The design of a humanoid robotic research platform

William Duckitt [1*], *João* de Freitas [1], *Connor* Robertson [1], and *Mathias* Smith[1]

[1] Department of Electrical & Electronic Engineering, University of Stellenbosch, South Africa

Abstract. A novel humanoid robotic research platform is presented with a total of 38 degrees of freedom (DOF), incorporating EtherCAT for real-time control, GPU-based high-performance computing for vision, motion tracking and inference, and large language model (LLM) reasoning for voice-based interaction. Building on a torso, two 6-DOF arms, and five-finger hands, the system demonstrates robust manipulation, intuitive voice commands, and extensibility toward bipedal locomotion. Preliminary tests demonstrate stable real-time performance, reliable AI-driven dialogue, and dexterous end-effector function for up to 1 kg loads. This platform aims to serve as a foundation for future research in humanoid robotics.

1 Introduction

Advances in humanoid robotics require the development of systems that not only execute complex, human-like movements but are also cost-competitive and modular for iterative improvement. Recent work emphasizes the integration of robust industrial protocols such as Ethernet for Control Automation Technology (EtherCAT) [1,2,3 and 4] and open-source control frameworks like the Experimental Physics and Industrial Control System (EPICS) [5] or Robot Operating System (ROS) [6]. At the same time, GPU-based computing has accelerated the use of advanced AI methods, including deep learning for vision and large language models (LLMs) for natural-language interaction [7,8 and 9].

Building on prior investigations into modular robotic arms [10] and dexterous robotic hands [11], we present a unified platform that incorporates EtherCAT for deterministic actuation, GPU-based motion tracking and inference, and an LLM-powered voice interface for human-robot communication. By extending the platform to 38 degrees of freedom (DOF) that approximates an upper-torso humanoid and hands, the design can support bipedal locomotion soon.

2 Related work

*William Duckitt: wdduckitt@sun.ac.za

2.1 Humanoid robotics platforms

Humanoid robotics is a subset of robotics where the robots are built to resemble the human form-factor. Unlike traditional industrial robots and other specialised robotic systems that are purpose built to accomplish a specific task, humanoid robots can navigate human spaces and use human-like dexterity to accomplish a wide range of tasks [12]. This versatility has the potential to unlock significant productivity gains, but progress has been limited by the software capabilities. Recent Advances in AI, particularly LLMs and reinforcement learning have helped accelerate progress towards general-purpose robotics [13].

2.1.1 Commercial humanoid robots

The commercial humanoid robots have received substantial investment in recent years leading to multiple different design philosophies. A key divergence is the choice of actuation. Figure AI's 02 and the Boston Dynamics Atlas Electric [14,15] use all rotary actuators which can be more backdrivable and compliant helping with dynamic movements such as walking. In contrast, Apptronik's Apollo and Tesla's Optimus [16,17] have implemented linear actuators for greater force density and stiffness which can help with holding poses and lifting. These platforms represent the cutting edge of humanoid robotics but their propriety hardware and software, as well as the large cost, limit their accessibility for broader research applications.

2.1.2 Academic research robots

In contrast to the closed nature of current commercial humanoids, the academic community has developed more accessible alternatives, to enable more research. UC Berkeley has released an open-source 3D printable humanoid robot which stands at 1 meter tall with the aim of providing a low-cost template for research [18]. Similarly, the TALOS robot is a torque-controlled platform designed for research into safe human-robot interaction and robust locomotion [19]. Additionally, while a commercial product, the Chinese Unitree G1 robot has been frequently used as software development platform due to low cost and developer friendly environment [20].

2.1.3 Comparison with our research platform

Commercial Platforms such at Atlas or Figure 02 offer high-performance operation but as mentioned are proprietary and inaccessible for research. Conversely, academic platforms often lack industrial-grade real-time control and integrated AI capabilities. Our project presents a humanoid robotic research platform which integrates modularity, cost-effective design, deterministic high-speed EtherCAT control and fully onboard AI system for motion tracking and natural language interaction. The combination of these design principles enables a cost-effective platform for humanoid robot AI research. Table 1 shows a comparison of existing humanoid robotic platforms.

Table 1. Humanoid robot comparison [14,18,19].

Our Feature	Our Platform	TALOS	Berkeley Lite	Figure 02
Control System	EtherCAT	ROS	Custom OS	Proprietary Control System
Actuators	RC Servo Motors	Custom Torque Controlled Actuators	Custom Rotary	Custom Rotary
Integrated AI	LLM	No	No	VLM
Accessibility	Open-Design	Purchasable for research	Open-Source	Proprietary
Payload	1kg	6kg	<1kg	20kg

2.2 Humanoid robotic hands

The design of humanoid robotic hands is shaped by various philosophies, often drawing inspiration from bionics and advanced prosthetics. These philosophies involve a trade-off between dexterity, control complexity, and robustness, resulting in different design paradigms. Common solutions include cable-driven systems, which allow for remote motor placement to reduce hand inertia, and more rigid mechanical linkages, which offer greater durability. Additionally, a key design decision lies between high-DOF (degree of freedom) hands low-DOF, under-actuated hands. High-DOF hands aim to replicate the full kinematic characteristics of human dexterity, whilst low-DOF hands use fewer actuators to control multiple joints, enabling adaptive grasping with simpler control schemes. [21]

These design paradigms are evident in several leading platforms. For instance, the Tesla Optimus Gen 2 hand is designed for industrial work, utilising a cable-driven, under-actuated system with 11 DOF and integrated touch sensors [22]. This design balances dexterity with the need for reliable manipulation in repetitive tasks. In contrast, the Psyonic Ability Hand, with its origins in the prosthetics field, uses a robust mechanical linkage system with 6 DOF and incorporates touch sensing [23]. Finally, representing the push for maximum dexterity in research, the open-source DexHand V1 features a complex, cable-driven design with 19 DOF [24]. Its design prioritises the exploration of intricate, human-like manipulation over the durability required for industrial or prosthetic applications. The DexHand project represents the push for maximum dexterity in research, to replicate the full kinematic dexterity of the human hand.

3 System architecture and implementation

Our humanoid robotic research platform, shown in Figure 1, currently features 38 DOF: a head and torso assembly, with two DOF in the neck, two arms each with 6 DOF, and five-finger hands replicated on each side with 12 DOF each.

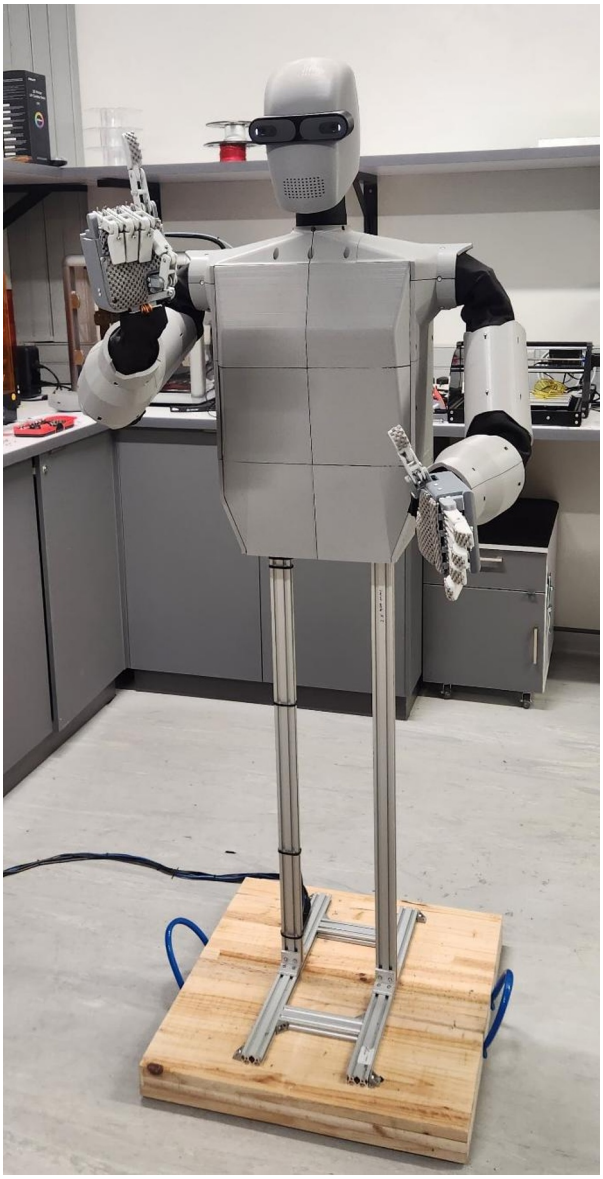

Fig. 1. Assembled humanoid robotic research platform featuring a modular torso, dual 6-DOF arms, and dexterous five-finger hands. The system integrates EtherCAT-based actuation, GPU-accelerated perception, and voice-based VLM interaction.

3.1 System design

This system design represents the convergence and evolution of two earlier subsystems developed at Stellenbosch University. The robotic arms were designed by Robertson [10] as modular, 6-DOF limbs using PG20 aluminium extrusions and 3D-printed bracket joints, optimized for human-like articulation. A photograph of the original 6 DOF arm with pincer

end effector is shown in Figure 2. Aluminium extrusions and 3D-printed brackets and covers provide a modular design that eases maintenance and rapid reconfiguration.

In parallel, Smith [11] developed a five-fingered tendon-driven robotic hand capable of performing multiple human grasp gestures. The hand design utilized a hybrid cable-linkage mechanism with palm-integrated routing, allowing for independent flexion of each digit. The system employed visual tracking to match hand gestures and demonstrated high-fidelity replication of common grasps without gloves or physical markers. The complete mechanical hand assembly is depicted in Figure 3.

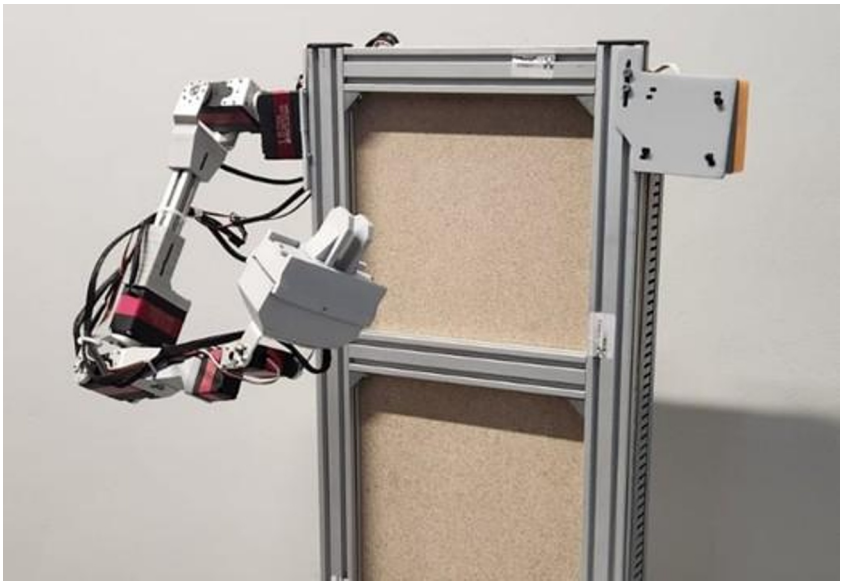

Fig. 2. Modular 6-DOF robotic arm with a pincer end effector, designed by Robertson. Constructed from PG20 extrusions and 3D-printed joints, it demonstrates structural modularity and human-like articulation.

Fig. 3. Five-fingered tendon-driven robotic hand developed by Smith. The design incorporates a hybrid cable-linkage system for digit articulation and demonstrates precise execution of human-like grasp gestures. Shown from left to right, is the top and bottom view of the right hand and the bottom and top view of the left hand.

The 6 DOF arm with gripper end effector was reproduced for both the left and right sides of the humanoid robot to create a symmetric, dual-arm configuration capable of performing bimanual tasks. In transitioning from the prototype to the integrated humanoid platform, the original pincer-style end effectors were replaced with tendon-driven robotic hands on each

side to enhance dexterity, enable more natural human-like interaction, and support a broader range of grasp types.

The torso of the prototype was designed with modularity in mind, utilising the aluminium extrusions seen in the 6 DOF arm. This allowed easy assembly and disassembly of the robot as well as attaching of components such as power supplies. The 3D printed brackets were redesigned to fit the proportions of the human body and to provide attachment points for the control electronics and covers. The covers were designed to protect the wiring and electronics housed in the arms. The covers for the rigid sections of the upper-limb and forearm were 3D printed, and the joints utilised a flexible mesh fabric to allow the joints to move without issues.

The head was designed to be modular with 2 DOF. It features two cameras for stereo vision and a mini USB speaker for audio output. The entirety of the head can be assembled with ease, with each part able to slide into each other and held together with self-tap-in screws. The neck servos allow for pitch and yaw rotation, and a printed mount goes onto the servo for easy head dismounting and maintenance.

Additionally, the control systems for both the arms and hands were upgraded from the initial microcontroller-based architecture to a unified EtherCAT system,

3.2 Control system

The robot's control system is designed on the EtherCAT industrial control standard. A block diagram depicting the interconnection of the subsystems of control system is shown in Figure 6.

A central computing unit in the torso hosts the EtherCAT master and an NVIDIA GPU, allowing millisecond cycle times to coordinate servo updates across the arms and hands. Each servo node on the network interfaces through EtherCAT for deterministic control. The GPU simultaneously handles vision tasks, LLM inference in real time, offloading these computationally heavy processes from the main CPU. Power is managed centrally and is distributed from the torso to each the limbs and head.

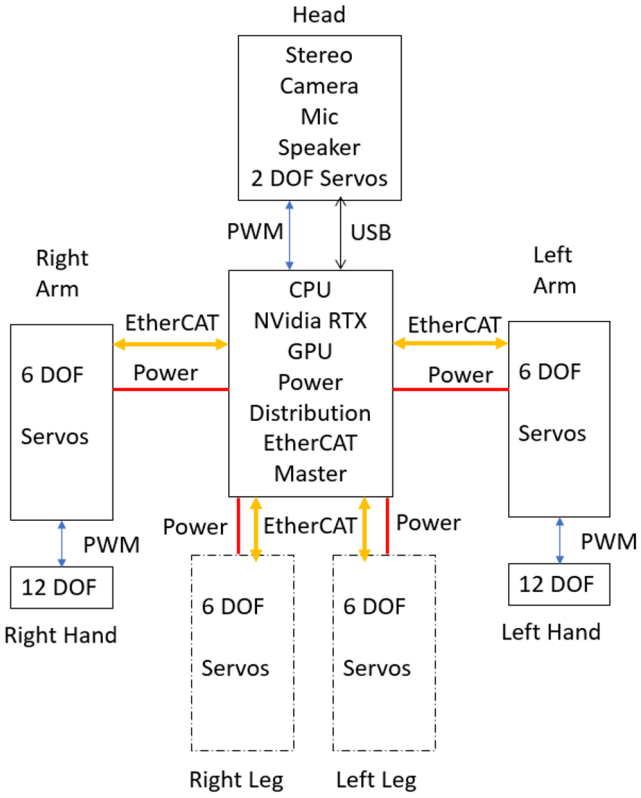

Fig. 4. Control system-level architecture of the humanoid robot platform. The CPU and GPU core handle vision, inference, and EtherCAT master control, interfacing with all limbs and the head. Right and left arms, hands, and future legs are shown.

The open source EtherCAT master [25] runs in Ubuntu Linux. Python and EPICS are used in the backend for control of the robot actuators and React-Automation-Studio [26] is employed as the frontend for all engineering user interfaces, offering browser-based access to live telemetry, motion tracking, actuator diagnostics, and system configuration tools.

The control system software is containerized using Docker and structured into modular microservices for scalability, maintainability, and isolation. Services include EtherCAT control, camera acquisition, vison language model inference, speech-to-text text-to-speech, and telemetry logging. These services are orchestrated using Docker Compose, with tailored deployment modes for development, testing, production operation on the physical robot, and simulation environments. This architecture allows seamless transitions between simulated and physical robots, while maintaining consistent interface and behaviour across platforms.

3.3 Voice-Enabled LLM Reasoning

The voice interaction subsystem integrates a vision-language model (VLM), specifically Gemma 3:4B [27], which enables multimodal reasoning by accepting audio, visual, and textual input streams.

The system includes a microphone for speech recognition, a stereo camera setup for visual context, and speakers for natural language responses. Speech-to-text (STT) conversion is performed on-device, triggering semantic parsing within the VLM. Commands are inferred through contextual grounding of both voice input and visual feeds—enabling interaction with dynamic environments. The model's tool-calling capability enables real-time invocation of robot control functions. This allows voice-initiated control of motor groups, gesture commands, or task sequencing using symbolic logic grounded in the visual state. The platform supports bidirectional dialogue using text-to-speech (TTS) for response synthesis.

The TTS, STT and VLM runs locally as Docker microservices on the same GPU, maintaining concurrent inference across modalities. Integrated tests confirmed high responsiveness to natural user commands, with robust understanding of compound and referential instructions in the presence of scene context.

3.4 Real-time motion tracking

For functional verification of the robot systems, an integrated pose tracking system was developed in React-Automation-Studio.

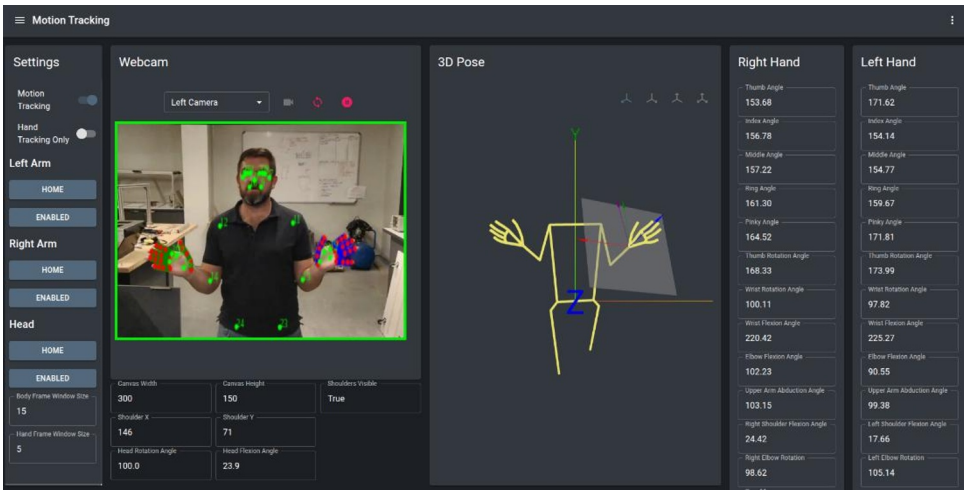

Fig. 5. The React-Automation-Studio based real-time motion tracking and inverse kinematics system.

The system shown in Figure 5, employs MediaPipe [28] to perform to perform pose estimation for the upper torso and hands. The inverse kinematics of the 3D pose is calculated, and the angles are mapped to each actuator. The source of camera feed can be controlled between the robot's vison cameras, a remote camera feed or a video. This allows the robot to be put into follow mode and track a subject in real time, whilst diagnostics and system performance is track allowing for simultaneous control system optimization. This system also serves as the basis for future work to teach the robot in real-time how to perform various tasks.

4 Experimental results

Our humanoid robotic research platform currently features 38 DOF: a head and torso assembly, with two DOF in the neck, two arms each with 6 DOF, and five-finger hands replicated on each side with 12 DOF each.

4.1 Real-time motion tracking

The DOF of the design is intended to approximate human upper-body dexterity, with future expansions planned for bipedal legs. The functioning of the upper body dexterity has verified using an inverse kinematics motion tracking. Various poses where the robot mimics the operator are shown in Figure 5.

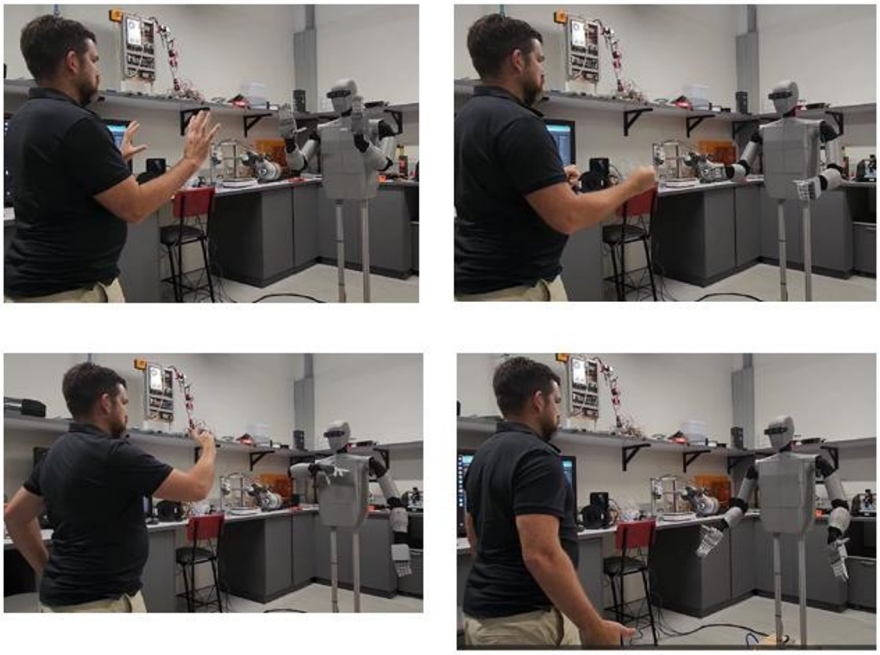

Fig. 6. Demonstration of real-time pose imitation by the humanoid robot. The robot mirrors the operator's hand gestures—including open palm, closed fist, running pose, and relaxed stance.

4.2 Hand results

4.2.1 Hand dexterity

The current version of the 12 DOF hands perform adequately and can perform various grasp types, including pinch, cylindrical, and spherical. In Figure 7, we demonstrate the robots left hand performing grasps including l lateral tripod grip with a puck, pen grasp, and cylindrical grip with a PVC pipe, and spherical grasp with ball, and an index finger extension grip with a metal rod.

Dexterity is not the only feature that makes the human hand unique – the speed of finger flexion and extension is also important. Testing the speed of the 12 DOF will verify whether the hand operates at human-like speeds. The test includes measuring the time moving from full extension to full flexion, and then vice versa in a separate test. The results are shown in

the table below. The results verify that the 12 DOF hand operates at near human-like speed. The closing time of the robotic hand (260ms) might seem incomparable to the human speed (190ms), but considering the time frame of the results, the speed is near identical to the human eye. Therefore, the test verifies the human-like speed of the 12 DOF hand.

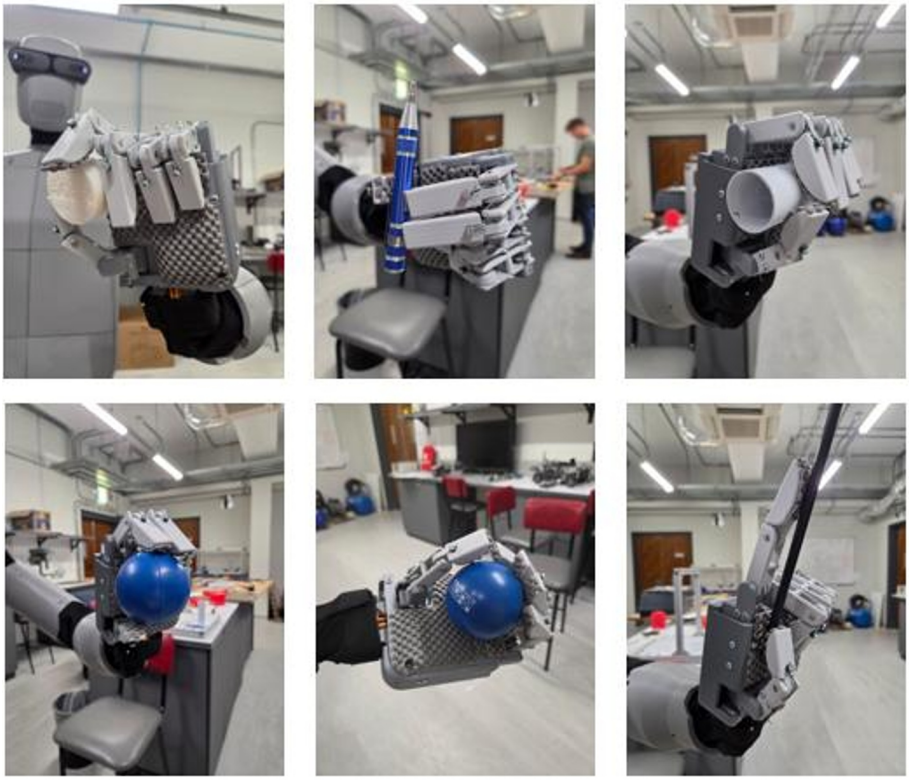

Fig. 7. Dexterous hand grasp demonstrations. The robot performs precision and power grasps including (top row, left to right): tripod grasp with a ball, thumb 2 finger pinch (pen grasp), and medium wrap with a PVC pipe (cylindrical grip); and (bottom row, left to right): tripod grasp with a ball, power sphere grasp with a ball, and an index finger extension grip with a metal rod.

Table 1. Finger speed test.

	Human	Robot
Closing Time (ms)	190	260
Opening Time (ms)	200	200

4.2.2 Hand strength

The strength of the hand was evaluated by conducting two tests. The first test determines the finger flexion load capacity by loading the finger with a mass at full extension and then observing the flexion. The second test determines the combined lift capacity of the four fingers by performing an inverted fist lift test, also known as a briefcase lift test. The results are shown in Table 3 and Table 4.

Table 2. Finger flexion load capacity test.

	Stall Torque (kgf.cm)	Rated Torque (kgf.cm)	Stall Mass (g)	Rated Mass (g)
Theoretical	1.599	0.3202	592.6	118.6
Actual	1.083	0.3510	401.0	130.0

Table 3. Inverted fist lift capacity test.

	Lift Capacity (g)
Theoretical	2370
Actual	2225
Target	2000

The results from the load capacity tests show satisfactory results. Deviation between theoretical and actual results are most likely due to tension loss in the cables connected to the servo motor pulleys and finger phalanges, as well as the servos not accurately representing their electromechanical specifications. Table 4 still shows the actual lift results exceeding the target capacity of 2000g. Therefore, the 12 DOF hand meets the 2000g target lift capacity specification.

4.3 Payload capacity

Preliminary tests confirm that the robotic arms demonstrate full human-like movement with a payload capacity consistent with design specifications. The robot arms can each lift objects weighing up to 1 kg at full extension of each joint, which is sufficient for lightweight tools and demonstration objects in research and teaching environments. However, the 3D printed brackets, lack the rigidity for accurate heavy lifting. Replacing the current 3D printed brackets with machined aluminium parts, will result in improved lifting accuracy [10]. Additionally, the current actuators will need to be replaced with high-torque servos or custom-designed actuators that maintain EtherCAT compatibility, meet thermal efficiency requirements, and deliver consistent torque under load without introducing backlash or compliance into the motion system.

4.4 Repeatability

To test the robotic arm's repeatability was measured. The robot was continuously picked up and placed a piece of extrusion onto a box. The first attempt position was marked as the intended location and thereafter each placement was marked and the distance from the correct location was measured. Performing 6 successive pick and place tasks, the maximum error was recorded to be 2.25mm with an average error of 1.8mm across the 6 trials [10]. Given the intended use of the robot, the repeatability was acceptable. As previously mentioned, the lack of rigidity of the 3D printed parts would cause a larger error with differing loads.

5 Future work

Future work will include increasing the payload capacity to 10 kg by redesigning joint brackets with machined aluminium parts and selecting higher-torque actuators to replace the existing servos. In parallel, custom in-house designed actuators will be developed to optimize both weight and torque density while ensuring EtherCAT compatibility and modular integration. The dexterity of the hands will be expanded from 12 DOF to 21 DOF, enabling independent articulation of all finger joints and thumb opposition, which is essential for complex object manipulation and tool usage. Additionally, we plan to develop and integrate bipedal legs to extend the platform's mobility and balance. The leg design will incorporate hip, knee, and ankle joints with appropriate torque-to-weight ratios and closed-loop feedback, allowing for dynamic locomotion, balance correction, and gait planning.

Furthermore, we aim to explore and integrate vision-language-action (VLA) models to enable the robot to autonomously interpret natural language commands in the context of its visual environment and execute multi-step actions. These models will be used to support task planning, real-time adaptation, and general autonomous control in unstructured settings. To enable robust and scalable behaviour, we also plan to implement reinforcement learning strategies for whole-body control, trained initially in simulation environments. Sim-to-real transfer techniques will then be applied to bridge the domain gap, allowing policies learned in simulation to generalize effectively to real-world robot operation. These capabilities are essential for developing a robot that can adaptively interact with dynamic environments and learn new skills over time.

6 Conclusion

We have developed a 38-DOF humanoid robot platform combining EtherCAT for deterministic actuation, GPU acceleration for real-time AI tasks, and an LLM-based voice interface for natural language interaction. This design enables dexterous arm and hand control, robust dialogue-driven commands, and, eventually, bipedal walking. Future work will incorporate multi-sensor fusion, force feedback and increased DOF in the hands, increase payload capacity, custom actuator integration and advanced gait control for dynamic locomotion as well as an exploration into state-of-the-art vision language action and reinforcement learning control of the platform. Our platform serves as a foundation for research into sophisticated, human-like robots capable of complex manipulation and fluid communication in real-world scenarios.

References

1. PC-Control: ”Rollin’ Justin” robot drags crates and serves tea. https://www.ethercat.org/download/documents/pcc_0210_dlr_e.pdf. (Accessed on 05/7/2024).
2. W. June et al. “Dual Channel EtherCAT Control System for 33-DOF Humanoid Robot TOCABI”. In: IEEE Access (2023).
3. F. Sygulla et al. “An EtherCAT-Based Real-Time Control System Architecture for Humanoid Robots”. In: 2018 IEEE 14th International Conference on Automation Science and Engineering (CASE). 2018, pp. 483–490. Doi: 10.1109/COASE.2018.8560532.

4. B. Fiebiger and H. Munz. KUKA KR C4 robot controller uses EtherCAT. https://www.pc-control.net/pdf/012014/solutions/pcc_0114_kuka_e.pdf.
5. EPICS - Experimental Physics and Industrial Control System. https://epics-controls.org/
6. ROS: Home. https://www.ros.org/.
7. Ollama. https://github.com/ollama/ollama
8. H. Touvron et al. “LLaMA: Open and Efficient Foundation Language Models.” https://arxiv.org/abs/2302.13971.
9. M. Kim et al. “OpenVLA: A Vision-Language-Action Platform.” https://arxiv.org/abs/2406.09246.
10. C. Robertson, “Design of a Modular Humanoid Robotic Arm,” Stellenbosch University, 2024.
11. M. Smith, “Humanoid Robotic Hand,” Stellenbosch University, 2024.
12. B. Adams, C. Breazeal, R. A. Brooks and B. Scassellati, "Humanoid robots: a new kind of tool," in *IEEE Intelligent Systems and their Applications*, vol. 15, no. 4, pp. 25-31, July-Aug. 2000, doi: 10.1109/5254.867909
13. Jeong, H.; Lee, H.; Kim, C.; Shin, S. A Survey of Robot Intelligence with Large Language Models. *Appl. Sci.* 2024, *14*, 8868. https://doi.org/10.3390/app14198868
14. Figure AI. https://www.figure.ai/.
15. Atlas and beyond: the world's most dynamic robotshttps://bostondynamics.com/atlas/
16. Apollo. https://apptronik.com/apollo
17. A Complete Review of Tesla’s Optimus Robot. https://briandcolwell.com/a-complete-review-of-teslas-optimus-robot/
18. Chi, Yufeng, et al. "Demonstrating Berkeley Humanoid Lite: An Open-source, Accessible, and Customizable 3D-printed Humanoid Robot." *arXiv preprint arXiv:2504.17249* (2025).
19. O. Stasse *et al*., "TALOS: A new humanoid research platform targeted for industrial applications," *2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*, Birmingham, UK, 2017, pp. 689-695, doi: 10.1109/HUMANOIDS.2017.8246947.
20. Unitree G1. https://www.unitree.com/g1
21. H Yang et al. “An Affordable Linkage-and-Tendon Hybrid-Driven Anthropomorphic Robotic Hand – MCR – Hand II”. In: Journal of Mechanisms and Robotics. DOI: 10.1115/1.4049744 https://pmc.ncbi.nlm.nih.gov/articles/PMC8558225/#:~:text=(2015)%2C%20the%20authors%20used,conform%20to%20the%20grasped%20object.
22. Tesla Optimus Gen 2 Review. https://blog.robozaps.com/b/tesla-optimus-gen-2-review#:~:text=This%20level%20of%20precision%20is,that%20rival%20the%20human%20hand.
23. Psyonic. https://www.psyonic.io/robots
24. DexHand. https://www.dexhand.org/
25. R. Mercado, I. J. Gillingham, J. Rowland, and K. G. Wilkinson, “Integrating EtherCAT based IO into EPICS at Diamond”, in Proc. 13th Int. Conf. on Accelerator and Large

Experimental Control Systems (ICALEPCS'11), Grenoble, France, Oct. 2011, paper WEMAU004, pp. 662–665.

26. W. Duckitt and J. K. Abraham, "React Automation Studio: A new face to control large scientific equipment", presented at the 22nd Int. Conf. on Cyclotrons and their Applications (Cyclotrons'19), Cape Town, South Africa, Sep. 2019, paper THA04.
27. A. Kamath et al., Gemma 3 Technical Report. 2025. [Online]. Available: https://arxiv.org/abs/2503.19786
28. C. Lugaresi et al., MediaPipe: A Framework for Building Perception Pipelines, 2019 [Online]. Available: https://arxiv.org/abs/1906.08172