, and Bradley Hayes. 2020. A Survey of Mental Modeling Techniques in Human–Robot Teaming. *Current Robotics Reports* 1 (2020), 259 – 267.
- [50] Yuichi Tazaki and Masaki Murooka. 2020. A survey of motion planning techniques for humanoid robots. *Adv. Robotics* 34, 21-22 (2020), 1370–1379.
- [51] Yuchuang Tong, Haotian Liu, and Zhengtao Zhang. 2024. Advancements in Humanoid Robots: A Comprehensive Review and Future Prospects. *IEEE/CAA Journal of Automatica Sinica* 11, JAS-2023-1103 (2024), 301.
- [52] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *NeurIPS 2017*. 5998–6008.
- [53] Michael Walker, Thao Phung, Tathagata Chakraborti, Tom Williams, and Daniel Szafrir. 2023. Virtual, Augmented, and Mixed Reality for Human-robot Interaction: A Survey and Virtual Design Element Taxonomy. *J. Hum.-Robot Interact.* 12, 4, Article 43 (2023), 39 pages.
- [54] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. 2022. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. In *NeurIPS 2022*.
- [55] Murphy Wonsick, Philip Long, Aykut Özgün Önel, Maozhen Wang, and Taskin Padir. 2021. A Holistic Approach to Human-Supervised Humanoid Robot Operations in Extreme Environments. *Frontiers Robotics AI* 8 (2021), 550644.
- [56] Yang Ye, Hengxu You, and Eric Jing Du. 2023. Improved Trust in Human-Robot Collaboration With ChatGPT. *IEEE Access* 11 (2023), 55748–55754.
- [57] Fanlong Zeng, Wensheng Gan, Yongheng Wang, Ning Liu, and Philip S. Yu. 2023. Large Language Models for Robotics: A Survey. *CoRR* abs/2311.07226 (2023). <https://doi.org/10.48550/ARXIV.2311.07226>
- [58] Jakub Zlotowski, Diane Proudfoot, Kumar Yogeewaran, and Christoph Bartneck. 2015. Anthropomorphism: Opportunities and Challenges in Human-Robot Interaction. *Int. J. Soc. Robotics* 7, 3 (2015), 347–360.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009