

Design of a Butler-Type Android Employing State-of-the-Art AI Techniques

Roumiana Ilieva

Department of Management and Business Information Systems
Faculty of Management
Technical University of Sofia
Sofia, Bulgaria
rilieva@tu-sofia.bg

Antoni Angelov

Department of Management and Business Information Systems
Faculty of Management
Technical University of Sofia
Sofia, Bulgaria
antonianangelov@tu-sofia.bg

Abstract— This paper is about the possible and likely soon creation of a functional Butler android via available AI technology and methods. “Available” in this case refers to as publically accessible to the general public. There are 2 types of AI abbreviated types of control mentioned in this paper and one of them could even easily be set up today. The more functional and also unrestricted would require additional training and possible training of additional models. It can also benefit of distillation of combined AI projects such as ChatGPT and Computer Vision AI. For the body, both the necessary kinematics and functionality already exist. There may be some gap in design optimality but the currently existing designs can be considered “Good enough”. Some are also publically available and several companies are already mass producing robot bodies even if the software is yet to be perfected. The paper goes through the basics and the architecture of achieving the control.

Keywords—AI, Artificial General Intelligence, Neural Network, AI Distillation, VR, Housekeeping, rudimentary tasks. Butler, Android;

I. INTRODUCTION TO THE PROJECT

This project is written within the currently available technologies of at the date it is written. It is not Sci Fi and thus the end results should be within realistic expectations. Then again, a walking talking butler android who can carry things around is still ahead of current normality, yet it is fully achievable. The real magic is the core around which the technology runs through and that actually happens to be ChatGPT. For some it may be a simple chat bot, for others a substitute to Google. The opportunities it provides however, and its actual capabilities actually go far, far beyond that. The rest of the parts are also AI projects we have seen implemented and publically tested in popular video format. Unfortunately, other than the animation of some plastic toys, they have yet to see proper use. At least according to what the public is allowed to see.

II. THE AIS OF THIS PROJECT

A. ChatGPT, more than a chatbot

First we start with the heart of the project, the Narrator. The way we would imagine using a neural network for this type of project and the way we actually do

are very different. Normally the Neural Network would be the brain and its end nodes would control the limbs of

the robot, and the Neural Network would grow and change and have memories and maybe even feelings. Science fiction is beautiful. Reality, close but not so much.

Instead of the Neural Network controlling the robot, it controls a story. A story of a humble robot butler that would never hurt his/her clients and follow their every command to the best of his abilities.

In this use the Neural Network is immutable and it's the characters within its story that are by design and whim of the owner of the hardware. Switching the Neural Network with a better one will not change the characters of the story significantly, although better writing is always welcome. Luckily, one of the things ChatGPT was fed during creation was book novels with all sorts of interesting made up characters. If you would ask ChatGPT to role play as one or several you would be amazed how deep, and realistic it can go.

B. The role of ChatGPT

ChatGPT plays the role of a narrator, controls the story and follows the input of the user. That would be great for a video game character and there are implementations of that, such as adding Darth Vader to the popular game “Fortnite” [1]. The character is incredibly responsive and acts strongly within parameters. In fact, it will keep the conversation within its comfortable zone and will ignore any attempts to be brought out of character as attempts of meaningless distraction.

Writing a character for ChatGPT is as simple as writing the rules an actor has to follow to get into character. It is done by writing a free text explaining what his role is, what he should do at certain conditions and so on. If there is any formatting, it's for the user not for ChatGPT. One of the uses for ChatGPT is tech support or someone to answer phone calls. It is done by combining Voice to text, ChatGPT and Text back to voice. It's end user will have a hard time detecting whether they talk to a real person or ChatGPT. Especially since the Text to Voice is made via a

Now to bring it out of virtual world and into reality, to do so we need to first understand what ChatGPT actually is. With all its intellect and versatility, it is still just a predictive algorithm that predicts the next word in a sentence. It's basically blind, deaf and can only see and communicate via text.

C. Necessary system for the creation of a butler android

How to bring a story to real life? If the only communication is via text, one of things the system would require is computer vision AI that will track the world through cameras and describe it in detailed text so that ChatGPT has an understanding of the location environment of its characters. There are many pre-trained Neural Networks for this task that we can chose from.

Vision AI

Online Demo

API Docs

Upgrade

Detect Objects:

astica GPT-4 Description

A man is standing in a grocery store, wearing a blue apron and a mask, working on a laptop at a counter. The store has various products displayed on shelves, including a vending machine, chairs, desks, and cabinets. The man appears focused on his work, surrounded by the bustling atmosphere of the store. The store is well-lit, with white walls and a clean, modern aesthetic. The man is wearing a blue apron, which is a common sight in grocery stores. The laptop is open on the counter, and he is looking at the screen. The shelves are filled with various products, including a vending machine, chairs, desks, and cabinets. The man is standing in the aisle, and the store is well-lit.

asticaVision - Image AI

asticaVision AI Output

Image Caption:
A Man Is Standing In A Store With A Counter.

[Detect Image Caption](#)

Facial Recognition:

No faces detected.

Object Detection:

Cabinet	28.00%	
Chair	29.00%	
Cabinet:shelf	51.00%	
Cabinet:shelf	37.00%	
Cabinet:shelf	43.00%	
Cabinet:shelf	43.00%	
Cabinet:shelf	43.00%	
Cabinet:shelf	48.00%	
Chair	30.00%	
Desk	58.00%	

Once the ChatGPT is no longer blind it will have an understanding of the environment around it and thus act in a responsive to the environment manner. Combined with Speech to text for the user and if it's able to recognize the source of the speech, it will have sufficient integration of the real world to its story line. Of course it will have to be extremely detailed and with explanation of how far each object and where exactly is located to the unit that is being controlled. It may be common sense that one server running ChatGPT may run several androids at a time and their storyline may even be combined.

On this question, the answer would be through stacking more AI projects. There are a few more projects that can be added to the system to complete the cycle. While the visual scanning may be great to describe the environment it is not enough to provide information needed for the mobility of the unit. For such task terrain recognition such as on figure 3. will be required.



The software created for SpotMini of Boston Dynamics is perfect for our project as well. Of course with certain adjustments. What it does is map the terrain around the unit creating and underlying a path through which the robot can step on. Whether it's a bipedal or 4 legged robot is of no consequences as long as it can keep its balance. It should also understand what material is it stepping on and whether it can step on it or not, but that's up to the computer vision that was previously described. Here it's all about geometry.

Additionally, a proper training algorithm can be created and trained via a neural network such as Gymnasium [4] as seen on Figure 4.

Humanoid



This environment is part of the Mujoco environments which contains general information about the environment.

Action Space	Box(-0.4, 0.4, (17,), float32)
Observation Space	Box(-inf, inf, (348,), float64)
import	gymnasium.make("Humanoid-v5")

Figure 4. Training environment for walking

The environment can train a neural network, not only to walk on an even surface, but also one with bumps and other obstacles and even self-correct when pushed by wind or other unit or person. A functional body with this type of neural network for versatile mobility can be called a Drone. Of course the word is already used to describe the flying drones but that's only because their functionality already allows for free mobility, thus making flying unironically easier to achieve than walking.

D. Summary and completion

So far the butler robot can walk, talk, understand a conversation and recognize people and objects. However, can it pick them up, can it manipulate the environment and can it do complicated tasks? Not yet.

The computer vision for walking can be further worked on to separate objects from one another and with computer vision for analysis to label them. That way it will not be just a 3D environment in which it can only walk around but it may be possible to manipulate it.

Using a pre-trained Neural Network to train another is called Distillation and that is exactly what is required to create the final piece of the puzzle. It will also require a large amount of videos that will be analyzed, both as tasks that the people in the video are doing. Skeleton meshes will be recognized of the people in the video thus giving suggestion of how the butler robot could move implementing the same movements and thus doing the same task. With enough material a fully functional butler

android may be achievable soon. However, for simple tasks such as grabbing things and carrying them such additional training may not be required. And carrying drinks and being there for show can be a sufficient use of the technology for today.

There is also an alternative that can provide in depth video data of the most common tasks, a butler android may be required to do and thus accelerating significantly the project. And on top of that it can provide the same functionality with no additional research required.

III. THE ALTERNATIVE

A. AI: Actual Indians [5]

With the already implemented software and pre-trained neural networks all that's left is the manual tasks which can be done by people with VR controllers who don't speak English and live in economy realms where their salaries are not difficult to provide. Walking, talking can be handled by the Artificial Intelligence. The manual tasks that are still difficult for the AI can be done by these people who are monitored by the AI and have signed NDAs and work for reputable local companies. Most people who would require the type of service of a butler could have already given that job to a human before and thus concern about privacy may not be in the highest regard. Never the less there are ways to ensure that privacy is respected and no harm is done. Just because it is humans, it alone should not be an issue. They will only follow written commands and be constantly monitored by AI at all time.

An example is on Figure 5.



Figure 5. XR + humans, a sufficient alternative to AI (for now)

The hand controller or hand detection is a great and accurate control mechanism for the upper body of the androids. The lower body can be moved via buttons or by stepping onto a designated circle around the user. Step forward will have the android walk forward until the user steps back in the middle and so on.

This way the humans can both provide the necessary data for the training of the AI in doing the manual tasks and can provide the completion of said tasks in the meantime. The XR users will not talk to the owners of the androids and

only do tasks written to them by ChatGPT as subtitles, usually as simple easy to understand commands. Communication is handled by the AI. Where to go can be handled by giving the XR personnel arrows to follow and thus completing the cycle. Recording devices will obviously be forbidden in the XR studio and with a signed NDA. With that alone the level of privacy is already higher than hiring a cleanup crews or butler / maid on salary.

IV CONCLUSION

Understanding the technology on the inside can help us understand its limitations on the outside. And they are currently many. Thus a work around is required. Using XR technology for tasks such as Dark Factories or in this case Butler / Maid androids may soon become a very common thing. XR is a great way to turn manual labor that previously was solely done in person into a remote work, that can even eventually be even replaced by AI. It may feel a bit negative for the worker to be replaced by AI and have to change job, but the benefit to the economy can make their life and of everyone around them better and thus it's not a worthless sacrifice. Especially if it's a dirty job that nobody really likes such as cleaning stuff. A better economy allows for pursuit of a job of passion instead of necessity. Yet, there are jobs that "Someone has to do" and if those go to the AI, the world will become a slightly better place.

ACKNOWLEDGMENT

ACKNOWLEDGE THE FINANCIAL SUPPORT OF THE PROJECT WITH FINANCING AGREEMENT No. PVU-44 OF 05.12.2024 UNDER PROJECT NO. BG-RRP-2.017-0011 "ECOLOGICAL COLLABORATIVE ROBOTS POWERED BY GREEN HYDROGEN" UNDER THE RECOVERY AND RESILIENCE MECHANISM FOR THE IMPLEMENTATION OF AN INVESTMENT UNDER C2I2 "INCREASING THE INNOVATION CAPACITY OF THE BULGARIAN ACADEMY OF SCIENCES (BAS) IN THE FIELD OF GREEN AND DIGITAL TECHNOLOGIES" FROM THE RECOVERY AND RESILIENCE PLAN, BULGARIA.

REFERENCES

- [1] Fortnite.com, May 16, 2025, <https://www.fortnite.com/news/this-will-be-a-day-long-remembered-speak-with-darth-vader-in-fortnite> (last visited 19.10.2025)
- [2] Computer Vision AI Describe and Caption, <https://astica.ai/vision/describe-images/> (last visited 19.10.2025)
- [3] Boston Dynamic's Dog Robot Can Now Run Autonomously, Interestingengineering.com <https://interestingengineering.com/videos/boston-dynamics-dog-robot-can-now-run-autonomously> (last visited 19.10.2025)
- [4] Gymnasium Documentation, <https://gymnasium.farama.org/environments/mujoco/humanoid/> (last visited 19.10.2025)
- [5] David Braue, The company whose 'AI' was actually 700 humans in India, Jun 05 2025, <https://ia.acs.org.au/article/2025/the-company-whose-ai-was-actually-700-humans-in-india.html> (last visited 19.10.2025)