Threepio: From Intelligence to Interaction in Embodied AI Systems

Samuel J. Potozkin

Chapman University

School of Communication

Fowler School of Engineering

April 2026

Abstract

This paper presents Threepio, an embodied artificial intelligence system designed to enable real-time, character-consistent human interaction through a physically constructed platform. The project combines speech recognition, large language model reasoning, and text-to-speech synthesis all contained within a modular runtime architecture, allowing for continuous conversational exchange in an ambient environment. This contrasts against traditional screen-based interfaces by exploring the role of embodiment in shaping how humans perceive artificial agents by drawing on communication theories such as the Media Equation and Social Information Processing Theory to inform system design.

This project integrates both software and hardware components, including a Raspberry Pi-based embedded system, a MEMS microphone, an exciter speaker, and a multi-layer automotive finishing process to achieve a visually and acoustically authentic representation of C-3PO from the Star Wars saga. Key engineering challenges included real-time audio processing, echo mitigation, constrained hardware integration, and precision surface finishing using chrome and translucent lacquer systems.

This work demonstrates how advanced intelligence can be translated into interactive, physically situated systems, highlighting a broader shift from computational capability toward experiential and embodied interaction.

Keywords: embodied AI, human-AI interaction, real-time systems, conversational agents, embedded systems

Prologue: Innovating Under Constraints

My father, Dr. Jerome R. Potozkin, showed me Star Wars Episode IV: A New Hope (1977) when I was three years old. At the time, it was just an entertaining story that excitedly told classic morals. As I got older, my appreciation for the series began to shift towards the way it was created.

In 1977, George Lucas set out to create a universe that could not be shown with the technology available at the time. Instead of compromising the vision to fit within current technology's constraints, Lucas assembled a team willing to experiment, iterate, and develop entirely new technologies and methods to achieve what had previously been considered impossible. This effort eventually led to the creation of Industrial Light & Magic (ILM), a studio built to push the technical boundaries of filmmaking and, in doing so, redefine what was possible in VFX (American Society of Cinematographers, 2025).

When existing tools are insufficient, new ones can be created. Innovation does not come from accepting constraints, but from questioning and challenging them. This perspective extends to fields like engineering, artificial intelligence, and entrepreneurship, where progress is driven by people willing to experiment and build what does not exist.

As my educational experience has shaped my interest in technology and business, I've found myself drawn towards this mindset more and more. Instead of seeing systems as limited by their current capabilities, I began to see them as flexible systems that could be extended, modified, and entirely reimagined.

This project is a direct reflection of that perspective. Characters like C-3PO were originally constrained to fiction. Not because the idea was impossible, but because the underlying technology had not yet caught up. Current advances in artificial intelligence, speech synthesis, and embedded systems have significantly reduced those limitations. Tasks that once required entire studios and specialized facilities can now be completed at an individual level.

This project began by asking a simple question: What happens when the constraints that once made something fictional begin to disappear?

Introduction: From fiction to feasibility

Over the past decade, the gap between broad ideas of science fiction and real-world technology has narrowed significantly. Advances in AI, speech synthesis, and embedded computing have opened the door to creating technology that processes language, generates responses, and engages in ongoing conversations with human users.

Tasks that once relied heavily on advanced research teams and specialized organizations are increasingly accessible at an individual level. Open-source resources, commercially available hardware, and widely accessible AI models have lowered the barrier to entry, allowing complex systems to be iterated on and prototyped outside of traditional institutional environments.

By extension, the way we, as humans, perceive this technology is also changing. Machines are no longer limited to one-dimensional tools restricted to carrying out commands. Today, machines are interactive systems capable of maintaining dialogue, adapting to context, and behaving like conversational partners rather than strictly mechanical interfaces.

This shift highlights an important question: If the underlying technology exists, why are these experiences confined to boxes with LCD screens rather than integrated or embodied in physical form?

This project explores that question through the development of a physically integrated, AI-powered embodied system inspired by the C-3PO character from the Star Wars universe. Rather than treating the character purely as fiction, it treats C-3PO as a reference for human-AI interaction; an example of how responsiveness, personality, and communication structure can be combined within a real-world system.

By integrating speech recognition, language modeling, personality design, and physical hardware into a unified architecture, this project illustrates the steps required to translate a fictional communicative system into a functional prototype.

Why Now: The Shift in AI

Systems like the one in this project are beginning to emerge as a broader shift in AI architecture unfolds. Recent discoveries and advancements across multiple layers of the AI ecosystem have accelerated the transition from an experimental technology to a consumer-facing tool.

This progression is often expressed as a layered stack, where each level builds on the one below it. At the lowest level lie the energy and infrastructure required to support these systems, followed by data centers and the necessary hardware, such as GPUs (Graphics Processing Units), which are responsible for their training and deployment. Large-scale AI models sit on top of that foundation, capable of processing language, generating responses, and maintaining context across interactions. This project sits at the application and embodiment layer of that stack, where underlying intelligence is translated into systems that humans can directly interact with. While much of the current AI landscape remains confined to screens, this layer is where intelligence becomes an experience. This distinction marks a transition from systems that strictly compute, to systems that interact.

Recent demonstrations of embodied AI systems further drive this point, highlighting a shift towards machines that can process information, communicate, respond, and exist within physical environments. Evidence for this direction exists in recent industry developments, including demonstrations of embodied, conversational AI presented at major conferences such as NVIDIA GTC. These improvements suggest that the next phase of Artificial Intelligence will not be defined by its intelligence, but by the experience and interaction users facilitate.

In this context, the creation of a conversational, physically embodied system like Threepio is not an isolated experiment, but an early example of that broader transition: the movement from abstract intelligence to interactive, human-centered systems.

Human-AI Interaction: A Theoretical Framework for Design

As artificial intelligence becomes more capable, the limiting factor isn't the intelligence itself, but how that intelligence is presented, interpreted, and experienced by humans. The design of this project was not solely grounded in engineering, but also in communication theory. Understanding how and why humans interpret interactions with artificial systems as social experiences is critical to designing effective human-AI experiences.

This lens builds on my prior academic work on human-AI interaction and communication theory, which examined the conditions under which communication with artificial intelligence is interpreted as interpersonal rather than purely technological.

Social Information Processing Theory suggests that interpersonal relationships can develop through communication over time, even when using channels that lack nonverbal cues.

Rather than strictly relying on physical presence, individuals construct the identity of the other speaker through patterns of interaction, responsiveness, and message exchange. This principle is directly applicable to conversational AI systems, which rely primarily on dialogue rather than physical expression (Walther, 1992).

Similar ideas are also explored in the Media Equation, which extends on how users respond to artificial systems in social ways. Research points to a subconscious set of social rules used when speaking to artificial systems, including politeness, reciprocity, and emotional interpretation. These rules are still employed when users know that the systems they are interacting with are artificial. When an artificial system uses the interpersonal frameworks through which humans interact, interactions are perceived similarly to how a communicative partner would be received (Reeves & Nass, 1996).

This project applies these theoretical principles in practice. By creating a system that emphasizes responsiveness, personality, and continuity across conversations, it attempts to recreate the conditions under which humans interpret interaction as social rather than mechanical. In this sense, the system is both an engineering implementation and an applied exploration of communication theory within an artificial context.

System Overview: Bridging Theory and Implementation

This project explores the translation of a fictional character into a real-world, embodied AI system. Throughout this paper, “C-3PO” refers to the fictional character, while “Threepio” refers to the embodied AI system developed for this project. Developing this system required translating and combining abstract concepts from both AI and communication theory into a functional, working architecture. While previous sections outlined why human-AI interactions can be interpreted as social, this section focuses on how those ideas are implemented within a physical system.

At a high level, this system operates as a layered pipeline that processes user input, generates a context-aware response, and delivers an output that resembles interpersonal communication. Each component contributes to the overall goal of creating a system that feels responsive, continuous, and socially aware. This structure reflects a broader architectural pattern in modern AI systems, where perception, reasoning, and output are separated into distinct layers, allowing each to be independently optimized while contributing to a unified interaction model.

The interaction starts with speech input, captured via speech recognition. That input is then passed to a language model capable of interpreting context, maintaining dialogue, and generating responses. Rather than transmitting raw responses directly, the system incorporates a "personality layer" that outlines how the response should sound and behave, ensuring it aligns with C-3PO's character and behavior.

Once a response is generated, it is converted to audio using a text-to-speech system and delivered via a built-in speaker. This interaction loop enables ongoing, real-time dialogue between the user and the system.

Additionally, the system is physically embodied within a hardware structure that includes lighting, audio components, and embedded computing hardware. This embodiment is not incidental, but central and essential to the design. By placing the system in a physical form, the interaction moves beyond screen-based communication and into a tangible experience.

Conceptually, the system can be understood as a combination of four key layers: perception, reasoning, personality, and embodiment. Perception is what enables the system to receive input, reasoning allows it to process responses, personality shapes how those responses are communicated, and embodiment defines how the system is interpreted and experienced in the physical world.

The intersection of these layers allows the project to move from a theoretical exploration into an applied implementation, demonstrating how principles of human-AI interaction can be realized within a working system.

Hardware Architecture: Physical Integration and System Constraints

The software system defines how Threepio thinks and communicates. The hardware determines how the system exists and operates within the physical world. The transition from a purely digital system to a physical, embodied one introduces new constraints, including space limitations, power management, thermal management, and audio clarity.

At the heart of the system is a compact, embedded computing platform designed to handle real-time audio processing, model interaction, and system orchestration. This platform is installed directly within a physical structure of Threepio's head, requiring careful consideration of placement and internal layout.

The audio system plays a central role in the interaction experience. A microphone captures user input, while an internal speaker delivers generated responses. The challenges of achieving clear input and output within an enclosed system include echo, feedback, and sound dispersion, which require both hardware positioning and software mitigation strategies to ensure reliable performance.

The integration of these components within a confined space requires iterative design and modification. Mounting solutions, internal supports, and component spacing were adjusted to balance functionality and accessibility, ensuring the system remained maintainable while preserving the external form factor.

This embodiment is not strictly aesthetic, but functional. It allows the system to coexist with the user in a shared physical environment, reemphasizing the principles of presence and interaction that underpin the overall design.

Embedded Systems: Audio, Optics, and Signal Chain

The embedded hardware platform was built around a Raspberry Pi 5. This decision was made due to its balance of computing power, small form factor, and compatibility with real-time audio input. This functioned as the system's "brain", handling microphone input, language-model communication, text-to-speech generation, digital signal processing, and LED control.

Audio input was handled via an INMP441 I2S MEMS microphone, which offered a compact form factor and a direct digital interface. This allows for clean digital transmission directly into the Pi without any analog noise.

A MAX98357A and a Dayton Audio DAEX25FHE-4 speaker exciter handled audio output. Unlike a traditional cone speaker, the exciter converts electrical signals into vibrational energy, which is then transferred directly to the housing's structure, effectively turning the entire head into a radiating acoustic surface.

From a physics standpoint, this occurs through the electromagnetic motion of a voice coil within a magnetic field. Alternating current induces oscillatory motion, causing the mounted surface to vibrate and emit acoustic pressure waves into the surrounding air. This approach was particularly effective in creating the illusion that the voice is coming from within the droid, rather than from an exposed speaker opening.

The LED eye system was implemented using six pre-wired 5V LEDs (3 LEDs per eye) housed in custom, outsourced ultra-clear resin lenses, which allowed the light to effectively spread throughout the eyes and preserve as much realistic light diffusion as possible, a task that would have been difficult to achieve through standard in-house printing workflows.

Software Architecture: Modular Design and Real-time Interaction

Threepio's system is designed as a modular software that supports real-time conversations across multiple environments. Instead of relying on a single, linear system that handles all tasks, the program offloads tasks around interchangeable components that handle input processing, reasoning, personality shaping, and output generation. This modular approach allows those subsystems to be developed, tested, and refined independently while remaining a part of a consistent interaction model.

The system operates as a pipeline that takes user input (speech) and produces a structured conversational response. In these voice-based interaction environments, audio is captured through a microphone and segmented using voice activity detection to isolate speech. This audio is transcribed into text using a speech recognition module and passed to the reasoning layer, where a language model generates a response based on conversational context and prior interaction history.

Before being delivered to the user, the response is passed through a dedicated persona layer. This layer is responsible for shaping tone, phrasing, and conversational behavior to ensure they align with the characteristics and values associated with C-3PO. By separating its personality from its core reasoning, the system maintains flexibility while ensuring consistent output.

The final stage of the interaction pipeline converts the system's response to an audio file using a text-to-speech system. That audio file is then passed through a series of effects that simulate the vocal processing by Ben Burtt, the ILM sound designer responsible for the effects applied to 3PO's voice in the Star Wars films. The resulting audio is then played through the system's speaker, completing the interaction loop and enabling a continuous dialogue.

The vocal system required both machine-learning-based synthesis and traditional digital signal processing. The raw speech output was generated by a neutral text-to-speech model that predicts phonetic timing, intonation, and vocal timbre from textual input. On a technical level, this process can be understood as a sequence-based acoustic modeling process, where the model estimates waveform characteristics across time to produce natural–sounding speech.

However, the generated voice did not initially align with the metallic vocal tone associated with C-3PO. To address this, a post-processing signal chain was developed using digital audio effects commonly employed in audio engineering and sound design. A short delay line was introduced to create tightly spaced temporal reflections, producing a subtle mechanical tone that had characteristics of synthetic vocal spaces. In parallel, a chorus effect was applied by slightly modulating the time and pitch of duplicated signal paths, creating the impression of layered vocal resonance. Together, these effects transformed the raw synthesized voice into a more convincing in-universe protocol droid vocal profile. The final result required repeated iteration of delay times, modulation depth, feedback levels, and filtering parameters, reflecting a workflow similar to that of professional audio engineers and sound design.

In addition to the core pipeline, the architecture supports multiple operational modes. A command-line interface gives users a deterministic interaction path for testing and tool-based responses, while an ambient mode enables continuous, microphone-driven interaction in a physical environment. A real-time streaming mode integrates with external API's to support low-latency conversations with built-in interruption handling.

This multi-layered mode design reflects a key characteristic of architectural decision: separating interaction logic from input and output mechanisms. Speech recognition, language modeling, persona shaping, and audio playback are separated into distinct components, allowing the system to adapt to different environments without changing its core behavior.

The architecture also incorporates practical considerations required for real-world deployment, including echo mitigation, speech gating, and interruptible playback. These features allow the system to operate reliably in a physical setting, where background noise, feedback, and overlapping input would otherwise disrupt interaction.

In summary, the software architecture demonstrates how a conversational system can be structured as a cohesive, extensible platform; one that integrates multiple subsystems into a single, unified experience designed for real-time human-centered interaction.

Electrical Integration and Board-level Assembly

Physical integration required a board-level electrical assembly to establish reliable communication between the computing system and peripheral hardware. This included voltage-supply pins, ground references, and multiple communication lines for transmitting digital audio and control signals between the microphone, the amplifier breakout board, the LED assemblies, and the speaker system.

A major part of this process involved soldering, the technique used to create permanent conductive bridges between electronic components and their corresponding connection points. In practical terms, soldering involves heating a low-melting-point conductive metal alloy and allowing it to flow between two electrical contacts, where it cools and solidifies into a stable conductive junction. Functionally, this acts as both an electrical and mechanical bond, allowing current and digital signals to travel reliably between boards, wires, and interface pins.

Engineering Challenges and Iterative Problem Solving: Constraint, Failure, and Refinement

The development of Threepio was defined as much by iteration and recovery from failure as by the initial design. Rather than a linear process, the project required repeated cycles of diagnosis, redesign, and refinement across the physical fabrication, electronic integration, and surface finishing processes.

One of the earliest challenges came about during the fabrication process. The rear dome of the head was accidentally dropped, resulting in a significant crack across the left side. Instead of reprinting the entire section, it was repaired using glazing spot putty. From a chemical standpoint, this material is typically a nitrocellulose-based or polyester-based filler compound engineered for low-shrinkage repairs of fine surface cracks, pinholes, and shallow structural defects. The filler's role was not cosmetic; it restored surface continuity at a microstructural level, allowing the repaired region to be feathered into the surrounding geometry through progressive sanding.

Surface finishing introduced one of the most technically demanding challenges in the entire build. Achieving a convincing chrome finish required a technical and meticulous workflow. Unlike standard metallic paints, the Alclad II Chrome system relies on an extremely thin lacquer suspension containing ultra-fine metallic flake pigments. These pigments are designed to reflect light in a near-specular manner. The reflective quality of this finish is highly dependent on the optical smoothness of the substrate beneath it. Because chrome lacquers do not hide surface imperfections but instead amplify them through reflected light, the final finish's success was determined almost entirely by the precision of the layers beneath it.

A significant portion of the finishing process required learning principles typically associated with automotive paint systems. The metallic gold appearance could not be convincingly achieved with standard opaque acrylics and enamels, as these would compromise the reflectivity of the chrome substrate. Instead, the process draws on automotive candy-coat finishing techniques, where transparent color layers are sprayed over a reflective metallic or chrome base to create depth, warmth, and cohesive optical richness.

At an industrial scale, this effect is typically achieved using an HVLP (high-volume, low-pressure) spray system to apply translucent candy colors over a polished metallic surface. To replicate this effect at a smaller scale, these same principles were adapted through an airbrush-based finishing workflow. A custom translucent lacquer mixture, using a 5:1 ratio of transparent yellow to transparent orange, was developed to emulate the warm gold tone of the original C-3PO prop.

Because this process relies on light passing through the tint layer, reflecting off the chrome, and returning through the pigment, the final appearance is highly sensitive to application consistency. Even minor variations in thickness or flow can result in visible defects, making this stage particularly difficult to execute reliably. This optical dependency made the process inherently unforgiving.

During application, several visible drips developed in the tinted lacquer layer. Under normal surface-finishing standards, this would have required a complete correction process involving sanding and reapplication. However, because the original C-3PO prop is canonically weathered and exhibits signs of age and use, these surface irregularities were ultimately evaluated as consistent with the design language of the character rather than as critical defects. This decision reflects an important design principle that emerged throughout the project: distinguishing between true defects and aesthetic deviations that remain consistent with the intended system.

The final protective layer involved the application of a two-component clear coat system, commonly referred to as a “2K-clear”. This is the same class of protective coating technology used in automotive paint systems to seal and protect a car's exterior finish. In practical terms, this is the hardened glossy surface people physically interact with every day when they touch a car door, open the trunk, or run their hand across the hood. It is important to distinguish the 2K clear coat from a conventional paint sealer or standard aerosol clear coat. While these terms are often used interchangeably, they represent fundamentally different material systems. A standard sealer or single-compound (“1K”) clear coat typically dries through solvent evaporation. In this process, the liquid carrier evaporates into the air, leaving behind a solid film on the surface. This process is effective when protecting from UV rays and enhancing the surface, but it neglects resistance to scratching, chemicals, and long-term wear. By contrast, a 2K (two-component) clear coat is a chemically curing polymer system. It consists of a resin base and a separate hardener, which are activated immediately before use. Once combined, these materials undergo a crosslinking reaction in which the polymer chains chemically bond to one another, forming a dense, hardened network. In practical terms, this significantly increased both the quality and the risk of the finishing process. The 2K layer was not merely sealing the paint beneath it, but was also creating a hardened structural shell over the chrome and candy layers, locking in gloss, depth, and long-term surface protection.

One major disadvantage of using a 2K clear-coat system is the development of orange peel. In this surface-texture issue, the cured finish develops a slightly dimpled appearance similar to the skin of an orange. In automotive finishing, this is typically caused by factors such as improper spray distance, poor atomization, or environmental conditions during application. Because the 2K layer served as the final surface, even minor surface-texture flaws disrupted the mirror-like reflectivity of the chrome and candy-coated layers beneath it by scattering light rather than allowing clean reflection, and correcting this required carefully wet-sanding the surface to level the raised texture without sanding through the hardened clear coat and exposing the layers beneath. This was one of the most technically stressful stages of the build, as excessive sanding could have compromised hours of prior finishing work. Once the surface was restored to optical smoothness, a fresh coat of 2K clear was reapplied to rebuild gloss depth and restore the final protective finish.

Integrating all of the parts posed a whole separate set of engineering challenges. The primary model/head was derived from a digital scan of the original film prop. The neck assembly, however, had to be redesigned from scratch after the provided neck model proved to be structurally insufficient. The revised neck was intentionally designed to support the embedded computer, incorporating a lid structure that securely houses the Raspberry Pi developer board while maintaining structural stability.

Ensuring that all computer components and breakout boards worked introduced additional complexity. Internal constraints required repeated prototyping of component placement, particularly for the microphone and speaker. Achieving reliable fitment while preserving audio clarity required multiple hardware layouts and adjustments.

The most daunting aspect of the project was acquiring entirely new technical skills, including soldering and board-level assembly. For a system that depended on reliable electrical connectivity, learning to solder components and wiring connections brought both technical and psychological challenges, particularly given the risk of damaging sensitive components.

In sum, these challenges highlight that the project was not a linear build process but an iterative engineering process ultimately defined by adaptation, failure recovery, and design refinement across a handful of technical domains.

Design Philosophy: Why C-3PO

The decision to center this project on C-3PO was not purely aesthetic but rooted in design philosophy and human-computer interaction. As a character, C-3PO stands in a unique position within popular culture: he is clearly artificial, yet consistently perceived as socially present, emotionally expressive, and human in the ways audiences perceive and relate to him. This distinction makes him an ideal reference point for exploring embodied conversational AI.

Unlike most robots that emphasize cold efficiency or machine-like detachment, C-3PO is defined by his personality, tone, and relational presence. His identity is conveyed less through physical realism and more through the way he communicates, through voice, phrasing, emotional cadence, and social responsiveness. These characteristics align directly with the communication theories discussed earlier in this report, particularly the idea that humans interpret interactions through social frameworks when communication patterns resemble interpersonal exchange.

From a systems-design perspective, C-3PO functions as more than a character reference; he serves as an interface model. The physical embodiment (Illuminated eyes, vocal delivery, and recognizable personality) all contribute to the user's perception of presence. These design choices help transform the system from a strictly functional tool into something that feels interactive, expressive, and socially meaningful.

From a broader perspective, the use of C-3PO reflects an intentional effort to make advanced AI systems feel approachable. Rather than presenting the technology as abstract or intimidating, the design leverages a familiar reference point to create immediate recognition and emotion. In this sense, character design becomes a bridge between technical capability and human acceptance.

This philosophy reinforces the central idea of the project: the future of embodied AI will not be defined solely by computational intelligence, but by the design of systems that people feel comfortable interacting with in everyday life.

Future Implications and Industry Context

While this project began as something inspired by science fiction, its broader significance lies in what it represents about the future of human-AI interaction. Recent developments across artificial intelligence suggest that systems are rapidly moving beyond screen-based interfaces and towards embodied, interactive forms of intelligence.

This shift has become increasingly evident in industry demonstrations of robotics, conversational agents, and physically integrated AI systems, including recent showcases at NVIDIA GTC, where embodied AI and interactive robotics have become major focuses (NVIDIA, 2026). These systems are no longer strictly defined by their ability to process information, but by their ability to communicate, respond, and operate within shared physical environments. In this context, the evolution of AI is no longer a question of increasing what it's capable of, but a shift from screen-based interactions to systems that exist within a shared physical and social environment.

In this context, Threepio serves as a small-scale exploration of a much larger technological direction: the convergence of language models, speech systems, embedded computing, and physical embodiment. The same principles applied here; real-time interaction, personalization shaping, presence cues, and hardware-software integration are directly relevant to the next generation of robotics and human-centered AI systems.

From a broader perspective, this project reflects a belief that the future of artificial intelligence will be defined not only by intelligence itself but also by how naturally these systems integrate into human life and social frameworks. As AI systems continue to move closer to everyday life, design decisions about personality, communication, and physical presence will become increasingly important.

In this sense, Threepio is not simply a standalone build, but an applied exploration of where conversational and embodied AI may be headed in the years ahead. The long-term trajectory of AI will likely be defined not just by what these systems can do, but by how naturally they integrate into the environments and relationships that define everyday human life.

Epilogue: From inspiration to implementation

This project began with a childhood memory: being introduced to Star Wars by my father and becoming captivated by the world it created. Over time, what initially began as a fascination with the story evolved into an appreciation for something deeper, the spirit of innovation that made that world possible in the first place.

The story of George Lucas and Industrial Light & Magic has been a recurring source of inspiration throughout this process. At its core, their work embodies a simple yet powerful idea: when existing tools are insufficient, new ones can be built. That principle became central to the way I approached this project.

Threepio ultimately became far more than a physical build. It required integrating concepts from communication theory, software architecture, embedded systems, materials engineering, and user-centered design into a single, cohesive system. In doing so, the project became an exercise in multidisciplinary problem-solving, requiring both technical skill and conceptual thinking.

In many ways, this project represents the transition from inspiration to implementation, from admiring what others built to developing the skills and confidence to build something of my own.

If the first generation of artificial intelligence was defined by intelligence itself, the next will be defined by integration and how seamlessly these systems exist within human environments, communicate within social frameworks, and become a part of everyday life.

Appendix: Technical Methodology, Development Notes, and Fabrication Workflow

Appendix A: Software Stack and Runtime Environment

The software stack was developed primarily in Python due to its iteration capabilities and mature ecosystem for artificial intelligence, audio processing, and embedded systems development.

Core configuration and environment management were handled with Pydantic, which provided environment-specific configuration layers that enabled modular switching between development and deployment environments.

Major software libraries used include:

- OpenAI Python SDK – language-model reasoning and optional text-to-speech
- Faster-Whisper – Local speech-to-text transcription optimized for CPU inference
- SoundDevice + NumPy – Low-latency audio capture and playback
- FFmpeg – digital signal processing and C-3PO vocal effects
- WebRTC VAD – frame-level voice activity detection
- ElevenLabs/OpenAI TTS – modular text-to-speech providers

The software stack was intentionally modular, allowing the speech, reasoning, and output systems to be independently tested and iterated without altering the broader runtime pipeline.

Appendix B: Runtime interaction pipeline

The runtime interaction pipeline starts with low-latency microphone capture through SoundDevice and NumPy-based audio buffers, which ingest live audio input directly from the embedded system.

Incoming audio is then processed by the voice activity detection (VAD) layer, which combines RMS energy thresholds with WebRTC-based frame classification. This allows the system to distinguish between intentional speech and silence and ambient noise.

Once valid speech is recognized, it is transcribed and then passed into a prompt-construction and memory layer. This layer accesses system instructions, conversation history, and personality guidelines to ensure all required information is assembled before sending it to the language model for response generation.

The generated response is then refined through a persona-shaping layer. This layer is designed to reflect the phrasing, tone, cadence, and conversational architecture that is consistent with how C-3PO communicates in the Star Wars films.

The final output is routed through a text-to-speech provider and into an FFmpeg-based digital signal processing chain, which applies equalization, compression, delay, and modulation effects to emulate the metallic voice created by sound designer Ben Burtt for the original character.

The processed signal is then sent to the exciter speaker for physical playback.

Appendix C: Materials, Finishing Stack, and Prop Construction Workflow

The physical construction of Threepio was done in a multi-stage workflow that combined fabrication, surface engineering, and embedded systems integration.

The primary geometry was created with 3d printing, using high-fidelity digital scans of the original film prop. All external surfaces were progressively sanded to remove all visible layer lines and prepare the substrate for finishing.

Surface correction involved using filler putty and glazing compounds, which were applied to seams, cracks, and print imperfections before being sanded again so that the entire surface was flush with surrounding geometry.

The final metallic appearance was achieved through a multi-layer finishing stack, derived from automotive painting workflows. This consisted of a gloss black reflective base, Alclad II chrome lacquer, and a 5:1 translucent yellow-to-orange tint coat applied by airbrush, and a final 2K automotive clear shell.

The translucent tint layer was specifically selected to preserve the reflectivity of the chrome beneath it, while reproducing the warm gold tone associated with C-3PO.

The eye assembly was intentionally outsourced so that the LED housing could be constructed in an ultra-clear resin to achieve a level of transparency and light diffusion that would have not been possible through standard in-house printing workflows.

The final assembly required the integration of the physical shell, embedded electronics, soldered breakout boards, the LED systems, and the acoustics into a unified embodied AI platform.

References

American Society of Cinematographers. (2025, June 23). Behind the scenes of Star Wars. https://theasc.com/articles/behind-the-scenes-of-star-wars

NVIDIA. (2026, March 16). NVIDIA announces open physical AI data factory blueprint to accelerate robotics, vision AI agents and autonomous vehicle development. https://nvidianews.nvidia.com/news/nvidia-announces-open-physical-ai-data-factory-blueprint-to-accelerate-robotics-vision-ai-agents-and-autonomous-vehicle-development

Reeves, B., & Nass, C. (1996). The media equation: How people treat computers, television, and new media like real people and places. Cambridge University Press.

Walther, J. B. (1992). Interpersonal effects in computer-mediated interaction: A relational perspective. Communication Research, 19(1), 52–90. https://doi.org/10.1177/009365092019001003

Note: Portions of this document were edited with the assistance of AI-based tools (e.g., ChatGPT and Grammarly) for clarity, grammar, and structure. All concepts, system design, and technical implementation reflect the author's original work.