AMERICAN SOCIETY FOR ENGINEERING EDUCATION

Paper ID #49607

Design of a Greeting Robot

Caleb Edward Scheideger, Ohio Northern University

Caleb Scheideger is a mechanical engineering student at Ohio Northern University with interests in bio-medical research.

Evan Louis Fisher, Ohio Northern University

Evan Fisher is a senior undergraduate student at Ohio Northern University as part of the Department of Mechanical Engineering. He plans to finish his B.S. in Mechanical Engineering and Mathematics in Spring 2025 and continue his education following graduation. He has research interests in solid mechanics, fluid mechanics, and heat transfer, specifically in computational methods.

Kaden Wince, Ohio Northern University

Kaden Wince is currently pursuing the bachelor's degree in computer engineering with Ohio Northern University, OH, USA, where he is currently a senior. His research interests are focused on the robotics field, including manipulation, control, and planning algorithms in dynamic environments.

Emma Pond, Ohio Northern University

Emma Pond is a senior undergraduate student at Ohio Northern University, soon to finish a B.S. in Mechanical Engineering. She currently works as a teaching assistant for the Department of Mechanical Engineering at Ohio Northern University.

Maxwell Phillips, Ohio Northern University

Maxwell Phillips is a senior undergraduate student at Ohio Northern University, soon to finish a B.S. in Computer Science. He currently works as a research assistant for the Department of Electrical and Computer Engineering and Computer Science at Ohio Northern University. His current research focuses on high-precision hardware design, particularly for integer arithmetic and encryption.

Abigail Collins, Ohio Northern University

Abigail Collins is an senior at Ohio Northern University, majoring in Engineering Education. She is interested in the best Engineering Process education methods and plans to teach Engineering at the high school level following graduation.

Rhys Francis Davies, Ohio Northern University

Rhys Davies is a senior undergraduate student at Ohio Northern University, soon to finish a B.S. in Electrical Engineering. He currently works part-time as an Automation Engineering intern at Crown Equipment Corporation and has accepted an offer to go full-time in June. He is interested in researching automating software development for speeding up the development process.

Design of a Greeting Robot

Introduction

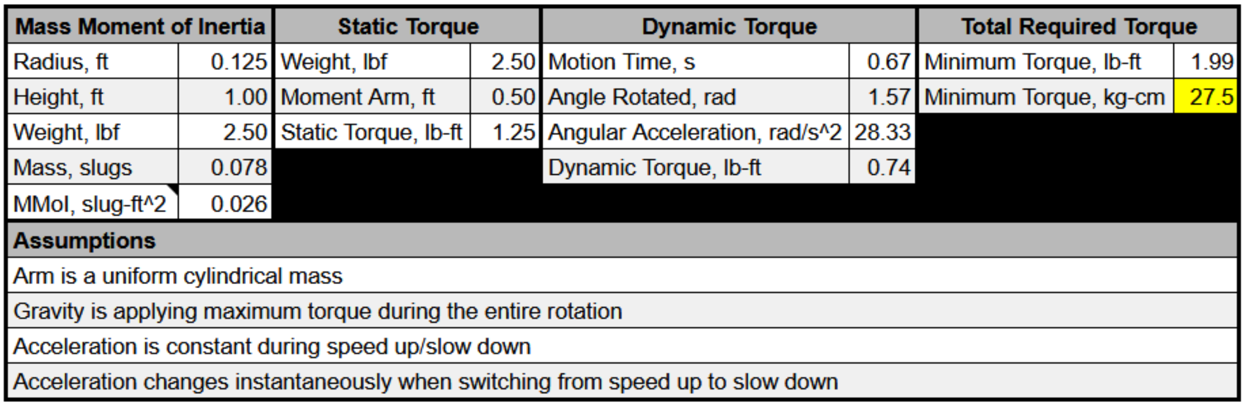
The Greeting Robot is a collegiate-level capstone design project aimed at developing a stationary, humanoid robot capable of interactive engagement. The project consists of designing and building a robot that recognizes and responds to verbal interactions, answers questions about the college it represents, performs simple gestures such as waving, and detects and mimics the movements of a nearby individual. The objectives of the project are to enhance outreach efforts by showcasing its represented college to prospective students and families, demonstrating the technical capabilities of current students, and inspiring interest in STEM fields.

The project is characterized by five main design factors: **(1)** motor and power transmission, which drives limb and head movement. This is achieved through compact, efficient, servo motors which will generate the torque required to drive the limbs smoothly and safely. **(2)** structural design, which has a crucial role in the assembly of the robot's physical structure, specifically the head, arms, and torso. The torso skeleton is constructed from lightweight but durable material such as aluminum, while the head, arms, and other aesthetic components are primarily 3D printed to reduce weight. (**3**) establishing electrical power and communication for control and coordination across all system components. DC power supplies and the Universal Asynchronous Receiver Transmitter (UART) protocol keep the motors, controllers, and sensors powered and synchronized for smooth limb movement. **(4)** human pose estimation, allowing the robot to track and imitate user movements. The robot uses a camera and extensive modeling software to recognize the position and movement of a person and map that to a set of points in real-time. This is then used to generate a series of joint angles to position the robot's limbs. **(5)** an artificial intelligence "pipeline" to power the robot's speech processing. The robot listens to an audio input and converts it to text using a speech-to-text (STT) model, then uses a large language model (LLM) to produce a dynamic response, and finally uses a text-to-speech (TTS) model to verbally respond to the user naturally. These core design challenges, when implemented cohesively, provide a fully functional, interactive greeting robot. This paper includes a detailed description of each of the five design factors, a summary of the process of designing the robot, and concluding remarks.

1. Motor and power transmission

The motor and power transmission subsystem is an integral part of the robot's movement, enabling precise, smooth, and reliable operation. It powers the motion of the robot's head and arms and allows the robot to perform gestures such as waving and mimicking human actions. In order to achieve these functional aspects, appropriate motors were selected and paired with efficient power transmission mechanisms. The chosen motors needed to provide high amounts of torque while remaining compact, lightweight, and efficient. After evaluating several different motor options, it was decided that servo motors would be best suited for this application due to their precision, durability, and high torque-to-mass ratio[1]. In order to size the servos, worst case static and dynamic loading conditions were assumed and used to estimate the maximum required torque. The calculations shown in Table 1 show 27.5 kg-cm was required for servos located in the shoulder. From this estimate, 30 kg-cm servos were selected to be used in all joints, noting that the assumptions made make this estimate an upper limit. The placement and number of motors were determined based on the desired degrees of freedom (DOF) for the robot. The objective was to minimize the overall weight of each limb while maintaining enough DOF for the robot to accurately imitate human movements. The final motor configuration, shown below in Figure 1, has seven total degrees of freedom: four in each arm and one in the head. This design reduces the number of motors required in each arm while maintaining a balance between minimizing weight and providing the necessary range of motion for realistic human mimicry.

Table 1: Servo sizing calculations based on worst case static and dynamic loading conditions.



Mass Moment of Inertia		**Static Torque**		**Dynamic Torque**		**Total Required Torque**	
Radius, ft	0.125	Weight, lbf	2.50	Motion Time, s	0.67	Minimum Torque, lb-ft	1.99
Height, ft	1.00	Moment Arm, ft	0.50	Angle Rotated, rad	1.57	Minimum Torque, kg-cm	27.5
Weight, lbf	2.50	Static Torque, lb-ft	1.25	Angular Acceleration, rad/s^2	28.33		
Mass, slugs	0.078			Dynamic Torque, lb-ft	0.74		
MMoI, slug-ft^2	0.026						

Assumptions
Arm is a uniform cylindrical mass
Gravity is applying maximum torque during the entire rotation
Acceleration is constant during speed up/slow down
Acceleration changes instantaneously when switching from speed up to slow down

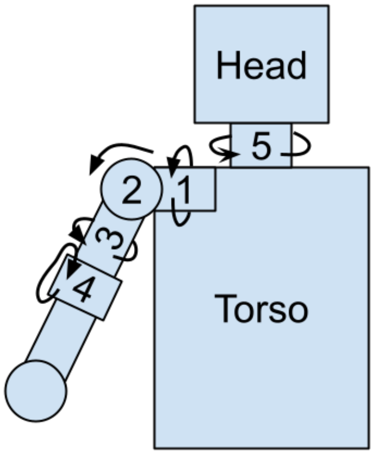


Figure 1: Kinematic motion and motor configuration.

With the motors and their configuration determined, the physical structure of the robot can be designed.

2. Structural design

The physical structure of the robot provides the framework that will house all of the mechanical and electrical components. The design must balance strength, weight, and visual appeal to create a robot that is both functional and engaging for those interacting with it. The core of the robot's structure will be a lightweight, modular frame made from durable materials such as angled aluminum along with various 3D printed components. The chosen materials will provide strength and durability, allowing the robot to withstand regular use and support the dynamic movements of the limbs and head. The lightweight frame minimizes the energy required for moving and decreases the output torque demand on each motor, allowing for the use of smaller more efficient motors. Finally, the modularity of the frame will be beneficial, allowing for easy assembly, maintenance, and upgrades after tests are completed.

The robot's head, torso, and limbs will be designed to reflect a humanoid appearance. This will make the robot approachable and engaging for users which is crucial. These components will be mostly fabricated through 3D printing which can create the aesthetic and intricate parts of the robot that will be visible. Finally, the robot will be representing its college, so the school's colors will be displayed on the robot.

3. Electrical power and communication

The NVIDIA Jetson is the "brain" of the robot and will process both interaction and movement. Interaction inputs will be collected using a camera and an external microphone array. The outputs will include speakers and a display screen with an adaptive face. These components will connect to the Jetson via USB, which will also supply power to them.

For movement, the Jetson will send control signals to each of the servo motors using the Universal Asynchronous Receiver Transmitter (UART) protocol. UART is a simple communication protocol that allows for asynchronous communication, meaning each device doesn't need to share a clock signal. Each motor will be able to send feedback from its internal magnetic encoder, allowing for high-precision control. Figure 2 summarizes the connections between the Jetson and other components for communication.

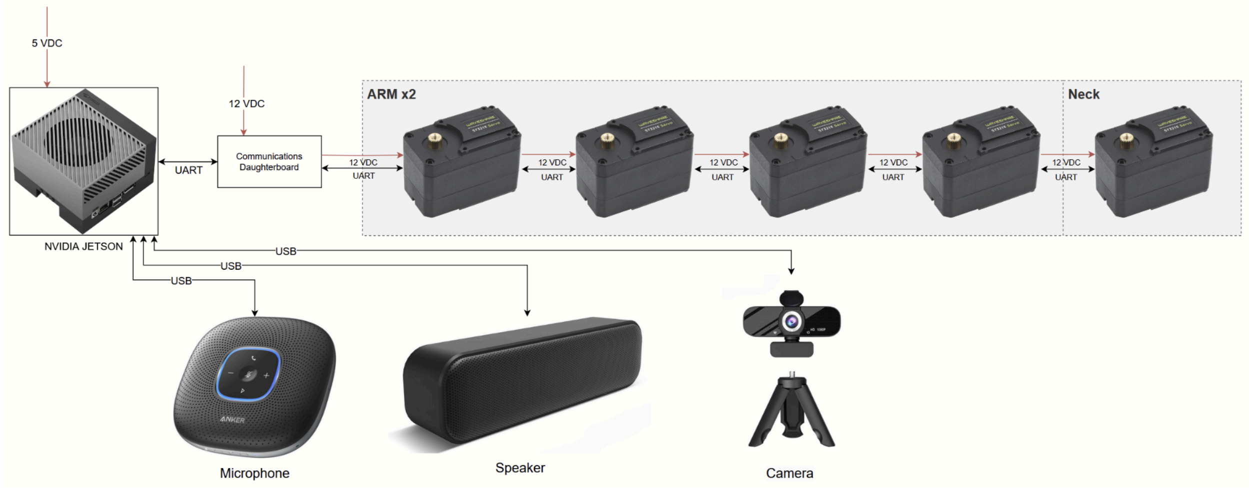


Figure 2: Communication Diagram

A power distribution system will supply power to the Jetson and motor controllers at different voltages. The robot will be powered from an AC wall outlet connected to a DC power supply. Using AC wall power is cost-effective and gives the robot unlimited runtime without the need for recharging batteries.

4. Human pose estimation

The greeting robot has a mimicry mode, meaning that the robot can mirror the movements of the guest it is interacting with. This requires a real-time sensing of the human pose in order to get the data to apply to the robot. However, determining the exact pose of a human using conventional image processing techniques can be extremely challenging and computationally infeasible in real-time. To address this problem and get exact data in real-time, the robot will use a neural network to convert an image to a list of key points and their exact location in the image. An example of these key points, forming a "pose skeleton," is shown in Figure 3.

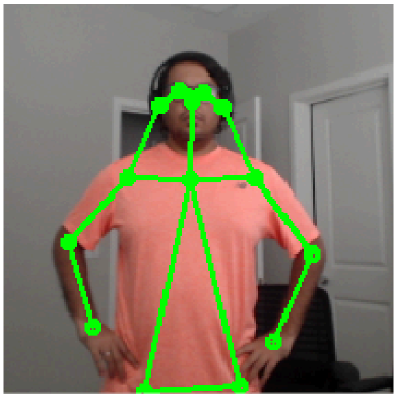


Figure 3: Pose Skeleton[2]

Using this pose skeleton, joint angles along specific axes of rotations can be determined using trigonometry and linear algebra. This would result in 3 angles for the shoulder and 1 angle for the elbow. These joint angles can then be applied to the limbs of our robot in order to mimic the pose, without the need for scaling between different limb lengths and human heights. For the limbs, only the shoulder, elbow, and wrist points would be required to satisfy the desired DOF specified above. The system handling Human Pose Estimation (HPE) and control of the robot will use Robot Operating System 2 (ROS2), an open-source framework for developing and controlling robots.[3] ROS2 simplifies complex robot programming with its modular design, extensive libraries, and strong community support, enabling accelerated development. This will allow for easy integration with our hardware components while also providing ease of controlling the robot using libraries.

5. Artificial intelligence pipeline

A key requirement is that the greeting robot must recognize and answer questions. However, it is impossible to know ahead of time every single question that could potentially be asked by a visitor. To meet the requirement and address this problem, artificial intelligence (AI) must be used to adapt and respond. A generative AI chatbot with built-in general knowledge will be used, supplemented by a database of known facts about the college. This allows the robot to dynamically adapt to any question that might be asked, and provide a reasonable response. Furthermore, the client stipulated that the robot should be able to answer approximately 20 questions about the college consistently and accurately.

The system built to address the design factors described above has been dubbed the “AI Pipeline” and consists of the following software components:

- Speech-to-Text (STT) Module – voice recognition/natural language processing (NLP)
- Large Language Model (LLM) – ex. ChatGPT, Google Gemini
- Text-to-Speech (TTS) Module – transforms text output from the LLM into audible speech

For a better understanding, Figure 4 shows a visual representation of the AI Pipeline:

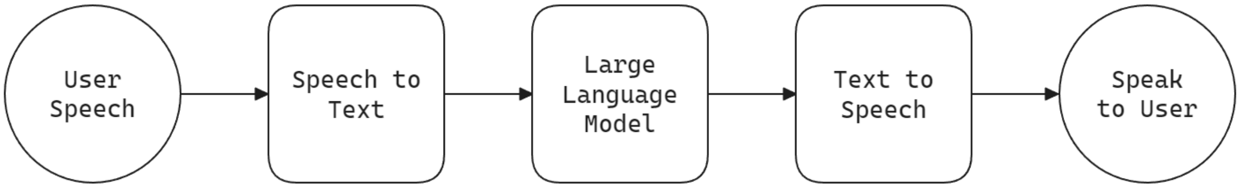
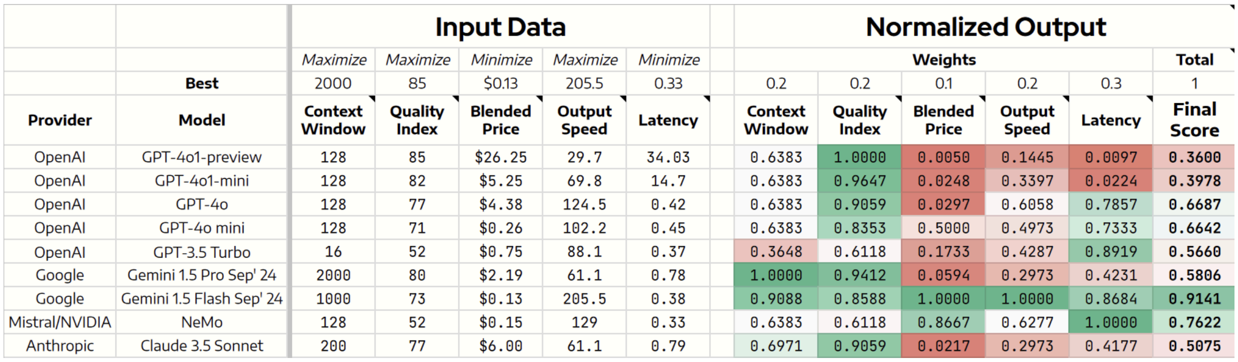


Figure 4: AI Pipeline Overview

To select an LLM for the AI Pipeline, the following five metrics were considered: **(1)** *context window (in thousands of tokens)* – the number of tokens (numerical representations of words) that an instance of generative AI can process at once. Essentially, it determines the amount of knowledge that can be added about the college and how much the LLM can “remember” in a single conversation. **(2)** *quality index* – a normalized relative metric[4] representing the relative response quality among different AI models. **(3)** *blended price* – LLM providers typically charge a different rate per input token than per output token. Also, there are typically more input tokens than output tokens, by a ratio of 3:1 (input tokens to output tokens).[4] This metric is the price per one million tokens at the 3:1 ratio. **(4)** *output speed* – the median output speed in tokens per second, while the LLM is generating tokens. Factors into total response time. **(5)** *latency* – the median latency until the first token is received; a major factor in total response time.

Data for these metrics[4] are combined in Table 2 below, normalized to the maximum value of that metric across all LLMs under consideration for usage in the project, and then weighted based on project design requirements.

Table 2: LLM Decision Matrix



		Input Data					Normalized Output					
		Maximize	*Maximize*	*Minimize*	*Maximize*	*Minimize*	Weights					Total
	Best	2000	85	$0.13	205.5	0.33	0.2	0.2	0.1	0.2	0.3	1
Provider	**Model**	**Context Window**	**Quality Index**	**Blended Price**	**Output Speed**	**Latency**	**Context Window**	**Quality Index**	**Blended Price**	**Output Speed**	**Latency**	**Final Score**
OpenAI	GPT-4o1-preview	128	85	$26.25	29.7	34.03	0.6383	1.0000	0.0050	0.1445	0.0097	**0.3600**
OpenAI	GPT-4o1-mini	128	82	$5.25	69.8	14.7	0.6383	0.9647	0.0248	0.3397	0.0224	**0.3978**
OpenAI	GPT-4o	128	77	$4.38	124.5	0.42	0.6383	0.9059	0.0297	0.6058	0.7857	**0.6687**
OpenAI	GPT-4o mini	128	71	$0.26	102.2	0.45	0.6383	0.8353	0.5000	0.4973	0.7333	**0.6642**
OpenAI	GPT-3.5 Turbo	16	52	$0.75	88.1	0.37	0.3648	0.6118	0.1733	0.4287	0.8919	**0.5660**
Google	Gemini 1.5 Pro Sep' 24	2000	80	$2.19	61.1	0.78	1.0000	0.9412	0.0594	0.2973	0.4231	**0.5806**
Google	Gemini 1.5 Flash Sep' 24	1000	73	$0.13	205.5	0.38	0.9088	0.8588	1.0000	1.0000	0.8684	**0.9141**
Mistral/NVIDIA	NeMo	128	52	$0.15	129	0.33	0.6383	0.6118	0.8667	0.6277	1.0000	**0.7622**
Anthropic	Claude 3.5 Sonnet	200	77	$6.00	61.1	0.79	0.6971	0.9059	0.0217	0.2973	0.4177	**0.5075**

For the greeting robot, the most important factor in LLM selection is total response time, so the highest weighted metrics are latency and output speed. Quality index and context window size are directly related to informational accuracy, which is also a key requirement specified by the client. Also, note that context window size is normalized logarithmically. Price is the least important out of the factors here, particularly since API calls are relatively cheap. It is clear from Table 2 that Google's Gemini 1.5 Flash (specifically the September 2024 version) is the best solution according to these metrics. It has the lowest price, highest output speed, and near highest context window size and quality, as well as nearly the lowest latency. Its final weighted score is by far the best – thus, it has been selected as the LLM for the AI Pipeline.

Figure 5 shows how the LLM is integrated into the wider system at a high level. Input comes from two places: 1. as audio data directly from the guest, recorded by the robot's microphone and processed into text by the STT model within the main interaction program (programmed in Python). 2. as textual facts from a database loaded into memory at runtime for fast retrieval. Google Sheets is used as the frontend of the fact database since it is cloud-hosted and easy to update for non-technical users (i.e., the client of the project, who will manage the greeting robot after the project concludes). Output is delivered in three ways: 1. mechanically through gestures by the robot's arms, 2. audibly through a speaker connected to the robot (the primary method), and 3. visually through the robot's face, primarily characterized by a set of dynamic eyes that will change as the robot speaks.

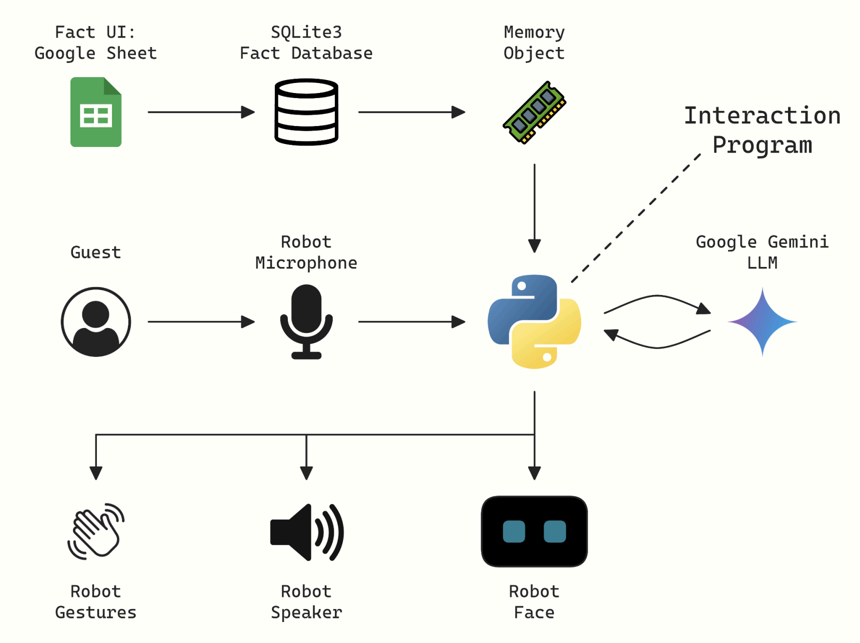


Figure 5: AI Pipeline System Diagram

Concluding Remarks

This capstone design project successfully incorporates various design factors into a greeting robot that will enhance outreach efforts for the represented college. Key design challenges were addressed including motors and power transmission, structural design, electrical power and communication, human pose estimation, and dynamic communication using AI. The project highlights the importance of collaboration between distinct disciplines to create a functioning system. The robot will interact with users in a meaningful way by responding to verbal interactions, providing specific information about the college, performing gestures, and mimicking human movements. Ultimately, the project will showcase the innovation of the college and its students while inspiring future generations to pursue STEM careers.

References

1. Advanced Motion Controls. "Advantages of Servo Motors." *Advanced Motion Controls*. Accessed January 13, 2025. https://www.a-m-c.com/servomotor-advantages/.
2. Open Robotics. "ROS 2 Documentation." Accessed December 15, 2024. https://docs.ros.org/en/jazzy/
3. NVIDIA_AI_IOT. "GitHub - NVIDIA-AI-IOT/ros2_trt_pose: ROS 2 package for 'trt_pose': real-time human pose estimation on NVIDIA Jetson Platform.*"* Accessed October 28, 2024. https://github.com/NVIDIA-AI-IOT/ros2_trt_pose
4. "AI Models Leaderboard | Artificial Analysis." *artificialanalysis.ai.* Accessed November 20, 2024. https://artificialanalysis.ai/leaderboards/models